KERNEL METHODS FOR BIOSENSING APPLICATIONS

By

Hassan Aqeel Khan

A DISSERTATION

Submitted
to Michigan State University
in partial fulfillment of the requirements
for the degree of

Electrical Engineering – Doctor of Philosophy

2015

ABSTRACT

KERNEL METHODS FOR BIOSENSING APPLICATIONS

By

Hassan Aqeel Khan

This thesis examines the design noise robust information retrieval techniques based on kernel methods. Algorithms are presented for two biosensing applications: (1) High throughput protein arrays and (2) Non-invasive respiratory signal estimation. Our primary objective in protein array design is to maximize the throughput by enabling detection of an extremely large number of protein targets while using a minimal number of receptor spots. This is accomplished by viewing the protein array as a communication channel and evaluating its information transmission capacity as a function of its receptor probes. In this framework, the channel capacity can be used as a tool to optimize probe design; the optimal probes being the ones that maximize capacity. The information capacity is first evaluated for a small scale protein array, with only a few protein targets. We believe this is the first effort to evaluate the capacity of a protein array channel. For this purpose models of the proteomic channel's noise characteristics and receptor non-idealities, based on experimental prototypes, are constructed. Kernel methods are employed to extend the capacity evaluation to larger sized protein arrays that can potentially have thousands of distinct protein targets. A specially designed kernel which we call the *Proteomic Kernel* is also proposed. This kernel incorporates knowledge about the biophysics of target and receptor interactions into the cost function employed for evaluation of channel capacity.

For respiratory estimation this thesis investigates estimation of breathing-rate and lung-volume using multiple non-invasive sensors under motion artifact and high noise conditions. A spirometer signal is used as the gold standard for evaluation of errors. A novel algorithm called the segregated envelope and carrier (SEC) estimation is proposed. This algorithm approximates the spirometer signal by an amplitude modulated signal and segregates the

estimation of the frequency and amplitude information. Results demonstrates that this approach enables effective estimation of both breathing rate and lung volume. An adaptive algorithm based on a combination of *Gini* kernel machines and wavelet filtering is also proposed. This algorithm is titled the wavelet-adaptive *Gini* (or WA*Gini*) algorithm, it employs a novel wavelet transform based feature extraction frontend to classify the subject's underlying respiratory state. This information is then employed to select the parameters of the adaptive kernel machine based on the subject's respiratory state. Results demonstrate significant improvement in breathing rate estimation when compared to traditional respiratory estimation techniques.

This thesis is dedicated to my parents.

ACKNOWLEDGEMENTS

I consider the 5 years in grad-school at Michigan State University to be the most challenging yet enlightening and exciting time of my life so far. I am truly blessed to have a wonderful family and academic advisers who have always provided their full support whenever I needed it. This thesis would not have been possible without the support and guidance of my adviser Professor *Shantanu Chakrabartty* and I am very grateful to him for exposing me to some very innovative and novel research problems. Dr. Chakrabartty is one of the smartest people I know and I am amazed by his ability to come up with an endless set of novel research ideas. I have thoroughly enjoyed my time working under his guidance at the AIM lab. I would also like to thank my committee members Dr. *Hayder Radha*, Dr. *Jonathan Hall* and Dr. *Evangelyn Alocilja* for their guidance. The knowledge that I gained in courses they taught was invaluable in helping me tackle the inter-disciplinary problems required for completion of this thesis. They were all very helpful whenever I needed their input on my research problems. I also thank the National Science Foundation and the National Institutes of Health for their generous research grants which enabled me to complete my PhD.

I am highly indebted to my parents for their unwavering love and support; they would only call me on weekends because they did not want to infringe upon my study time. Thank you *Ammee* and *Abbu*, I wish I can become as amazing a parent to my daughter as you both have been to me. My wife *Tooba* has been very supportive throughout this time. Thank you honey for not complaining about the long hours and weekends spent in the lab. Thank you also for taking care of our daughter, Abeer, all on your own back home in Pakistan over the last one year. The most wonderful gift that I have received during my PhD is my daughter; thank you *Abeer Zahra* your wonderful smile and unconditional love make me forget all my worries and troubles. Thanks are also due to my brother *Hassan Jamil* and my sister *Beenish* for taking care of our parents for all the years I have been away from

home. I would also like to thank all my friends and colleagues at MSU, NUST and UET for their help, support and ideas whenever I needed them. Thank you *Jawad, Faraz, Shahzad, Afshan, Samina, Zubair, Momina, Ahmad* and *Abhinav* for being part of some of my most wonderful memories at MSU. Thank you *Usman Ilyas* and *Mohammad Aghagolzadeh* at the WAVES-Lab and *Liang Zhou* at the AIM-Lab for always providing attentive ears to my ideas and giving valuable feedback. Thanks are also due to *Syed Ali Khayam* and Dr. *Arshad Ali* for being wonderful mentors during my time at NUST. A big shout out to my undergrad friends from UET Taxila, the *J-Gang*: *Rao, Iffi, Saqib, Roomi, Qazi, Moodaman, Zim-X, Beela Bhae, Chaudhry* and *Bukhari*; you guys rock and I can't wait to be back in your company.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Whenever possible the following notation will be employed throughout this thesis:

n	–	Lowercase symbols represent scalar values. Generally employed to represent indices of vectors and matrices.
N	–	Uppercase symbols also represent scalar values. Generally employed to represent sizes of vectors and matrices.
$\mathbf{x}$	–	Lowercase boldface symbols represent vectors.
$\mathbf{X}$	–	Uppercase boldface symbols represent matrices.
$\mathbf{x}_i$	–	Represents the i-th vector.
x_i	–	Represents the i-th element of vector $\mathbf{x}$.
x_{ij}	–	Represents the element (i, j) of matrix $\mathbf{X}$.
$\mathbb{R}^D$	–	Represents the D-dimensional space of real numbers.
$\mathbb{R}$	–	Represents the 1-dimensional space of real numbers.
$\mathbb{R}^D_{\geq 0}$	–	Represents the D-dimensional space of positive real numbers, including 0.
$\mathbb{Z}$	–	Represents the 1-dimensional space of integers.
$\mathbb{Z}^D_{\geq 0}$	–	Represents the D-dimensional space of positive integers, including 0.

CHAPTER 1

INTRODUCTION

Modern statistical and pattern recognition techniques have found applications in a diverse range of scientific disciplines. Modern biology for example has benefited tremendously from these inter-disciplinary interactions resulting eventually in the emergence of new disciplines such as bioinformatics and computational biology. This thesis explores the application of statistical learning approaches to enhance the performance of biological and medical sensing systems. In particular, this work focuses on the use of signal processing and kernel methods for reliable information extraction from high-dimensional and noisy data found in protein and respiratory sensing data.

The block diagram of a typical protein sensing framework is shown in Figure 1.1 (a). The goal in this application is to simultaneously detect the concentration levels of multiple target proteins within a test sample. From a communication theoretic perspective this setup can be viewed as multiplexed communication channel. The throughput of this channel depends on a number of factors which may or may not be under our control. These include: diffusion noise, receptor response and saturation characteristics and the Hook effect etc. This thesis is primarily concerned with the impact of diffusion noise and the receiver response characteristics. Ideally, we would like to obtain expressions that demonstrate the effect of receptor parameters on the throughput of multiplexed protein array platforms in the presence of channel irregularities such as diffusion noise. Information and communication theory provide a number of useful tools for evaluating the performance limits of any communication channel in the presence of noise; the foremost being the channel capacity. Enhancing the throughput of any communication channel entails the maximization of its information-theoretic capacity, C, which depends on the conditional probability distribution of the channel, $p(y|x)$. It is generally very difficult to find closed-form expressions for this distribution; this can attributed

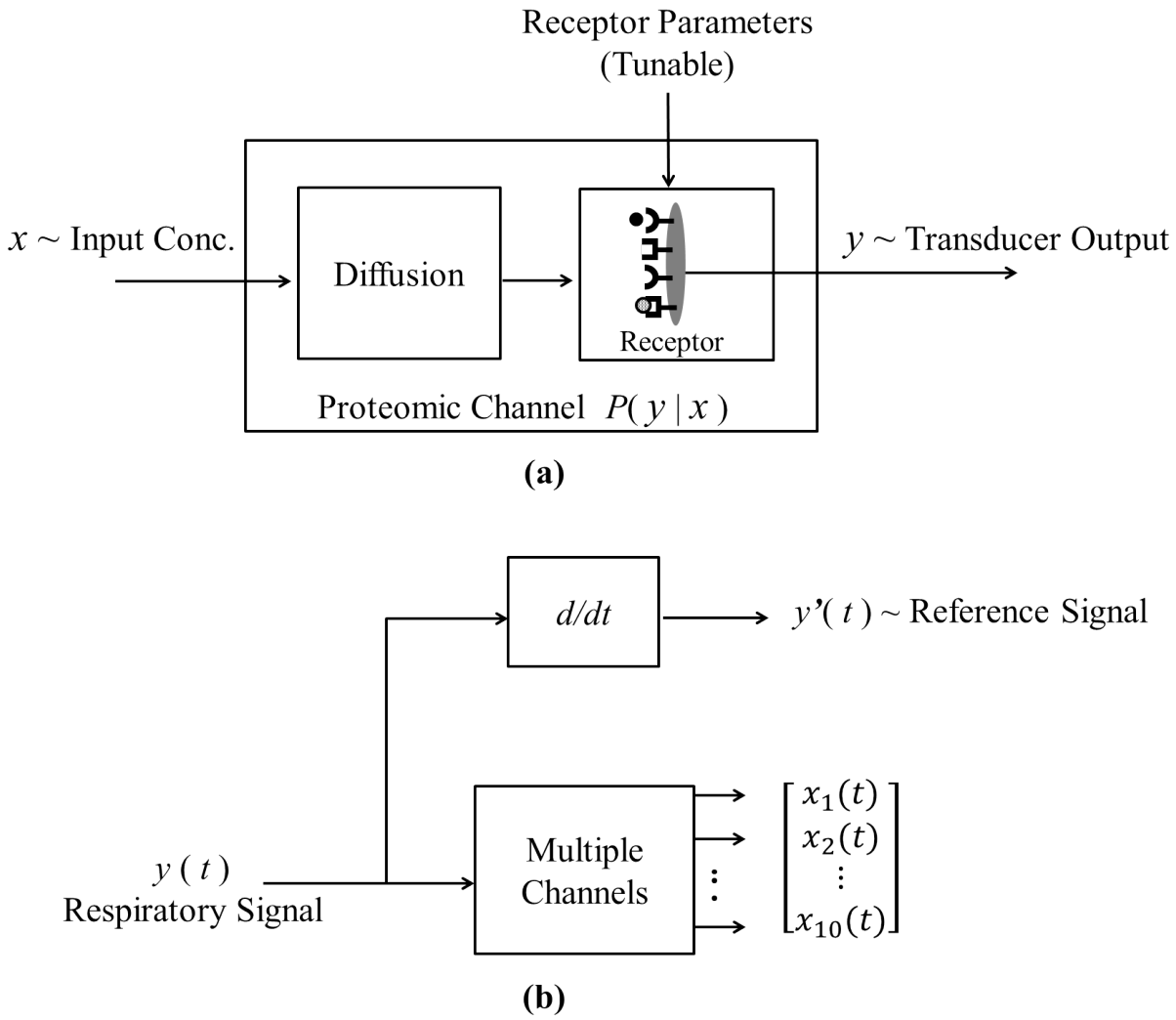


Figure 1.1: Block diagrams of (a) a protein sensing channel and; (b) a respiratory signal estimator.

to the complexity of the diffusion channel and the non-linear nature of the protein receptors. Therefore, the conditional distribution, $p(y|x)$, must be computed via numerical techniques. The numerical techniques employed in this thesis primarily employ Monte-Carlo simulations and kernel methods for evaluation of capacity and similar measure of information.

Figure 1.1 (b) shows the block diagram of a multi-electrode respiratory sensing system. The objective of such a system is to measure a human subject's respiratory parameters such as: *Breathing Rate* (BR) , indicating the frequency at which the subject inhales and exhales, and *Lung Volume* (LV) which corresponds to the volume of air contained within the lungs at

any given instance of time. Such a framework should preferably employ non-invasive sensors in order to avoid causing discomfort to the subject being monitored. Impedance plethysmography is a popular non-invasive respiratory signal estimation technique which operates by placing plethysmographic sensors over the subject's chest and abdomen areas. Under normal breathing conditions the cross-section of the chest and abdomen areas increases during inhalation and returns to a baseline during exhalation [1], this causes a change in the impedance of the attached electrodes resulting in output signals from which respiratory parameters of interest can be extracted. Unfortunately, these sensors suffer from motion artifacts and noise making it difficult to the estimate breathing rate and lung volume especially when the subject performs some physical activity. A potential solution to this problem is to employ multiple sensors so that information from multiple sources may be combined to obtain an artifact free estimate of the desired respiratory information. This thesis uses a number of signal processing and pattern recognition techniques to demonstrates that it is indeed possible to minimize (or diminish) the impact of noise and channel artifacts by using multiple plethysmographic sensors. Also proposed are algorithms based on kernel methods for robust recovery of respiratory parameters in the presence of motion-artifacts. The reference respiratory signal for comparison, and training, is obtained from a *Spirometer* which is immune to motion-artifacts but is invasive and therefore, not feasible for long-term subject monitoring.

This thesis is organized as follows: Motivation for both applications is presented in section 1.1 and section 1.2. Contributions are listed in section 1.3. Chapter 2 presents an approach which employs DCT based filtering and pattern recognition for estimation of breathing rate and (tidal) lung volume. Chapter 3 examines accurate breathing rate estimation using kernel methods. Also presented is an innovative wavelet filtering based front-end which enables detection of different respiratory and physical states of human subjects with high accuracy. Coupled with kernel methods this technique demonstrates significant reduction in the error obtained when estimating breathing rate from impedance-plethysmographic

electrode channels. An information theoretic analysis of the protein array channel is conducted in chapter 4. Optimal probe configurations that maximize information exchange across the proteomic channel are also investigated. Chapter 5 propose a framework based on kernel methods for evaluating the quadratic capacity of the proteomic channel. Chapter 5 also presents a novel *proteomic* kernel which is based on the bio-physical interaction of the receptor probes and the target particles. Conclusions and future work are presented in chapter 6.

1.1 Impedance Plythesmorgaphy For Respirator Signal Estimation

Chronic obstructive pulmonary disease (COPD) is the 3rd leading cause of death worldwide [2] and is a major cause of disability affecting more than 12 million people in the United States [3]. Over 5 million people in the United States (US) are affected by Heart Failure (HF) which accounts for 300,000 deaths per year in the US [4]. Difficulty in breathing and shortness of breath are early indicators of deteriorating patient conditions in both these diseases. There are currently no cures for HF and COPD therefore, continuous monitoring of the respiratory condition in these patients can enable caregivers to intervene at an early stage and manage disease symptoms, forestalling catastrophic events and avoiding loss of precious lives. The two most important parameters extracted from the respiratory signal are the *respiration-rate* and *lung-volume*. Lung-volume is indicative of the size of lungs and the volume of air a patient can breathe in or out, it is the most important factor for detection of COPD [5]. Tachypnoea, or an increase in respiration-rate, can be representative of an attempt by the body to compensate for poor pulmonary gas exchange and/or poor cardiac circulation. It has been demonstrated to be a significant factor in the prediction of cardiac arrest in the ICU [6]. Depression of the respiratory center due to severe deterioration of the patient or narcotic overmedication often corresponds to a decreased respiratory-rate [7]. However, despite the significance of monitoring patient breathing patterns and respiratory rates, these measurements are frequently ignored in clinical practice [1]. Studies of ICU

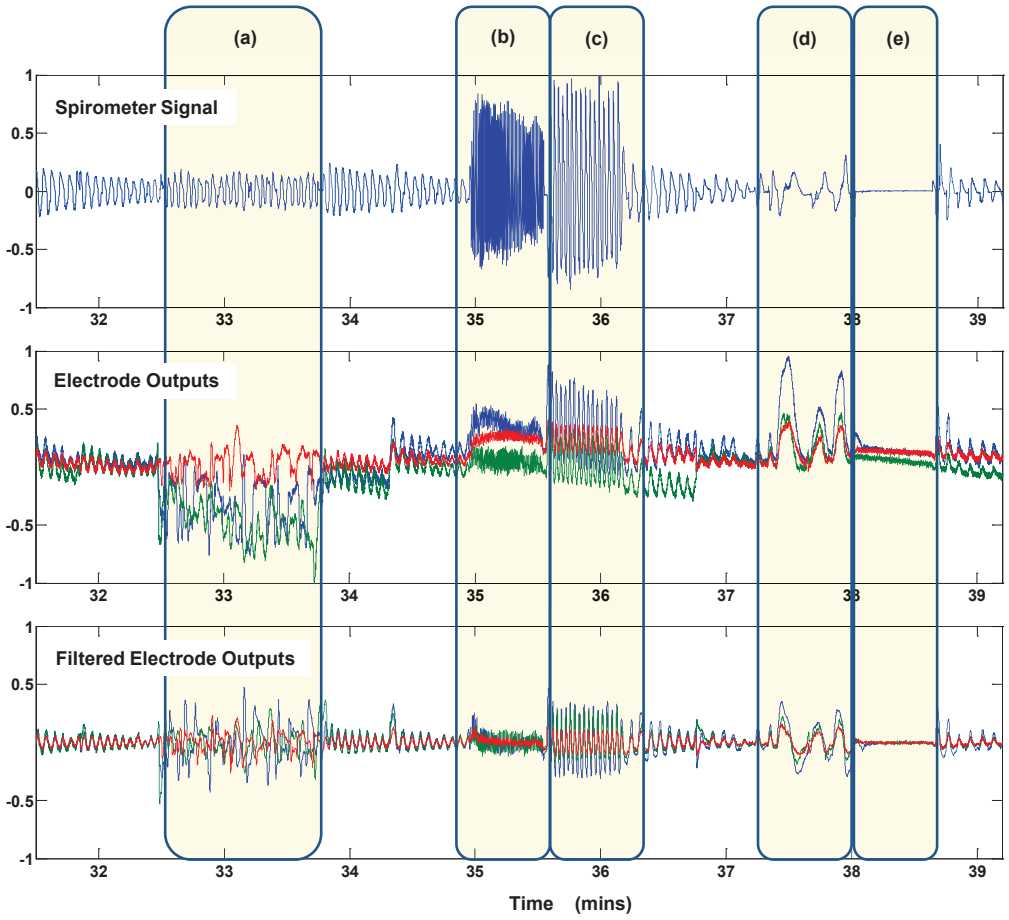


Figure 1.2: Reference Spirometer signal and Electrode outputs (raw and filtered) under different motion and breathing conditions (a) reaching for object; (b) shallow fast breathing; (c) deep fast breathing; (d) deep slow breathing; (e) holding breath.

practices have revealed that in spite installation of vital signs based early warning scoring systems respiratory measurements are neglected over 40% of the time [8]. This can be attributed to the fact that the most accurate respiratory rate measurement methods, such as CO_2 sensors and flow sensors, are difficult to administer and often intolerable for non-intubated ambulatory patients. An alternative is to measure respiratory conditions using impedance based electrodes. This method is not invasive and thus, is more likely to be accepted by both clinicians and patients due to the already routine use of electrode based monitoring systems at hospitals.

Unfortunately, electrode-based respiratory measurements are noisy and therefore, require

post-processing to minimize the impact of noise irregularities. The second major objective of this thesis will be to employ multiple electrodes placed at different spatial locations over the bodies of human subjects to estimate the respiratory signal. A *spirometer* (invasive flow sensor) is used as the "Gold standard" reference signal. The top plot in Figure 1.2 displays the output of a spirometer in different respiratory states. The corresponding outputs of three distinct electrodes are shown the middle plot. Notice, that the electrodes introduce a slow varying DC baseline in the respiratory signal. Furthermore, there is significant distortion in region-(a) due to motion-artifacts that occur when subject reaches for an object. The electrode output after bandpass filtering is shown in the bottom plot. This thesis proposes a number of different techniques for obtaining an accurate estimate of the spirometer/ respiratory signal from electrode outputs by formulating the task at hand as a machine learning problem. A novel technique called "*Segregated Envelope Carrier*" (SEC) estimation is proposed. This approach is based on the hypothesis that the respiration information lies on two distinct manifolds: (1) a high frequency manifold and; (2) a low frequency manifold. The details of this approach are presented in chapter 2. The SEC enables the estimation of both breathing rate and lung volume. It is highlighted that traditional non-invasive approaches to respiratory signal estimation concentrate solely on respiration-rate estimation and generally do not cater for lung-volume estimation. For example a Kalman filter framework was proposed in [9]. This approach estimates the respiration-rate by combining information from multiple physiological sources. Respiration rate can also be derived from the electrocardiogram (ECG); such approaches generally employ algorithms based on the R-peak amplitude (RPA) modulation [10] or the respiratory sinus arrhythmia (RSA) [11], [12]. More recently, an approach combining both RSA and RPA has been proposed in [13]. Although, all the aforementioned techniques achieve high accuracy (in estimating respiration-rate only), they are tested on data collected from non-ambulatory subjects generally resting in a supine position. In contrast, the database employed for this thesis has been created under more challenging conditions and records respiratory signal in both ambulatory and non-ambulatory condi-

Breast Cancer	Sialyl Lewis^x, C3, C4, C5, IL-8, TM-peptide, IL-5, IL-7, MCP-3 CXCL8, IL-8, CXCL1, GRO
Ovarian Cancer	IL-6, IL-8, VEGF, EGF, MCP-1, CA-125 Leptin, Prolactin, Osteopontin, IGF-1, MIF
Prostate Cancer	MCP-1, IL-6, IL-8, GRO-α, ENA-78, CXL-16

Table 1.1: List of cytokines/proteins employed for cancer detection.

tions. There are only a small number of very recent studies that measure both lung-volume and respiration-rate which can be found in [14] and [15].

In addition to the SEC an innovative approach based on the combination of wavelet filtering and kernel methods is proposed in chapter 3. This technique achieves a significant reduction in the breathing rate error under noise and artifact conditions.

1.2 High Throughput Protein Arrays

The human body is thought to contain more than 2 million proteins, each associated with a different biological function [16]. Decoding of these complex biological functions requires detecting and measuring the state of numerous proteins simultaneously. In this regard, high-throughput protein microarrays have become an essential tool which enables rapid, direct, quantitative and multiplexed detection of a multitude of proteins. Applications of the protein microarray technology range from drug development to disease detection and diagnosis. Consider, for example the case for detecting “cytokines" which are signaling proteins that are collectively responsible for a number of physiologic functions and play an important role in many detecting onset of diseases [17]. Using protein microarrays, researchers have been able to uncover new and improved, cytokine based, biomarkers for a number of diseases such as: Alzheimers [18] Parkinsons diseases [19] and many other types of cancers [20]. Table 1.1, lists some examples of cytokine and protein targets that could be used as biomarkers for different types of cancers (Refs: [17, 21–24]). One of the trends and corresponding challenges in the design of protein assays is to be able to simultaneously detect as many biomarkers as possible while minimizing the volume of the sample required for analysis.

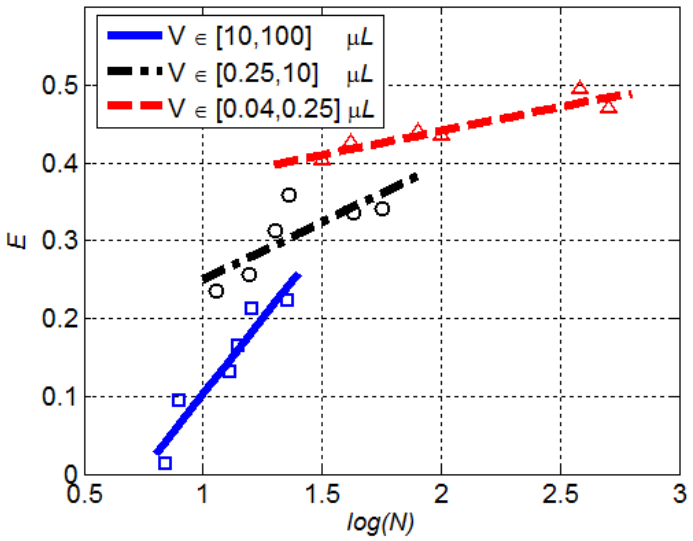


Figure 1.3: Recent trends in protein array construction. N is the number of target proteins; V is the sample volume (in μL) required for a single test; E=N/S is the efficiency of the protein array, where, S represents the total spots on the microarray.

Figure 1.3 shows the specifications of some of the protein microarrays that have been reported during the last 15 years [25–34]. The plot compares the total number of targets that can be simultaneously detected versus the efficiency of the array given by the different ranges of test sample volumes. Figure 1.3 clearly shows that the overall trend has been to enhance the throughput and multiplexing capability of protein arrays by detecting a large number of target proteins while consuming as little of the test sample volume as possible. For instance, one of first the protein arrays was proposed in [25], it employed 504 spots for multiplexed detection of 7 targets ensuring very high redundancy at cost of achieving very low efficiency (plotted on bottom left of Figure 1.3). Recently developed microarrays however, generally employ approximately 2 spots per target and therefore, have an efficiency value around 1/2. Figure 1.3 also indicates that the best arrays (in terms of E and N) also consume the smallest amount of sample volume per target per test.

This thesis investigates the limits of multi-analyte detection capability of a generic proteomic microarray platform based on information theoretic and computational modeling techniques. From an information theoretic point of view, a proteomic platform can be viewed

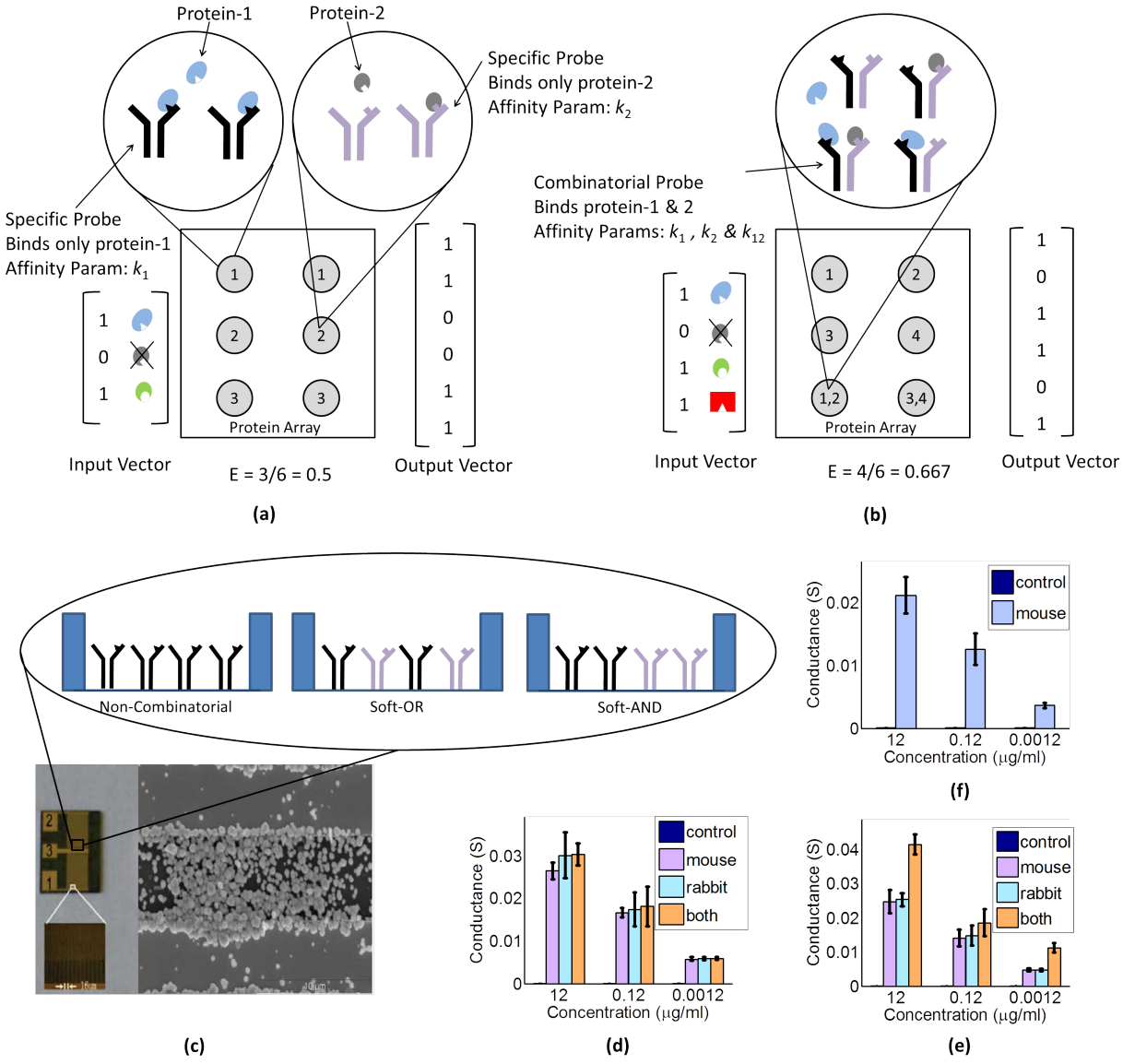


Figure 1.4: (a) Traditional array with microspots specific to only a single protein. Maximum efficiency equal to 0.5 (b) Combinatorial array, contains both specific and combinatorial microspots. Can achieve efficiency greater than 0.5 (c) Scanning electron microscope (SEM) image of previously reported combinatorial spot. Different logic elements are plotted on top of the SEM. Experimentally measured conductance across: (d) a soft-OR receptor for mouse and rabbit IgG; (e) a soft-AND receptor for mouse and rabbit IgG; (f) a conventional (non-combinatorial) receptor specific only to mouse IgG (Figure (c) to (f) adapted from [35,36]).

as a biosensing channel where target proteins (with different concentration levels) constitute the signal being transmitted and the channel noise arise due to different biosensing artifacts like non-specific binding, saturation, hook effect, spot corruption and measurement noise [37, 38]. This concept is illustrated in Figure 1.4(a) for a small scale assay where each of the spots (labeled 1-3) are immobilized by target specific antibody probes. For the sake of simplicity the input to the assay is a binary vector with a "0" indicating absence and a "1" indicating presence of the target protein. Each of the probes comprises of epitopes which are recognition sites that bind with the target protein with some degree of affinity. Thus, a protein-probe hybridization can be viewed as an equivalent "inner-product" between the assay matrix and the input "protein" vector, with the resulting output being a measurable electrical or an optical signal vector.

Conventional microassays and microarrays use multiple spots of antibody probes to improve the reliability of detecting a single target. Thus, from a channel coding point of view this procedure can be viewed as using a *repetition* block-code and existing microarray platform use a *repetition* code to combat channel errors. However, it is well known that the channel capacity of a repetition code is not efficient, especially if the size of the block-code becomes large. In this regard, using a "combinatorial" probe that can bind with different target proteins with different affinities (as shown in Figure 1.4(b)) could be used to enhance the capacity of the assay. Investigating this principle using a computationally efficient approach is one of the main objective of this thesis. This requires development of suitable channel models followed by the evaluation of the channel probability distributions which are required for computing the channel capacity. Models of the proteomic diffusion channel and different types of receptors are derived in chapter 4. Numerical results indicate that capacity of the proteomic channel can indeed be enhanced using combinatorial probes. Furthermore, efficiency of numerical computation of the channel distributions can be improved employing kernel methods.

1.3 Contributions

The primary motivation for this thesis is to employ the principles of pattern recognition and information and signal processing for noise robust retrieval of information from emerging biosensing applications. In this respect this thesis focuses on two applications areas namely: (1) Non-invasive respiratory signal estimation and (2) High throughput protein detection arrays. In both these application significant emphasis is placed on kernel methods. The key contributions are listed below:

1. This thesis employs a novel dataset recorded by General Electric Global research in Niskayuna New York for estimation of respiratory signal parameters. This dataset contains respiratory signals recorded from multiple non-invasive sensors from 19 human subjects. In comparison to datasets employed in existing literature our dataset is unique in the sense that it contains multiple instances of subjects performing various physical activities. Our dataset contains multiple instances of the subjects in different respiratory states such as: apnea, accelerated breathing and hyper-ventilation, slow breathing etc. Existing dataset in contrast generally focus on only one or two types of respiratory states. Therefore, **the respiratory dataset employed in this thesis is unique in the sense that it contains a diverse set of respiratory states and physical activities.**

2. Existing literature in respiratory estimation focuses primarily on breathing rate estimation alone. Lung volume estimation is generally ignored and there are only a few works that have investigated the estimation of lung volume from non-invasive sensors [14] and [15]. However, the dataset employed in these works do not contain any motion artifacts. **This thesis proposes a novel approach called the Segregated Envelope Carrier (SEC) estimation which examines the estimation of both breathing rate and lung volume from non-invasive sensors under both artifact and artifact-free conditions.**

3. A set of novel features based on the discrete wavelet transform is proposed. These features provide a simple method for classification of the subject's respiratory and physical states. This thesis employs these features for detection of artifacts, apnea, accelerated and normal breathing regions. **To the best of the author's knowledge this is the first time that these type of features have been employed for classification of respiratory and physical states.**

4. An adaptive framework based on *Gini* kernel machines is proposed for detection of breathing rate estimation. This is titled the Wavelet-Adaptive-*Gini* (or WA*Gini*) algorithm for breathing rate estimation. This algorithm employs wavelet based features for respiratory state classification and uses the classifier's decision for selecting a kernel machine that has been trained specifically for the underlying respiratory state. **Evaluation of the output indicates that the WA*Gini* algorithm enables significant reduction in the breathing rate estimation error. The performance improvement obtained is significant in comparison to standard rate estimation technqiues.**

5. For protein array sensing this thesis evaluates the impact of various channel irregularities on information transfer between the input and output of the affinity based protein array sensors. For this purpose a protein array is viewed as a communication channel and its channel capacity is evaluated. **To the best of the author's knowledge this is the first effort undertaken to evaluate the information transmission capacity of a protein array channel.**

6. Capacity evaluation of the protein array channel entails modeling of the various irregularities that can have an adverse impact on the information of interest. **For this purpose models of diffusion processes and receptor artifacts, based on experimental prototypes constructed in lab, are presented.** Existing literature investigating the capacity of biological communication channels generally employ ideal

models of receptors; see for example [39].

7. Evaluation of shannon's channel capacity becomes challenging when dealing with non-linear channels with continuous and high-dimensional input alphabets. An alternative can be to employ metrics such as a quadratic form of mutual information which in practice are more amenable to optimization. In this context an optimization framework based on a quadratic information measure is proposed in chapter 5. Of particular importance in this framework is the use of a novel kernel which we call the *Proteomic Kernel*. **The proteomic kernel is designed specifically to capture the biophysical interactions of the receptor probes and the target protein particles. This enables the easy extension of protein array designs to a large number of target proteins.** It is envisioned that the theoretical discussion provided in this thesis will eventually lead to software tools that will enable prompt and cost-effective design of high-throughput protein arrays.

CHAPTER 2

RESPIRATORY SIGNAL ESTIMATION

Accurate respiratory signal estimation using impedance plethysmography can be challenging under certain conditions such as; during patient motion or under noisy conditions at high breathing rates. This chapter discusses a number of different signal processing and learning techniques for estimation of breathing rate and lung volume from multiple non-invasive impedance plethysmographic electrode channels. The organization of this chapter is as follows: section 2.1 describes the hardware employed for obtaining respiratory data from different human subjects; it also details the different conditions under which the data was collected. Section 2.2 discusses the salient characteristics of the respiratory signal, this is done to provide theoretical background and gain insights about what strategy to employ to obtain a good estimate of the respiratory signal from the electrode outputs. Section 2.3 contains details of the different regression techniques employed to predict the respiration/spriometer signal from the electrode outputs. A total of four different regression techniques have been employed; they are listed in sections 2.3.1 to 2.4. The first two approaches are based on conventional techniques such as Support Vector regression (SVR) and Gaussian mixture regression (GMR). A simple scheme based on DCT filtering is discussed in section 2.3.3. A novel approach titled:"Segregated Envelope Carrier" (SEC) is proposed in section 2.4. It operates on the assumption that respiratory information is contained in two distinct manifolds: (1) Envelope Manifold containing slow varying temporal information and (2) The Carrier Manifold containing the relatively faster varying temporal information. Consolidated results for all human subjects are summarized in section 2.5.

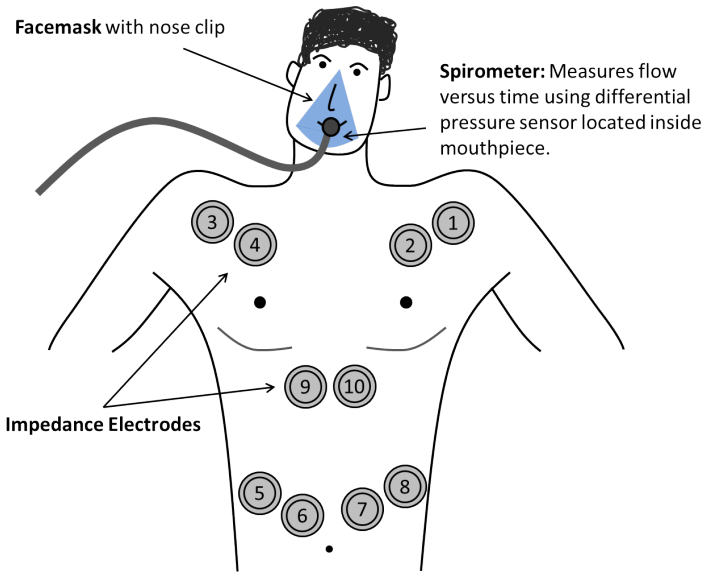


Figure 2.1: Configuration employed for measurement of respiratory signal from human subjects. The *spirometer* employs a differential pressure sensor (placed inside a tube over the mouth) to measure flow versus time. Multiple *impedance plethysmographic sensors* placed over the torso measure changes in lung volume versus time.

2.1 Multi-lead impedance plethysmography

The most reliable and accurate methods of measuring the breathing rate employ instruments such as spirometers that measure the changes in the airflow directly from the patient's airway. The spirometer (or flowmeter) setup employed during data recording is illustrated in Figure 2.1; it uses a differential pressure sensor placed inside a tube located over the subject's mouth. The subject's nose is blocked using a nose clip so that only the air flow to and from the mouth is captured. The difference in airflow to and from the mouth is measured by the differential pressure sensor which the produces the time-series signal $y(t)$ corresponding to the variation of airflow into and out of the lungs as function of time. An alternate sensing mechanism that can be employed to measure the respiratory signal uses impedance-plethysmographic sensors placed over the subject's torso. These sensors operate by capturing a subject's chest motion as it inflates and deflates during inspiration and expiration. An impedance plethysmographic electrode sensor measures variations of the changes in the air volume inside the subject's lungs as a function of time and its output $x(t)$

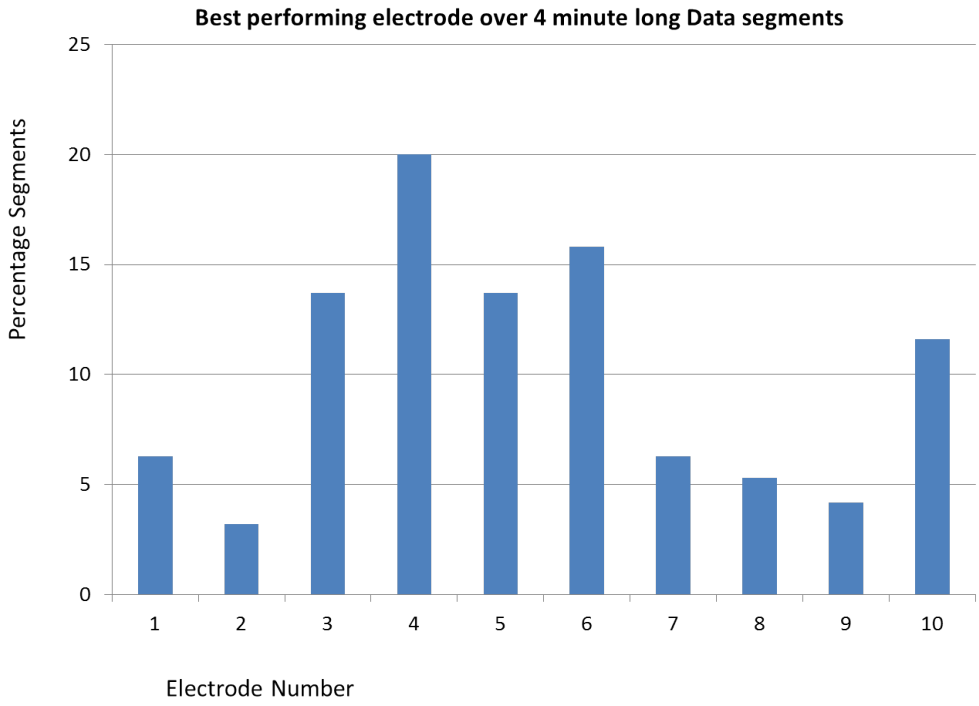


Figure 2.2: Percentage of data during which different impedance-electrodes give the lowest error rate.

is therefore, the integral of the flow signal output from the spirometer:

$$x(t) = \int y(t)dt \tag{2.1}$$

Due to its non-invasive nature the plethysmographic sensing mechanism is very appealing for long-term and remote monitoring of patients. Unfortunately, this mechanism is prone to motion artifacts [40] since essentially any activity by the subject such as arm movement etc can also be measured by a plethysmographic sensor and can therefore interfere with the respiratory information. A potential solution to this problem is to employ multiple sensing electrodes placed at different spatial locations over the patient's body. Since the respiratory signal is correlated among all the impedance electrodes and the patient movements are sporadic and generally not correlated we should be able to separate the respiratory signal from motion artifacts using a multi-sensor setup. A simple procedure to demonstrate the potential advantage of using multiple impedance-electodes is to divide the respiratory data in to, non-overlapping, segments of equal time duration and then compute the percentage of

segments during which the breathing rate obtained from any particular chest-sensor is closest to the *reference* breathing rate obtained from the spirometer. In this way we can compute that percentage of segments during which a certain electrode gives the best performance (or the smallest breathing rate error). The bar-graph in Figure 2.2 plots the percentage of segments during which each of the 10 chest-sensors gives the best performance. It can be observed that there is no clear winner and the most accurate chest-sensor gives the best performance in only about 20% of the total segments. This is not unexpected since the data is not artifact free and the degree of impact of a motion artifact on an impedance-electrode depends on the nature of the underlying physical activity and the electrode's location. For example, a sudden right arm movement is more likely to distort outputs of electrodes on the right side of the torso than it is to distort the left side sensors. As a result there seems to be no single electrode-sensor that gives the best performance across all the different activities contained in the various segments of data. Therefore, it seems likely that multiple impedance-electrodes may enable us to minimize/eliminate the impact of motion-artifacts on respiratory signal estimation.

2.1.1 Data Collection And Pre-processing

The experiments in this chapter are based on 11 respiratory datasets each of which was recorded from a distinct adult human subject. An additional 8 subjects are added for the results in the next chapter. *Data was collected by General Electric (GE) global research at their Niskayuna NY location. The data collection protocol was approved by GE's institutional review board.* A total of 10 impedance electrodes were placed at different spatial locations on a human subject's torso as shown in Figure 2.1. As mentioned previously, a *spirometer* (Model RX137F Biopac Inc. Goleta, CA) is used as the reference respiratory signal. The spirometer and electrodes were switched on and off by two different operators and the output signals were aligned manually. Excess pre- and post-samples were truncated. The spirometer and electrode hardware have different sampling rates therefore, interpolation was employed to

waveforms with identical sampling rates. Each individual dataset is approximately 50 minutes in duration. During recording the human subjects were instructed to maintain different positions/postures such as: sitting in chair, laying face-up on bed and standing etc. Furthermore, subject was told to achieve distinct *respiratory-states* such as: normal breathing, deep breathing, shallow fast breathing, deep fast breathing, coughing, yawning and holding breath etc; while simultaneously maintaining different postures/positions . Each dataset also incorporated motion artifacts by recording the respiratory signal while the patient performed different *physical-activities* such as: reading, eating, walking, reaching to grab object etc. Consider for example, interval-(a) in Figure 1.2 (on page 6) where the subject is reaching for an object while breathing normally (as indicated by the spirometer signal) in a seated position. Motion artifacts in this interval cause significant distortion in the electrode signals. During intervals-(b) through (e) the subject is laying, face-up, on a bed and maintaining different respiratory-states each for a duration of approximately 30 seconds.

In addition to motion artifacts impedance electrode outputs include a changing DC-baseline (see Figure 1.2 middle plot). This can be attributed to slight shifts in electrode position over time. Therefore, the first pre-processing step is to apply a DC blocking filter to the electrode outputs. After this a lowpass FIR filter is applied to eliminate high frequency interference and noise. The bottom plot in Figure 1.2 displays the filtered electrode signals and demonstrates that simple filtering eliminates the DC-baseline and high frequency noise.

2.2 Respiratory Signal Characteristics

This section provides a brief background about the important information contained within the respiratory signal output from the spirometer and examine its time-frequency characteristics. A typical spirometer employs a mouthpiece to directly measure the airflow in the lungs during inspiration and expiration [41]. The breathing-rate is contained in the spirometer frequency whereas, the lung-volume can be obtained by integrating the spirometer ouput. Therefore, it is critical to preserve both the frequency and amplitude of spirometer in order

to produce an accurate estimate of the respiration-rate and lung volume. As a result, we approximate the spirometer output by an *Amplitude-Modulated* (AM) signal. The harmonic nature of the Spirometer signal implies that it can be represented efficiently using a harmonic basis such as the Discrete Fourier Transform (DFT) or the Discrete Cosine Transform (DCT). If true this may provide us with a method to construct sparse features for signal regression. To examine the harmonic nature of the respiratory signal we take a sample Spirometer output and subdivide it into, non-overlapping, frames (or windows) of length $M = 200$ samples each. Given a total of F non-overlapping frames in the spirometer output, it is transformed using a DCT basis

$$\mathbf{z}_i = \mathbf{T}\mathbf{u}_i \quad \text{for } i = [1, ..., F] \tag{2.2}$$

where, $\mathbf{T}$ represents the $(M \times M)$ DCT basis. The vector $\mathbf{u}_i \in \mathbb{R}^M$ represents the i-th frame of the spirometer waveform and $\mathbf{z}_i \in \mathbb{R}^M$ represents the corresponding vector of coefficients obtained after application of the DCT transform. A quantized/ distorted estimate of the spirometer signal is then obtained using the inverse DCT transform as below

$$\widetilde{\mathbf{u}}_i = \mathbf{T}^{-1}\widetilde{\mathbf{z}}_i \quad \text{for } i = [1, ..., F] \tag{2.3}$$

where, $\widetilde{\mathbf{z}}_i$ is the vector that is obtained by retaining only the $N \leq M$ coefficients in $\mathbf{z}_i$ that have the largest absolute values; the remaining coefficients are set to zero. For a given value of N quality of the reconstructed signal is evaluated by computing its average Signal-to-Distortion Ratio (SDR) as below:

$$\text{SDR} = \frac{1}{F}\sum_{i=1}^{F} 20\log\frac{||\widetilde{\mathbf{u}}_i||}{||\mathbf{u}_i - \widetilde{\mathbf{u}}_i||} \tag{2.4}$$

where, $||\ ||$ represents the l_2-norm. Figure 2.3 displays a plot of the Signal-to-Distortion Ratio (SDR) for different values of N.

The length M of each frame $\mathbf{u}_i$ is equal to 200 samples and therefore, the maximum value of N can be equal to 200. However, we are interested in the SDR values at small values of N, hence Figure 2.3 displays SDR only upto a maximum of $N = 30$ coefficients.

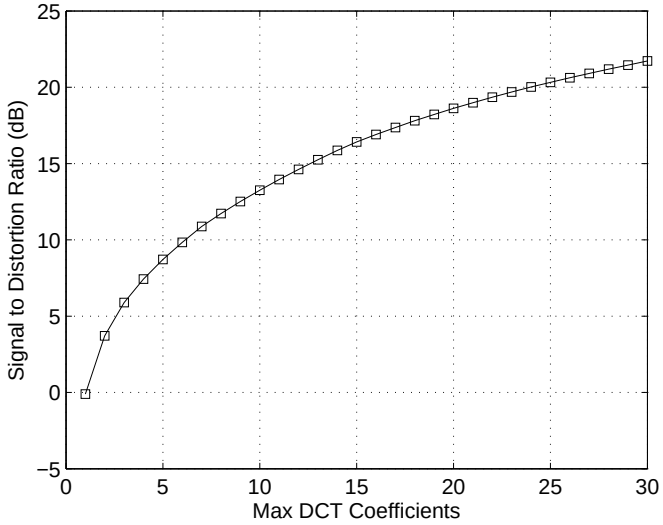


Figure 2.3: Signal-to-Distortion Ratio of a Spirometer Signal Reconstructed for a fixed number of DCT coefficients.

Figure 2.3 indicates that a reasonable SDR (>10 dB) can be achieved even using a small number (10 to 15) of DCT coefficients. It is highlighted here that we are primarily interested in extracting the frequency and the amplitude; other factors such as the exact shape of the waveform are not critical and therefore, can be compromised at the cost of preserving these two parameters. Therefore, the SDR may not be the best metric to measure the quality of the spirometer signal since even a low quality SDR signal may be acceptable as long as it preserves the critical parameters. Hence, SDR is only employed in this section for demonstration purposes, the final results are evaluated using different metrics (described in section-2.3).

A demonstration of the AM approximation is presented in Figure 2.4 where the top figure contains the reference spirometer signal. Figure 2.4 (b) contains a plot of the quantized spirometer signal, $\widetilde{\mathbf{u}} = [\widetilde{\mathbf{u}}_1, \ldots, \widetilde{\mathbf{u}}_F]$, obtained by retaining only the single largest DCT coefficient in each frame i.e.; $N = 1$. The SDR obtained for this signal is equal to -0.104 dB. The plot in Figure 2.4 (c) shows the signal obtained via the AM approximation. In this case the carrier (breathing rate) component is obtained by setting the spirometer DCT coefficient with the largest absolute value (in each frame) to 1. The remaining $M - 1$

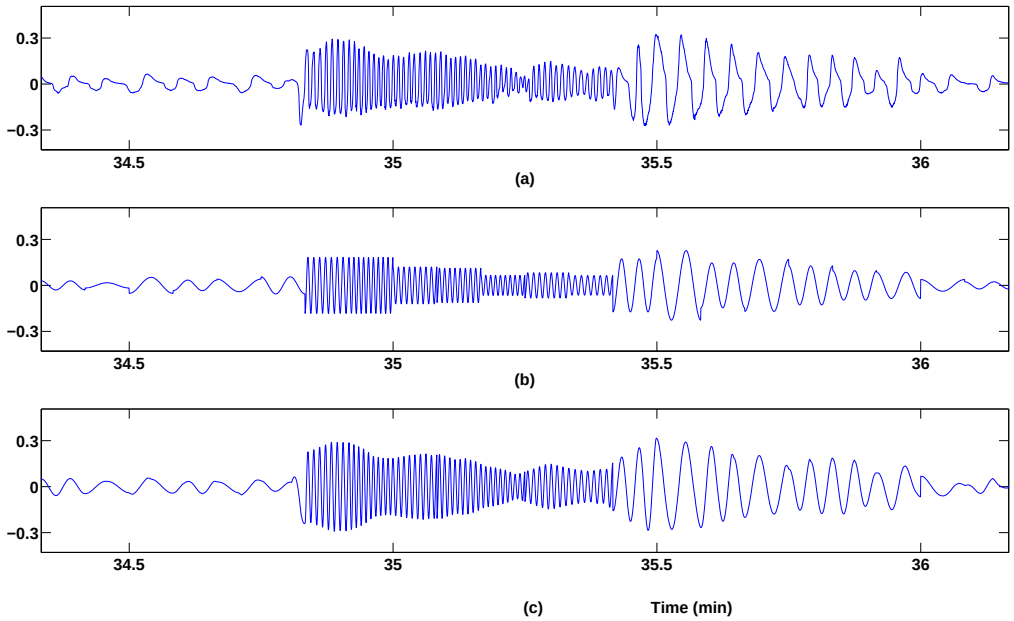


Figure 2.4: (a) Original Spirometer signal. (b) Reconstructed signal using only 1 DCT coefficient; SDR = -0.104 dB. (c) Reconstructed signal using the AM approximation; SDR = 5.483 dB.

coefficients are set to 0. This carrier component is then multiplied with the, ideal, envelope obtained from the spirometer signal in Figure 2.4 (a) to obtain the AM approximation displayed in Figure 2.4 (c). The SDR in this case is equal to 5.483 dB therefore, moving from the single DCT coefficient to the AM-approximation results in a gain of about 5.5 dB. The AM-approximation also uses only a single DCT coefficient however, it also multiplies the coefficient with the envelope. Therefore, it seems that the AM approximation is a fair assumption. Note that use of the "ideal" envelope here is only for demonstration purposes; for the SEC algorithm proposed in this chapter the envelope component is learnt from the spirometer signal during the training phase and predicted from electrode outputs during the test phase.

2.3 Spirometer Signal Regression

The multi-lead plethysmography system employed for data collection consists of a total of 10 impedance electrodes strategically placed at different spatial locations over the patient's body. As mentioned previously, a spirometer was employed to capture the patient's true respiratory state. On average each subject's dataset consisted of 95,000 samples. Each dataset was split into 9 non-overlapping sets out of which 8 were employed for training and 1 was used for testing at one time. A number of different regression techniques were tested to obtain the best estimate of the spirometer signal. The following subsections describe in detail some of the regression approaches that were employed. Due to space constraints it is not possible to plot all of the reconstructed test signals. Therefore, only 4 test signals for each regression technique are plotted here. For consistency and judicious comparison the same set of test signals is plotted for all the regression methods presented in the following subsections. Results for all the Datasets are summarized at the end in Tables 2.2 and 2.1. The critical parameters here are the signal's *breathing rate* and *envelope*. Therefore, to evaluate the quality of the estimated respiratory signal the following performance metrics are employed:

2.3.0.1 Average Breathing-Rate Error ($BR_{\mathbf{err}}$)

The accuracy of the estimated respiration rate is evaluated by comparing the predicted signal with the reference spirometer signal. More specifically both, the estimated and reference, signals are divided into 60 sec frames and the respiration rate is calculated by identifying the highest energy frequency component in their respective spectrograms. The spectrogram was evaluated using 60 sec long Gaussian windows with an overlap of 25 sec between successive windows. The average breathing rate error (BR_{err}), in breaths-per-minute (BPM), is then computed by first taking the absolute difference between the reference and estimated breathing-rate curves and then averaging over the total number of frames.

2.3.0.2 Envelope Correlation Coefficient (E_ρ)

The correlation-coefficient is employed to quantify the relationship between the *temporal variations* of the envelopes, of the reference and the estimated respiratory signals. In particular, this metric is critical for evaluation of test signals that contain a mixture of different respiratory states (such as the signal in Figure 1.2) where the envelope exhibits significant temporal variations.

2.3.1 Support Vector Regression (SVR)

The description of support vector machines in this section is based on Smola et. al's tutorial [42]. Support vector machines (SVM) [43], [44] are amongst the most popular and widely applied tools in regression problems. Given a set of input training vectors, $[\mathbf{x}_1, ..., \mathbf{x}_l]$ belonging to input space $\mathcal{X}\left(= \mathbb{R}^d \text{ in the current context}\right)$, and corresponding training labels, given by $[y_1, ..., y_l] \in \mathbb{R}$. Support-Vector Regression attempts to finds a function $f(\mathbf{x})$ that has at most ϵ deviation from all training values, y_i, and is as flat as possible [42]. In a linear formulation the function $f(\mathbf{x})$ is assumed to have the following form:

$$f(\mathbf{x}) = \mathbf{w}^t\mathbf{x} + b \tag{2.5}$$

where, $\mathbf{w} \in \mathcal{X}$ and $b \in \mathbb{R}$. For functions of the type in (2.5) *Flatness* corresponds to seeking a small $\mathbf{w}$. This may be achieved by minimizing the norm of $\mathbf{w}$. Thus one can solve for $\mathbf{w}$ within the framework of convex optimization. However, it is possible that a function, $f(\mathbf{x})$, that satisfies ϵ-deviation constraint for all pairs $(\mathbf{x}_i, y_i)$ may not exist. Therefore, slack variables ξ_i, ξ_i^* are introduced to tolerate some errors to make the optimization feasible. Minimization of $|\mathbf{x}|$ subject to the constraints discussed above can now be formulated as the following optimization problem [43]:

$$\min \frac{1}{2}|\mathbf{w}|^2 \;+\; C\sum_{i=1}^{l}(\xi_i+\xi_i^*) \tag{2.6}$$

$$\text{subject to}\begin{cases} y_i-\mathbf{w}^t\mathbf{x}_i-b & \leq \epsilon+\xi_i \\ \mathbf{w}^t\mathbf{x}_i+b-y_i & \leq \epsilon+\xi_i^* \\ \xi_i\xi_i^* & \geq 0 \end{cases}$$

The trade-off between the error-tolerance and the flatness of $f(\boldsymbol{x})$ is controlled by the constant $C>0$ [42]. *Non-Linear* support vector regression operates by mapping the training instance $\mathbf{x}_i$ into a (generally higher dimensional) feature space $\mathcal{S}$ using the map $\mathbf{\Phi}:\mathcal{X}\rightarrow\mathcal{S}$ and then applying the standard support vector regression algorithm.

The *primal* objective function of (2.6) can be solved more easily in its dual form by making use of a Lagrangian function. The *dual* of the primal in (2.6) generalized to the non-linear case is given by:

$$\max\begin{cases} -1/2\sum_{i,j=1}^{l}\left(\alpha_i-\alpha_i^*\right)K\left(\mathbf{x}_i,\mathbf{x}_j\right) \\ \\ -\epsilon\sum_{i=1}^{l}\left(\alpha_i+\alpha_i^*\right)+\sum_{i=1}^{l}y_i\left(\alpha_i-\alpha_i^*\right) \end{cases} \tag{2.7}$$

$$\text{subject to }\sum_{i=1}^{l}\left(\alpha_i-\alpha_i^*\right)\text{and }\alpha_i,\alpha_i^*\in[0,C]$$

where, $\alpha_i\geq 0\text{and}\alpha_i^*\geq 0$ represent the Lagrange multipliers; $K\left(\mathbf{x}_i,\mathbf{x}_j\right):=\mathbf{\Phi}\left(\mathbf{x}_i\right)^t\mathbf{\Phi}\left(\mathbf{x}_j\right)$ is called the *Kernel*-function. The dual in (2.7) depends only on the dot product in the feature space and therefore, can be solved without explicitly computing of $\mathbf{\Phi}\left(\mathbf{x}_i\right)$. In the non-linear case, $\boldsymbol{x}\text{and }f(\boldsymbol{x})$ take the following form:

$$\mathbf{w}=\sum_{i=1}^{l}\left(\alpha_i-\alpha_i^*\right) \tag{2.8}$$

$$f(\mathbf{x})=\sum_{i=1}^{l}\left(\alpha_i-\alpha_i^*\right)K\left(\mathbf{x}_i,\mathbf{x}\right)+b \tag{2.9}$$

Therefore; after training, predictions using future, or test, data vectors can be made using equation (2.9). For SV regression the open-source LIBSVM toolbox [45] is employed. The kernel used is the radial basis function (RBF) kernel, $K(\mathbf{x}_i, \mathbf{x}) := exp(-\gamma|\mathbf{x}_i - \mathbf{x}|)$. The dimension, d, of the input feature vectors is equal to 20. Each feature vector contains 10 electrode samples (one corresponding to each of the 10 electrodes) and an additional 10 values containing the delta coefficients of each electrode. The delta coefficients correspond to the 1^{st} derivative of the electrode time-series outputs and are employed here to capture the temporal dependence on the preceding samples. In order to capture temporal dependence we also, experimented with features based on the auto-regressive (AR) model however, the results were not encouraging and therefore, are not presented here. The parameter C and γ parameter, of the RBF kernel, were selected by performing a grid-search on a large collections of test signals from different patients; the values that gave the best performance (in terms of the performance metrics described above) were selected for the rest of the simulations.

The estimated time series obtained via SVR for four test signals are shown in Figures 2.5(b) - 2.8 (b). The reference spirometer output is also shown for comparison. Respiratory signal estimation is generally more challenging at higher breathing rates therefore, we selected test signals that contain instances of apnea, normal and accelerated breathing. Furthermore, to demonstrate the impact of motion artifacts, three out of the four test signals also contain regions where the subjects are performing a physical activity. For Test Signal-1 (Figure 2.5) SVR results in an overall $RR_{err} = 4.58BPM$ however, the envelope correlation coefficient is only 0.338. Additionally, there seems to be significant degradation in the envelope and rate estimation in due to motion artifacts when the subject is physically active. For Test Signal-2 the $RR_{err} = 17.61BPM$ which is quite high and there also seems to be noticeable degradation in the motion artifact region. For Test Signal-3 the RR_{err} is almost zero however, this signal does not contain any regions of physical activity and envelope correlation correlation coefficient is still quite low. Test Signal-4 follows a similar trend and there appears to be significant degradation in the region containing physical activity.

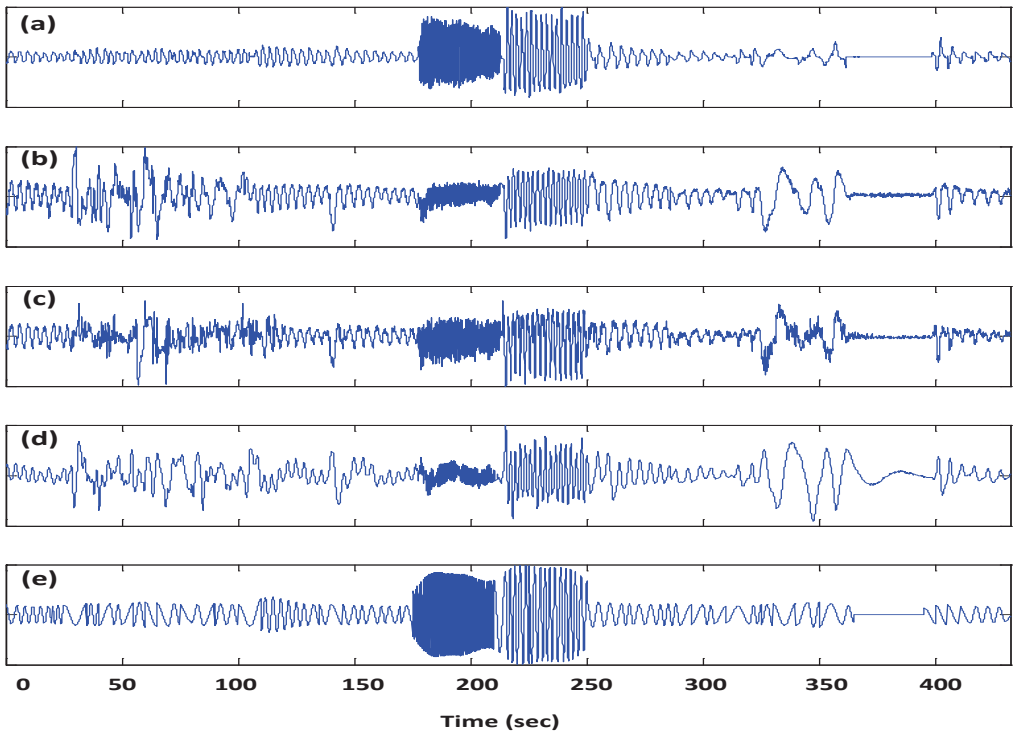


Figure 2.5: *Test Signal-1*; time series obtained from: **(a)** Reference spirometer **(b)** SVR ($RR_{\text{err}} = 4.58$ BPM, $E_\rho = 0.338$) **(c)** GMR ($RR_{\text{err}} = 5.92$ BPM, $E_\rho = 0.771$) **(d)** DCT based estimation, ($RR_{\text{err}} = 6.58$ BPM, $E_\rho = 0.327$) and **(e)** SEC ($RR_{\text{err}} = 2.79$ BPM, $E_\rho = 0.989$). Subject performing physical activity between 0 to 100 sec.

Therefore, there seems to be a significant margin for improvement and other alternatives must be investigated.

2.3.2 Gaussian Mixture Regression (GMR)

Given the harmonic nature of the respiratory signal, a Gaussian Mixture based approach seems to be a very appealing option. For example, Gaussian mixture models (GMMs) are one of the most widely employed methods in speech based applications [46], [47] where the underlying signal contains complex harmonic information. GMMs are based on the assumption that the underlying distribution of the data can be approximated by a multi-modal Gaussian distribution. A single instance of the feature vector, $\mathbf{x}$, at the input of the Gaussian mixture model is $d(= 20)$ dimensional and consists of the electrode outputs

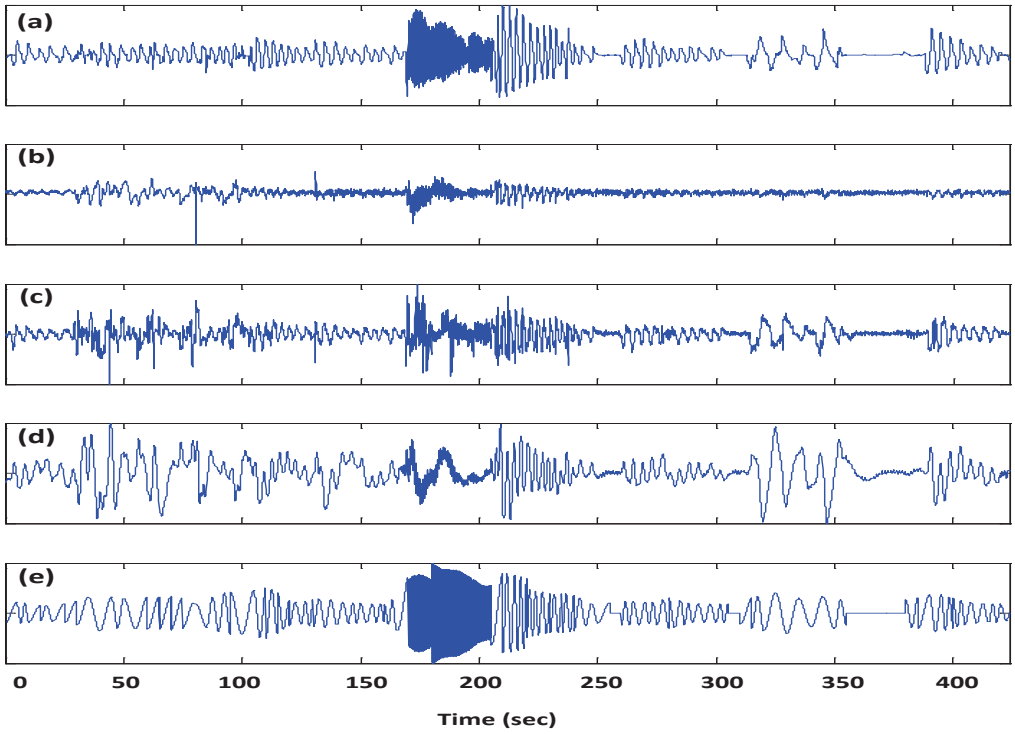


Figure 2.6: *Test Signal-2*; time series obtained from: **(a)** Reference spirometer **(b)** SVR ($RR_{\text{err}} = 17.61$ BPM, $E_\rho = 0.572$) **(c)** GMR ($RR_{\text{err}} = 9.38$ BPM, $E_\rho = 0.712$) **(d)** DCT based estimation, ($RR_{\text{err}} = 13.83$ BPM, $E_\rho = 0.384$) and **(e)** SEC ($RR_{\text{err}} = 2.80$ BPM, $E_\rho = 0.868$). Subject performing physical activity between 0 to 100 sec.

plus their corresponding delta coefficients (at one time sample). Thus the feature extraction block is identical to the one employed for SVM regression in section 2.3.1. The Gaussian mixture density of an $(d+1)$ dimensional multivariate random variable $\mathbf{\Phi} = [\mathbf{x}, y]$, obtained by concatenating $\mathbf{x}$ and the (1-dimensional) spirometer output y, is given by [48]:

$$p(\mathbf{\Phi}) = p(\mathbf{x}, y) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{\Phi}; \boldsymbol{\mu}_k, \mathbf{\Sigma}_k) \tag{2.10}$$

where, K is the total number of Gaussian components, π_k are non-negative mixing components with $\sum_{k=1}^{K} \pi_k = 1$ and $\mathcal{N}(\mathbf{\Phi}; \boldsymbol{\mu}_k, \mathbf{\Sigma}_k)$ represents a multi-variate Gaussian density. Furthermore, $\boldsymbol{\mu}_k$ denotes the mean vector and is given by:

$$\boldsymbol{\mu}_k = \begin{pmatrix} \boldsymbol{\mu}_{k\mathbf{x}} \\ \boldsymbol{\mu}_{ky} \end{pmatrix} \tag{2.11}$$

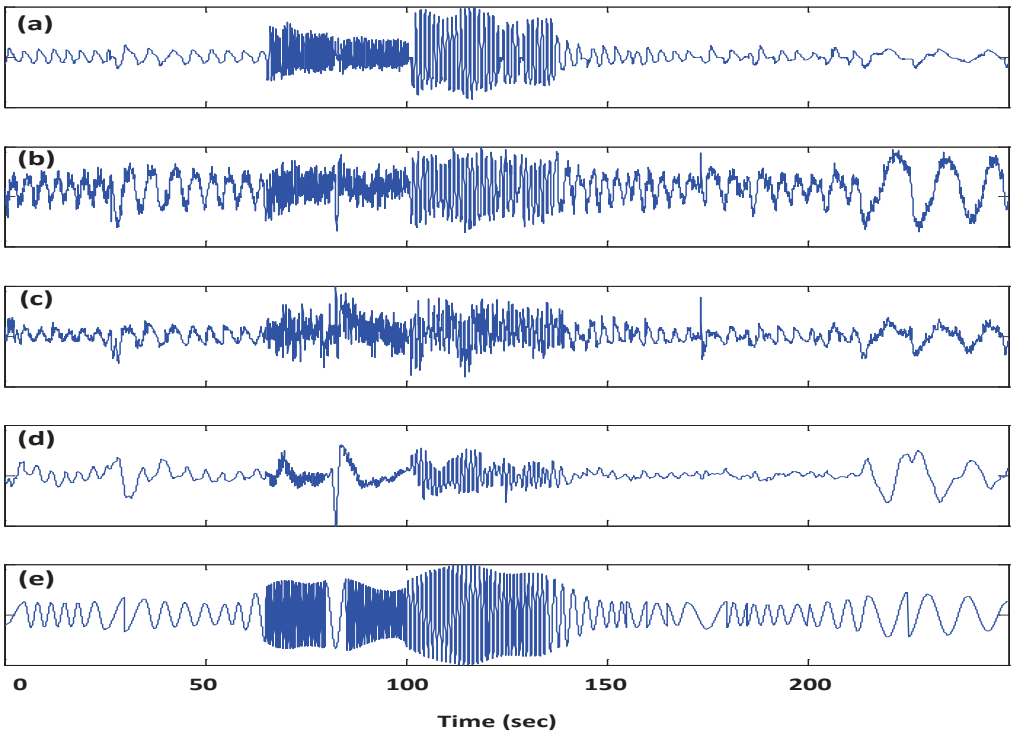


Figure 2.7: *Test Signal-3*; time series obtained from: **(a)** Reference spirometer **(b)** SVR ($RR_{\text{err}} = 0.396$ BPM, $E_\rho = 0.417$) **(c)** GMR ($RR_{\text{err}} = 16.21$ BPM, $E_\rho = 0.846$) **(d)** DCT based estimation, ($RR_{\text{err}} = 16.21$ BPM, $E_\rho = 0.343$) and **(e)** SEC ($RR_{\text{err}} = 3.03$ BPM, $E_\rho = 0.947$). No physical activity at any time.

$\mathbf{\Sigma}_k$ represents the covariance and is given by:

$$\mathbf{\Sigma}_k = \begin{pmatrix} \mathbf{\Sigma}_{k\mathbf{xx}} & \mathbf{\Sigma}_{k\mathbf{x}y} \\ \mathbf{\Sigma}_{ky\mathbf{x}} & \mathbf{\Sigma}_{kyy} \end{pmatrix} \tag{2.12}$$

After initialization using K-means clustering [49] the EM algorithm [50] is employed to find the parameters of the Gaussian mixture distribution(of equation (2.10)) that best fits the training data. The number of mixture components (K) is determined using the Bayesian-Information Criterion (BIC).

The conditional distribution, $p_k(y|\mathbf{x})$, of a component k, of the spirometer output y given the input feature vector $\mathbf{x}$ is determined by dividing the joint distribution, $p_k(\mathbf{x}, y)$, by the

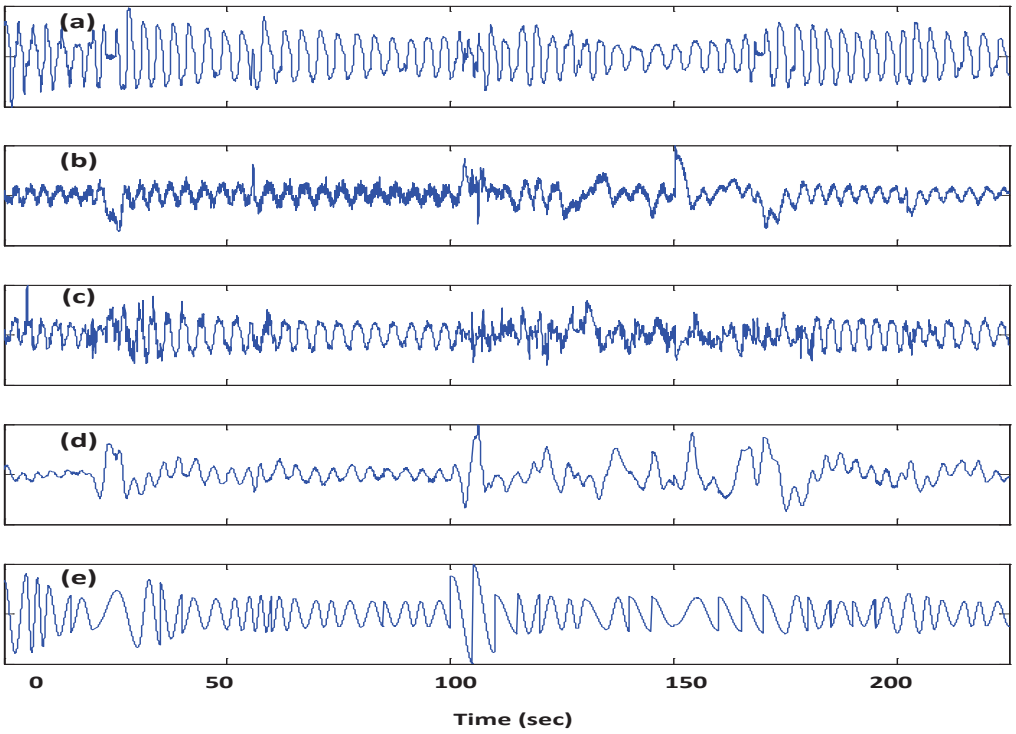


Figure 2.8: *Test Signal-4*; time series obtained from: **(a)** Reference spirometer **(b)** SVR (RR_{err} = 4.80 BPM, E_ρ = -0.041) **(c)** GMR (RR_{err} = 4.91 BPM, E_ρ = 0.646) **(d)** DCT based estimation, (RR_{err} = 7.81 BPM, E_ρ = -0.081) and **(e)** SEC (RR_{err} = 3.24 BPM, E_ρ = 0.291). Subject performing physical activity between 100 to 180 sec.

marginal distribution, $p_k(\mathbf{x})$ [48]:

$$p_k(y|\mathbf{x}) = \frac{p_k(\mathbf{x}, y)}{p_k(\mathbf{x})} = \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}_{ky|\mathbf{x}}, \boldsymbol{\Lambda}_{kyy}^{-1}\right) \tag{2.13}$$

where, $\boldsymbol{\Lambda}_{kyy}$ is a submatrix of the matrix $\boldsymbol{\Lambda}_k = \boldsymbol{\Sigma}_k^{-1}$ given by:

$$\boldsymbol{\Lambda}_k = \begin{pmatrix} \boldsymbol{\Lambda}_{k\mathbf{xx}} & \boldsymbol{\Lambda}_{k\mathbf{x}y} \\ \boldsymbol{\Lambda}_{ky\mathbf{x}} & \boldsymbol{\Lambda}_{kyy} \end{pmatrix} \tag{2.14}$$

The conditional mean, $\boldsymbol{\mu}_{ky|\mathbf{x}}$, is given by:

$$\boldsymbol{\mu}_{ky|\mathbf{x}} = \boldsymbol{\mu}_{ky} - \boldsymbol{\Lambda}_{kyy}^{-1}\boldsymbol{\Lambda}_{ky\mathbf{x}}\left(\mathbf{x} - \boldsymbol{\mu}_{k\mathbf{x}}\right) \tag{2.15}$$

During the test phase the spirometer output is predicted from the feature vectors using the Gaussian mixture distribution learnt during the training phase. More specifically; given a

test vector $\mathbf{x}$ the estimate $\hat{y}$ of the spirometer is equal to the expectation of the conditional distribution $p(y|\mathbf{x})$:

$$\hat{y} = \mathbb{E}\left[p(y|\mathbf{x})\right] \tag{2.16}$$

The spirometer test signals estimated using Gaussian Mixture Regression (GMR) are plotted in Figure 2.5 (c) - 2.8 (c). It seems that GMR, as compared to SVR, gives a better estimate of the envelope as indicated by both the shape of the GMR estimates and the higher values of E_ρ. However, in terms of breathing rate estimation it seems that GMR also degrades in high breathing rate and motion artifact regions. Therefore, overall it seems that GMR does result in performance improvement over SVR, especially at normal respiration rates. However, its performance at higher respiration rates is not satisfactory and there is still margin for improvement.

2.3.3 DCT Based Estimation

In general, it is difficult to optimize machine learning classifiers to give optimal performance in regression applications that have significant temporal correlation between neighboring samples. This is because the underlying theory, more often than not, assumes that the data points are independent and identically distributed. A potential solution to this problem can be to consider techniques such as *Markov Models* that explicitly cater for the temporal dependence or employ a feature extraction front-end that outputs feature vectors that effectively capture temporal dependence. The addition of delta features in SVR and GMR did ameliorate the situation to some extent however, there is still room for improvement. This subsection, presents a very simple time-frequency technique that operates on temporal frames of finite length and estimates the spirometer signal on a frame-by-frame basis, in contrast to the sample-by-sample estimation approach employed by SVR and GMR. The output signal from each electrode is split into non-overlapping frames of length M temporal samples. Different frame lengths were experimented with; $M = 200$ samples was found to give the best performance. Assuming that maximum number of frames in a given test signal

equals F then for the i-th frame the $(M \times N)$ matrix $\mathbf{X}_i$, whose columns contain the time samples from the $N(=10)$ electrodes, the DCT coefficient matrix $\mathbf{C}_i$ is obtained by

$$\mathbf{C}_i = \mathbf{T}\mathbf{X}_i \quad \text{for } i = [1, ..., F] \tag{2.17}$$

where, $\mathbf{T}$ represents the $(M \times M)$ DCT basis. In the next step the quantized coefficient matrix $\widetilde{\mathbf{C}}_i$ is obtained by setting all, but the $P(< M)$ largest magnitude coefficients in every column of $\mathbf{C}_i$ to zero. The mean coefficient vector $\widetilde{\mathbf{x}}_i$ is then obtained by taking the average over all the electrode signals:

$$\widetilde{\mathbf{x}}_i = \frac{1}{N}\sum_{n=1}^{N} \tilde{\mathbf{c}}_i^n \quad \text{for } i = [1, ..., F] \tag{2.18}$$

where, $\tilde{\mathbf{c}}_i^n$ represents the n-th column vector of $\widetilde{\mathbf{C}}_i$. Finally, the estimated spirometer time samples, for frame-i are obtained via the inverse DCT transform:

$$\hat{\mathbf{y}}_i = \mathbf{T}^{-1}\widetilde{\mathbf{x}}_i \quad \text{for } i = [1, ..., F] \tag{2.19}$$

The reconstructed test signals are plotted in Figure 2.5 (d) - 2.8 (d) and demonstrate that SVR and GMR are better in terms of envelope estimation than DCT based reconstruction. In terms of breathing rate estimation however; the time series plots of the estimated signals indicate that the (much simpler) DCT based approach gives a performance similar to that of SVR and GMR. This observation, although seemingly minor, has significant implications. It demonstrates that it is possible to estimate the respiration rate by a simple procedure based on identifying the dominant frequency components in the electrode signals (without requiring any training data). Although the problems with envelope estimation are still unresolved, the DCT based approach paves the way for the next approach (the SEC, presented in section 2.4) in which envelope and respiration rate estimation are segregated and treated as separate problems. Breathing rate estimation in the SEC is a more refined version of the DCT based estimation whereas, envelope estimation employs regression models similar to the ones presents in sections 2.3.1 and 2.3.2 with the difference that it estimates only the envelope and is not responsible for breathing rate estimation.

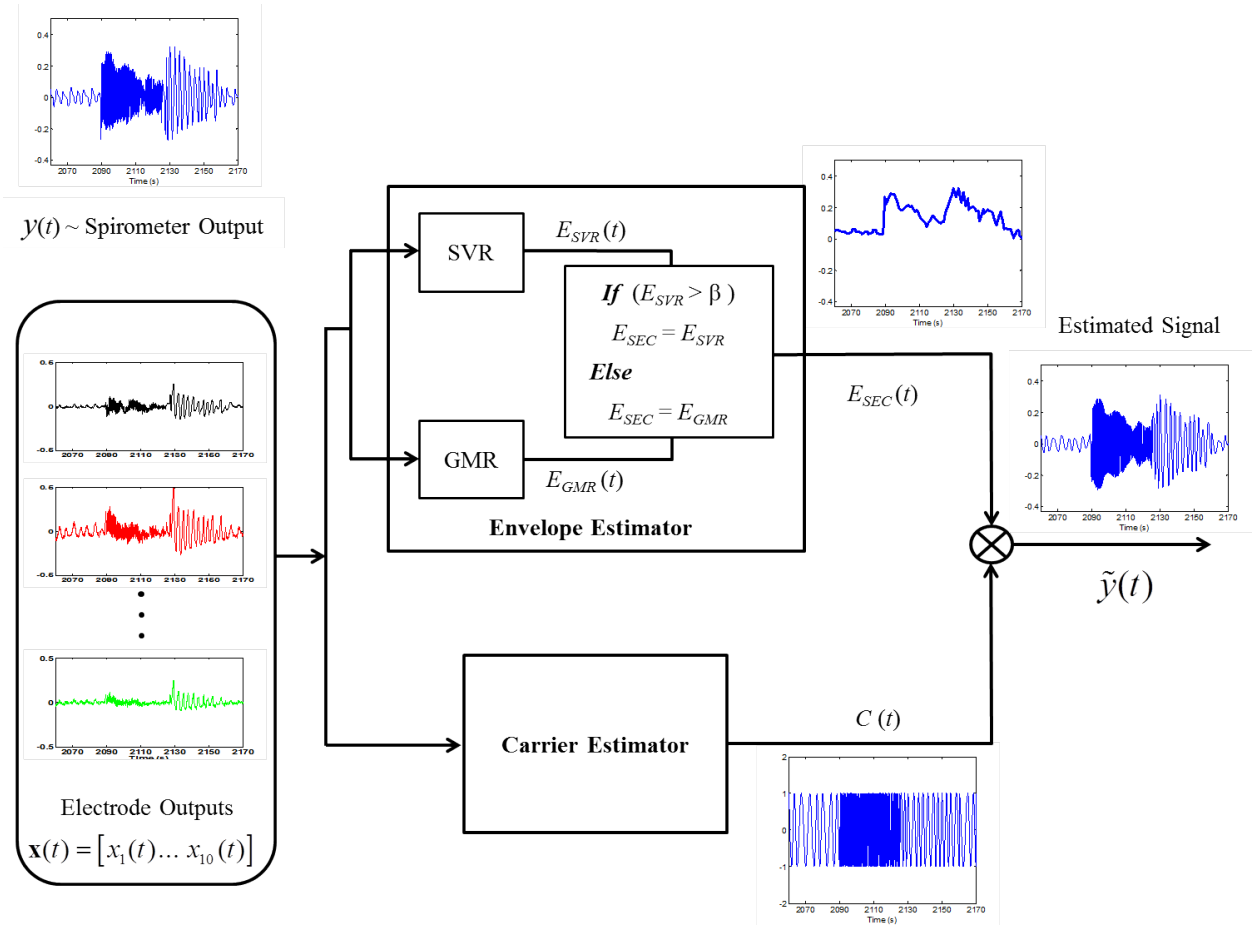


Figure 2.9: Block diagram of SEC estimation using the AM approximation.

2.4 SEC Estimation Using the AM Aproximation

This section introduces a novel respiratory signal estimation approach titled **S**egregated **E**nvelope and **C**arrier (SEC) Estimation. SEC estimation is based on the observation that the spirometer can signal be approximated by an Amplitude Modulated (AM) signal. AM is one of the oldest and simplest modulation techniques and is obtained by multiplying a (high-frequency) carrier with a (lower-frequency) information signal. The result is a signal centered around the carrier frequency whose amplitude/envelope varies in proportion to the information signal. Therefore, the SEC views the spirometer output as an AM signal, albeit with a time varying carrier component.

The primary advantage of the AM approximation is that it enables separate evaluation

of the envelope and respiration rate thus preventing errors in estimation of one parameter from effecting the other. The block diagram of a framework based on the AM approximation is shown in Figure 2.9, the *Envelope Estimator* and the *Carrier Estimator* blocks provide predictions of the signal's amplitude and respiration rate. The outputs of both blocks are then multiplied to obtain the the estimated spirometer signal. This approach may also be motivated physiologically in the sense that physical activity by the subject generally introduces large amplitude distortions in the electrode signals while the frequency information (in some, if not all) of the electrode signals still remains intact. In such a scenario we may benefit by segregating the estimation of the signal's rate and envelope. Instead of the sample by sample approach employed in the SVR and GMR, estimation of the signal envelope over a relatively longer temporal window should enable mitigation of the large amplitude variations observed in the motion artifact regions of Test Signals-1, 2 and 4. The envelope and carrier estimation blocks are described in detail below.

2.4.1 Envelope Estimation

The envelope estimation block employs regression techniques similar to those used in subsections 2.3.1 and 2.3.2 with the exception that training is performed using the envelope of the spirometer signal instead of it's actual sample values. The input features remain that same as those employed in the preceding sections. By focusing solely on the envelope (and not the carrier) it is hoped that the classifier maybe be able to learn the temporal variations in the signal amplitude with ease. Results indicate that this is indeed the case.

During the training phase the envelope is obtained by first locating the local-maxima in the spirometer signal and then interpolating to produce a signal that is the same length as the original spirometer output. SVM and GMM models are then trained to predict the envelope using the electrode signals. Analysis of the results of SVR and GMR based envelope estimation indicate that for the majority of Test Data, GMR performs better at lower amplitude values of envelope (commonly found in normal breathing regions) whereas

the SVR is better higher amplitude values of the envelope (commonly found in accelerated breathing regions). This is consistent with the trend in sections 2.3.1 and 2.3.2 where it was observed that the GMR based approach performed better at lower amplitude, lower frequency regions. The reason why SVR based envelope estimation outperforms GMR based envelope estimation (at higher envelope amplitudes) may be attributed to the fact that in general, SVMs are better suited for scenarios where the number of training instances are small, which is somewhat true here since the bulk of the data consists of subjects breathing normally. Therefore, given a larger amount of data it is possible that the performance obtained via GMR may be improved however, collecting so much data is often difficult. The final, SEC, envelope is obtained by combining the envelopes output by the SVR and GMR based envelope estimators. To elaborate, the GMR envelope was employed at lower amplitudes and the SVR envelope was employed at higher amplitude values. The threshold to switch between low and high amplitudes was determined by examining large instances of the data. It is highlighted that the same threshold was employed for all test signals and it was not changed for different test signals. Results for SEC estimation (Figures 2.5 (e) - 2.8 (e)) indicate that envelope estimation at higher respiration rates is significantly better than all previous methods. This can be attributed to employing a classifier dedicated for only envelope estimation.

2.4.2 Carrier Estimation

As highlighted earlier, the results in section 2.3.3 indicate that respiration rate information can be extracted using a simple procedure. The carrier extraction process is a modified version of the DCT based technique introduced in section 2.3.3. The primary difference being that only a single DCT coefficient is retained and assigned a value of 1. However, this approach can cause problems during time intervals where electrode signals contain noise. Consider for example the reference spirometer signal in Figure 2.10-(a) where the subject is holding his breath during some intervals. Picking the strongest frequency component

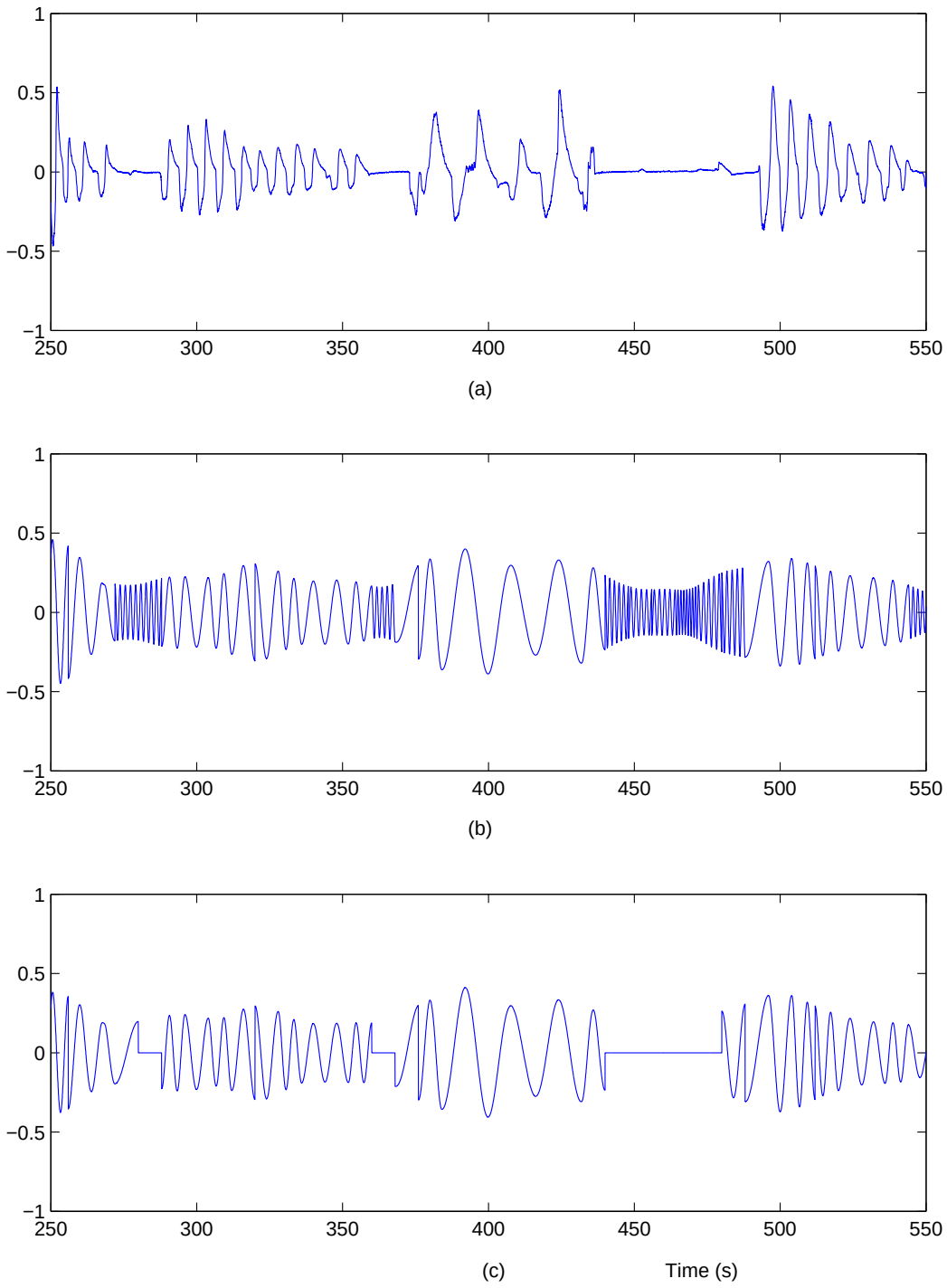


Figure 2.10: (a) Original Spirometer signal. (b) Signal estimate using the largest magnitude DCT coefficient (c) Signal estimate via noisy frame suppression.

in the electrodes results electrode noise frequency being selected in some noisy frames as demonstrated in Figure 2.10-(b). This can be remedied by exploiting the observation that the majority of noisy electrode frames have low energy compared to the true frames containing the respiratory signal. Thus we can differentiate between noisy and true respiratory frames by comparing the energy of each test frame with the average frame energy, ξ, computed from the training data. A test frame with energy significantly below ξ is assumed to be a zero frame and all DCT coefficients are set to zero. Formally, we first compute the $(M \times 1)$ vector $\mathbf{c}_i^{Tst}$ by adding adding the absolute values of the coefficients of the N electrodes

$$\mathbf{c}_i^{Tst} = \frac{1}{N} \sum_{n=1}^{N} |\tilde{\mathbf{c}}_i^n| \quad i = [1, ..., F_1] \tag{2.20}$$

Here, F_1 is the total number of frames in the test signal. The average training frame energy ξ is computed as below:

$$\xi = \frac{1}{F_2} \sum_{i=1}^{F_2} \|\mathbf{c}_i^{Trn}\| \tag{2.21}$$

where, $\mathbf{c}_i^{Trn}$ represents the mean coefficient vector of the i-th training data frame, F_2 is the total number of frames in the training data, and $\|\ \|$ represents the l_2-norm. Now, if

$$\|\mathbf{c}_i^{Tst}\| > \eta\xi \tag{2.22}$$

the largest magnitude element of $\mathbf{c}_i^{Tst}$ is assigned a magnitude of 1 (sign remains unchanged); the remaining elements are set to zero. If on the other hand condition (2.22) is not satisfied, then all coefficients are set to zero. The parameter $\eta = 0.15$ and its value was determined heuristically from the data. Figure 2.10-(c) demonstrates the spirometer estimate obtained via the noisy frame suppression procedure outlined above.

Results for Test Signals-1 to 4 are displayed in Figures 2.5 (e) - 2.8 (e). It can be observed that the performance improvement is significant. For Test Signal-1 SEC outperforms all other techniques by a large margin both terms of envelope and breathing rate estimation. The envelope correlation coefficient for this signal is almost equal to 1, $\text{RR}_{err} = 2.80BPM$. Additionally, SEC seems to be more robust to the effect of motion artifacts as is apparent

Subject	Low Breathing Rates				High Breathing Rates			
	SVR	GMR	DCT	SEC	SVR	GMR	DCT	SEC
1	3.392	3.022	3.331	**1.789**	**10.045**	16.574	14.900	10.212
2	8.265	2.529	3.269	**1.11**	35.877	25.06	31.731	**2.163**
3	17.33	2.131	2.699	**0.781**	14.531	**0.469**	15.469	0.938
4	11.435	5.185	6.747	**2.983**	22.461	21.289	20.313	**2.930**
5	3.196	2.983	4.19	**1.918**	10.637	13.882	9.195	**1.082**
6	8.594	6.605	4.759	**2.557**	17.578	**0.586**	20.733	2.524
7	35.938	7.244	4.972	**2.344**	33.545	12.744	9.229	**7.178**
8	4.688	3.764	4.230	**1.847**	27.604	20.182	30.599	**18.359**
9	**3.480**	9.375	3.835	4.261	11.953	9.258	4.922	**4.805**
10	3.385	**1.910**	4.514	2.17	6.563	12.891	7.031	**3.984**
11	5.122	2.951	3.733	**2.431**	11.484	3.281	10.078	**2.344**
Mean:	10.135	4.579	4.205	**2.199**	18.389	12.383	15.930	**5.359**

Table 2.1: Average Respiration Rate Error (RR_{err}) for 11 different human subjects.

from its estimate between 0 to 100 sec. Both SVR and GMR suffer from degradation in this region. Similar trends are observed in Test Signal-2 as well. For Test Signal-3 the $RR_{err} = 3.03BPM$ for SEC is slightly higher than that achieved by using SVR however; SEC is still significantly better in terms of envelope estimation. For Test Signal-4 the envelope correlation coefficient value is not the best amongst the four techniques mentioned in this work. However, this metric is not perfect and visually it seems that SEC estimation gives an acceptable performance. In terms of RR_{err} still gives the best performance and again seems to be more tolerant of motion artifacts than the other three approaches.

2.5 Results

Due to space constraints it is not possible to plot all the Test signals and the performance obtained by all methods for 11 different human subjects are summarized in Tables 2.1 and 2.2. As mentioned at the start of section 2.3, each patient's data was divided into 9 non-overlapping subsets. At one time a single subset was used as the test signal and the rest were used for training. In this fashion each subset was used as a test signal and compared

Subject	Low Breathing Rates				High Breathing Rates			
	SVR	GMR	DCT	SEC	SVR	GMR	DCT	SEC
1	0.182	**0.397**	0.367	0.281	0.036	0.286	0.343	**0.557**
2	0.261	0.368	0.209	**0.426**	0.062	0.501	0.091	**0.627**
3	0.122	0.210	**0.436**	0.359	0.053	0.395	0.258	**0.631**
4	0.431	0.506	0.421	**0.615**	0.592	0.856	0.441	**0.897**
5	**0.445**	0.353	0.438	0.369	0.565	0.545	0.257	**0.876**
6	**0.495**	0.450	0.441	0.399	0.572	**0.953**	0.138	0.918
7	0.420	**0.602**	0.555	0.548	0.763	0.757	0.512	**0.850**
8	0.505	0.506	**0.570**	0.534	0.165	0.472	0.162	**0.603**
9	0.183	**0.221**	0.131	0.164	0.141	0.195	**0.318**	0.082
10	0.534	0.346	**0.645**	0.458	0.123	0.187	0.311	**0.454**
11	0.492	0.507	**0.615**	0.592	0.174	0.592	0.150	**0.701**
Mean	0.370	0.406	**0.439**	0.431	0.295	0.521	0.271	**0.654**

Table 2.2: Envelope correlation Coefficient (E_ρ) for 11 different human subjects.

with the spirometer reference signal. The respiration rate and envelope performance metrics were computed; after this the subset was replaced in the training set and the next subset was selected and training and testing phases were repeated. Furthermore, we subdivided each subject's Test Signals in to two groups depending on whether they contained Low or High respiration rates. For example all signals such as Test Signal-1 were labeled as Low respiration rate signals whereas, signals such as Test Signal-2, 3 and 4 were labeled as High respiration rate signals. Each metric listed in Table 2.1 and Table 2.2 was computed by averaging over all the Low or High breathing rate test signals for that particular human subject. For example, the average respiration rate error for subject-1 using the SEC is 1.789 BPM for low breathing rate test signals and 10.212 BPM for high breathing rate test signals.

It can be observed from Table 2.1 that in terms of respiration rate, the AM based SEC approach gives the best performance for majority of subjects. Most practical applications require that the respiration rate error (RR_{err}) must be less than 10 BPM at all times. We employ a stricter threshold of 5 BPM here. At low breathing rates all approaches, except SVR, give low errors. However, overall SEC gives a more consistent performance and its

RR_{err} is never greater than 5 BPM. Whereas, RR_{err} for GMR exceeds 5 BPM for 4 out of the 11 subjects. DCT and SVR estimation exceed the 5 BPM threshold for 1 and 6 subjects respectively. At high respiration rates SEC gives the best performance on average and its RR_{err} exceeds the 5 BPM for 3 subjects. SVR and GMR exceed the threshold for 11 and 8 subjects respectively. Whereas the RR_{err} for DCT based estimation exceeds 5 BPM for all but one of the 11 subjects. Overall, the values in Table 2.1 demonstrate a trend similar to that observed for Test Signals-1 to 4 i.e., GMR gives reasonable performance at low respiration rates however, at high respiration rates it gives a reasonable performance in a few cases but completely misses the mark in the majority of cases. SEC, in contrast, delivers a much more consistent performance.

Envelope correlation coefficients (E_ρ) are listed in Table 2.2. At high breathing rates, SEC gives the best performance for all subjects except subjects 6 and 9. For human subject-6 GMR performs slightly better however, SEC also results in a very high correlation coefficient (0.918). For human subject-9, the envelope estimation performance for all 4 techniques is subpar; this may be due to extraordinarily high levels of noise or interference during data collection. At low breathing rates there is not a significant difference between the performance of all the four techniques in terms of envelope estimation. On average DCT gives the best performance followed by the SEC. However, it is highlighted that at low breathing rates the signal envelope does not exhibit significant variations in the shape of the envelope. The majority of signals at low respiration rates are similar to Test Signal-1 and therefore, have an almost flat envelope with an amplitude close to the subject's average lung volume. In these scenarios a few false peaks in the estimated envelope may result in significant variations in E_ρ. Consider for example TS-1, although the correlation coefficient of the SEC estimate of this signal is lower than that of the GMR estimate it can be observed visually that there is not a significant difference between the two envelopes. At high breathing rates however, the reference signals exhibits significant variations in lung volume shape and correlation coefficient gains much more importance in these regions. The results demonstrate

that SEC delivers significant performance improvements over all other approaches.

CHAPTER 3

BREATHING RATE ESTIMATION USING KERNEL METHODS

The previous chapter discussed the SEC which is a framework for estimation both the breathing rate and lung volume. The technique for estimation of the breathing rate in the SEC is quite simple technique for breathing rate estimation. Although it gives reasonable performance there still is room for improvement. In this chapter the emphasis is shifted primarily to the estimation of beathing rate esimtation alone. There are two primary reasons for this: (1) Breathing rate is considered to be a much more important in clinical practice than lung volume; and (2) Improvement in rate estimation is bound to benefit lung volume estimation as well since accurate estimation of breathing rate is critical for lung volume estimation using the SEC. To elaborate further, the breathing rate estimation approach proposed in this chapter can be used to replace the rate (carrier) estimation technique employed in section 2.4.2. In this chapter kernel machines are employed to for robust breathing rate estimation. The best performing technique employs an innovative set of features constructed from the discrete wavelet transform to differentiate between various respiratory states which enable the learning algorithm adapt based on the underlying state (such as Apnea, fast breathing or normal breathing etc).

A detailed analysis of the results obtained from the, DCT filter based, breathing rate estimation employed in the previous chapter reveals that (other than artifact regions) the most challenging respiratory states for rate estimation using impedance electrodes are hyperventilation and apnea especially when electrode outputs are noisy. Two example cases are presented in Figures 3.1 and 3.2. Detection of breathing rate during hyperventilation is challenging primarily because the subject is taking very shallow breaths while breathing at a very rapid rate. Therefore, the changes in the lung volume are quite small and can missed easily especially when the electrode signal contains noise. Consider for example, the

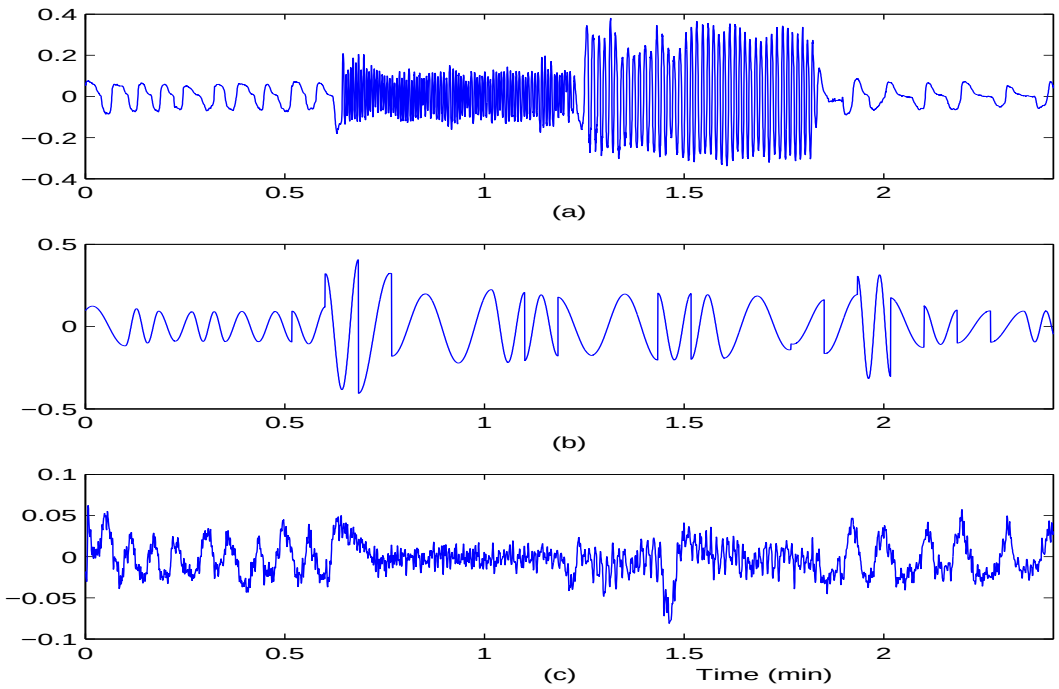


Figure 3.1: Breathing rate estimation during *hyperventilation* under high noisy conditions. (a) Reference Spirometer. (b) SEC output. (c) Electrode-2 output.

respiratory signal shown in Figure 3.1 (a); the subject starts breathing at an accelerated rate beyond the 30 second mark. Examination of only the spirometer signal seems to indicate that there shouldn't be much difficulty in estimating breathing rates at any rate because the amplitude of the spirometer signal during accelerated breathing is comparable to its amplitude under normal breathing conditions. However, the spirometer measures *flow*; our task is to estimate the breathing rate from the electrodes, which measure the lung volume directly which does not change significantly during the shallow breathing hyperventilation conditions. Therefore, when noise power is high it becomes very difficult to differentiate between noise and genuine breath signals. This can be observed from the plot of electrode-2 shown in Figure 3.1 (c), it can be seen that amplitude of the electrode output in the hyperventilation region (between 0.5 min to 2 min) is quite low; as a result a spectral energy based rate estimation algorithm may under or over-estimate the breathing rate as shown in

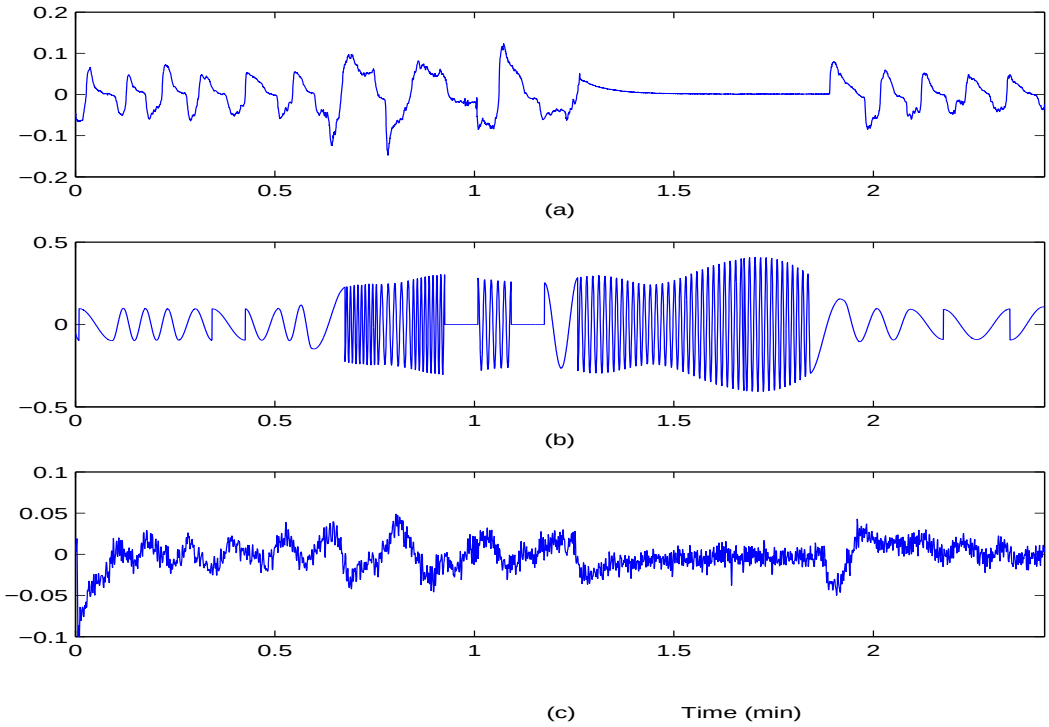


Figure 3.2: Breathing rate estimation during *apnea* under high noisy conditions. (a) Reference Spirometer. (b) SEC output. (c) Electrode-2 output.

Figure 3.1 (b). Similarly, a high level of noise may also cause falsely treating an apnea region as a hyperventilation region as shown in Figure 3.2 (c).

Although the SEC employs noisy-frame suppression (described in section2.4.2) to mitigate the impact of noise it is dependent on the accurate estimation of the envelope and therefore, may not work in case envelope estimation is not accurate. Therefore, this chapter presents a technique which employs kernel machines for accurate estimation of breathing rate; results demonstrate improvements in performance.

3.1 *Gini* Kernel Machines for Breathing Rate Estimation

As discussed above the breathing rate estimation algorithm employed by the SEC degrade in the presence of motion artifacts and high noise power. Training kernel machines on a dataset

that contains example scenarios containing artifacts and noise may enable better estimation of the breathing rate. The training dataset contains a number of representative noise and artifact scenarios and also contains multiple electrode channels therefore, it is anticipated that the kernel machine should be able to learn the mapping to actual breathing signal even when some of the electrode channels are corrupted. The subsections that follow describe in detail the optimization techniques employed for training *Gini* kernel machines and how they can be used for breathing rate estimation. This optimization framework was presented in [51]; it is described here for convenience and is modified according to the demands of the application where necessary.

3.1.1 Supervised Learning Using *Gini*-Kernel Machines

In a general supervised learning framework the learner is trained using a set of feature vectors $\mathcal{T} \subset \mathcal{X} : \mathcal{T} = \mathbf{x}_i, i = 1, .., N$ independently drawn from a fixed distribution $P(\mathbf{x})$, with $\mathbf{x} \in \mathcal{X}$. Furthermore, the learner is also provided with a set of soft (or hard) labels $y_{ik} = P(C_k|\mathbf{x}_i)$ that represent the conditional probability measures representing the probability of observing class-k given feature vector $\mathbf{x}_i$. The set of classes is discrete ($k \in [1, \ldots, M]$) and the labels therefore, are normalized to satisfy the condition $\sum_{k=1}^{M} y_{ik} = 1$. For breathing rate estimation the number of classes $M = 2$; with class-1 representing *inspiration* and class-2 representing *expiration*. The soft labels for each class are derived using the logistic transformation as described in section 3.1.2. The task of the learner is to search for useful patterns in the training data and use them to select a finite set of regression functions $\tilde{P} = \{\tilde{P}_k(\mathbf{x})\}, k = 1, .., M$ that are accurate estimates of the true conditional probabilities $P(C_k|\mathbf{x})$. The learner accomplishes this by incorporating prior knowledge of the topology of the feature space using a distance metric $D_Q : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$. In addition to D_Q, the learner also employs an agnostic (or non-informative) distance metric $D_I : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$ which assumes no knowledge of the training set. Use of the agnostic prior enables enforcement of the smoothness constraints on the regression function $\tilde{P}_k(\mathbf{x})$. Smoothing is consistent

with the principles of maximum entropy [52] and is required because the prior labels y_{ik} are based only on the training data; use of D_Q alone results in an over-fitted solution that does not generalize well to unseen data. Therefore, the training procedure for estimation of the probability functions $\tilde{P} = \{\tilde{P}_k(\mathbf{x})\}$ is generally performed by minimization of a joint distance metric as below

$$\min_{\tilde{P}} G(\tilde{P}) = \min_{\tilde{P}} [D_Q(Y, \tilde{P}) + \gamma D_I(\tilde{P}, U)]. \tag{3.1}$$

Where $Y : \mathbb{R}^N \times \mathbb{R}^M$ is a matrix of prior labels $y_{ik} = P(C_k|\mathbf{x}_i)$, with $i \in [1, .., N]$ and $k \in [1, .., M]$. U denotes a uniform distribution given by $U_k(\mathbf{x}) = 1/M, \ \forall k = 1, .., M$. $\gamma > 0$ is a hyper-parameter that controls the trade-off between the prior (D_Q) and agnostic (D_I) distance metrics. The solution obtained by minimizing the cost function 3.1 is close to both prior distribution with respect to the distance metric $D_Q(.,.)$ and the agnostic (non-informative) uniform distribution U. The maximum entropy framework [52] also permits the imposition of linear constraints on the optimization problem 3.1. These constraints should be in terms of cumulative statistics defined on the training set. The first linear constraint imposes the equality condition between the frequencies of occurrence of a class $k = 1, .., M$ under the distribution $\tilde{P}$ to an equivalent measure under the prior distribution y_{ik}. This first constraint expresses equivalence between average estimated probabilities and empirical frequencies for each class over the training set

$$\sum_{i=1}^{N} \tilde{P}_k(\mathbf{x}_i) = \sum_{i=1}^{N} y_{ik}, \quad k = 1, \dots M \tag{3.2}$$

The underlying assumption here is that all features $\mathbf{x} \in \mathcal{X}$ are equally likely. The normalization and boundary conditions for valid probability distributions are expressed using a second set of linear constraints

$$\tilde{P}_k(\mathbf{x}) \geq 0, \quad k = 1, \dots M, \tag{3.3}$$

$$\sum_{k=1}^{M} \tilde{P}_k(\mathbf{x}_i) = 1 \tag{3.4}$$

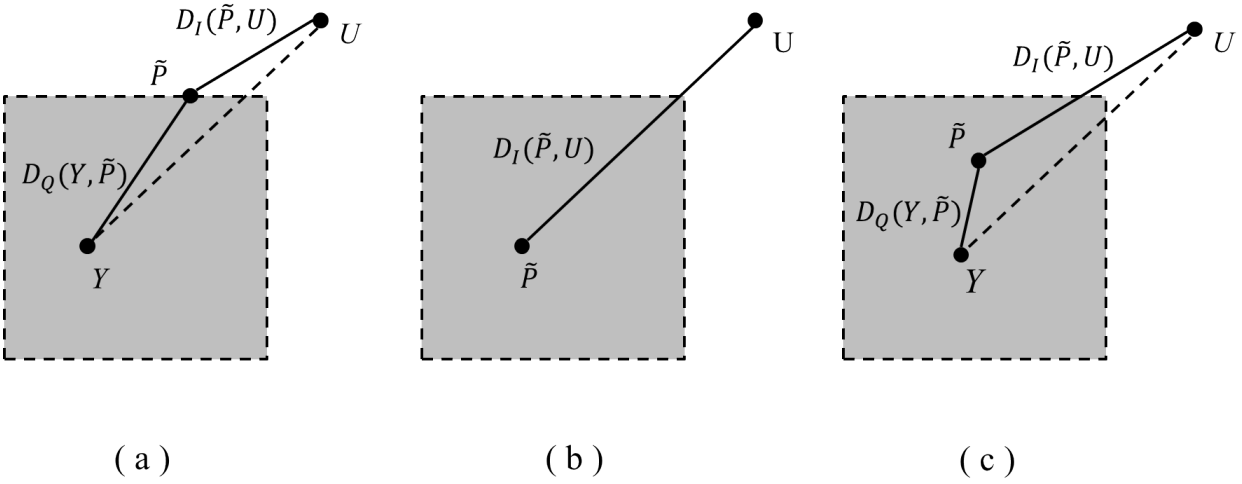


Figure 3.3: Maximum entropy regression for supervised learning; the square region represents the constraint space. (a)$\gamma \to \infty$: Solution is the projection of U onto the constraint space. (b)$\gamma = 0$: Solution $\tilde{P}$ is equal to Y. (c) Non-extreme values of γ: Solution $\tilde{P}$ lies at a location within the constraint space that minimizes the total distance to the prior distribution Y and the agnostic distribution U.

where the normalizing equality constraint subsumes the additional inequality constraint $\tilde{P}_k(\mathbf{x}) \leq 1,\ k = 1, \ldots M$.

An illustration of the solution to the optimization problem in equation (3.1) is shown in Figure 3.3. For the purpose of illustration the linear constraints (5.8), (3.3) and (3.4) are represented by the shaded square region. As a result any solutions to (3.1) must lie within or at the boundary of the constraint space. The proximity of the, learned, distribution $\tilde{P}$ to the prior empirical distribution Y is determined by the distance $D_Q(Y, \tilde{P})$. The distance $D_I(\tilde{P}, U)$ that defines an agnostic model which assumes zero prior knowledge. This framework is similar to the maximum entropy approach [52, 53]. Note that the prior distribution Y lies within the constraint space whereas the agnostic U distribution will lie outside the constraint space under non-degenerate conditions. The location of the solution $\tilde{P}$ with respect to the prior Y and agnostic U distributions is influenced by the value of the hyper-parameter $\gamma > 0$. This parameter also determines the generalization performance and sparsity of classifiers defined by $\tilde{P}$. As demonstrated in Figure 3.3, for $\gamma = 0$, the solution

overlaps with the prior distribution Y resulting in over-fitting of the training set. When $\gamma \to \infty$, the maximum entropy entropy solution is achieved which is the projection of the agnostic distribution U on the constraint space.

After application of first order Karush-Kuhn-Tucker (KKT) conditions [54] the optimization problem in (3.1) along with the constraints (5.8) to (3.3) can be represented by a Lagrangian function L

$$L(G, b_k, \beta_k, z) = G(\tilde{P}) - b_k \sum_{i=1}^{N} \left(\tilde{P}_k - y_{ik} \right) - \beta_k \tilde{P}_k(\mathbf{x}) - z \left(1 - \sum_{k=1}^{M} \tilde{P}_k(\mathbf{x}) \right) \tag{3.5}$$

Here b_k and β_k represent Lagrange multipliers corresponding to frequency constraints (5.8) and the inequality constraints (3.3) respectively. $z(\mathbf{x})$ corresponds to Lagrange multiplier for the normalization constraint (3.4). Minimization with respect to the probability function $\tilde{P} = \{\tilde{P}_k(\mathbf{x})\}$ can be achieved by taking the gradient and setting the result to zero.

$$\gamma \frac{\partial D_I(\tilde{P}, U)}{\partial \tilde{P}_k(\mathbf{x})} = -\frac{\partial D_Q(Y, \tilde{P})}{\partial \tilde{P}_k(\mathbf{x})} + b_k - z(\mathbf{x}) + \beta_k(\mathbf{x}). \tag{3.6}$$

For simplicity purposes it is assumed that $D_I(P, U)$ has a form that can be decomposed into independent and identically distributed (i.i.d.) components. Furthermore, a quadratic formulation is employed as a distance metric. This leads to the following form for $D_I(\tilde{P}, U)$

$$D_I(\tilde{P}, U) = \sum_{k=1}^{M} \sum_{i=1}^{N} \frac{1}{2} \left(\tilde{P}_k(\mathbf{x}_i) - U_{ik} \right)^2 \tag{3.7}$$

The agnostic distribution $U_{ik} (\equiv 1/M)$ is uniform, and upon substitution yields the following form

$$D_I(\tilde{P}, U) = \frac{1}{2} \sum_{k=1}^{M} \sum_{i=1}^{N} \tilde{P}_k(\mathbf{x}_i)^2 - \frac{N}{2M} \tag{3.8}$$

The Lagrange function in (3.5) can now be rearranged to give the class-conditional probability for any vector $\mathbf{x}$

$$\tilde{P}_k(\mathbf{x}) = \frac{1}{\gamma} \left[-\frac{\partial D_Q(Y, \tilde{P})}{\partial \tilde{P}_k(\mathbf{x})} + b_k - z(\mathbf{x}) + \beta_k(\mathbf{x}) \right] \tag{3.9}$$

There are a number of choices for $D_Q(.,.)$, the prior distance metric, amongst the most popular is the quadratic distance which employed widely in kernel methods [55] and in

Bayesian methods (as covariance functions) [56]. For two distributions $\hat{P} = \{\hat{P}_k(\mathbf{x})\}$ and $\tilde{P} = \{\tilde{P}_k(\mathbf{x})\}$ the quadratic distance is given by

$$D_Q(\hat{P}, \tilde{P}) = \frac{C}{2} \sum_{k=1}^{M} \sum_{\mathbf{x},\mathbf{v} \in \mathcal{T}} K(\mathbf{x}, \mathbf{v}) \left[\hat{P}_k(\mathbf{x}) - \tilde{P}_k(\mathbf{x})\right] \left[\hat{P}_k(\mathbf{v}) - \tilde{P}_k(\mathbf{v})\right]. \tag{3.10}$$

Where $K : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ represents a symmetric, positive definite kernel that satisfies Mercer's criterion,[1]. Some popular kernels employed in machine learning applications include the Gaussian radial basis function and polynomial splines [55, 57]. The kernel $K(\mathbf{x}, \mathbf{v})$ quantifies the topology of the metric space for the points $\mathbf{x}, \mathbf{v} \in \mathcal{X}$ and therefore, embeds prior knowledge into the distance $D_Q(.,.)$. The gradient of the quadratic distance $D_Q(.,.)$ of (5.12) with respect to $\tilde{P}_k(\mathbf{x})$ is given by

$$\frac{\partial D_Q(Y, \tilde{P})}{\partial \tilde{P}_k(\mathbf{x})} = -\frac{C}{2} \sum_{\mathbf{v} \in \mathcal{T}} K(\mathbf{x}, \mathbf{v}) \left[\hat{P}_k(\mathbf{v}) - \tilde{P}_k(\mathbf{v})\right] \tag{3.11}$$

To avoid complexity in notation, $\mathbf{v}$ in equation (3.11) is rewritten as an indexed vector $\mathbf{x}_i$ to give:

$$\frac{\partial D_Q(Y, \tilde{P})}{\partial \tilde{P}_k(\mathbf{x})} = -\frac{C}{2} \sum_{i=1}^{N} K(\mathbf{x}, \mathbf{x}_i) \left[\hat{P}_k(\mathbf{x}_i) - \tilde{P}_k(\mathbf{x}_i)\right] \tag{3.12}$$

The first order conditions (3.9) for the quadratic form $D_Q(.,.)$ in equation (5.12) can now be rewritten as

$$\tilde{P}_k(\mathbf{x}) = \frac{C}{2\gamma}[f_k(\mathbf{x}) - z(\mathbf{x}) + \beta_k(\mathbf{x})] \tag{3.13}$$

where

$$f_k(\mathbf{x}) = \sum_{i=1}^{N} \lambda_k^i K(\mathbf{x}_i, \mathbf{x}) + b_k$$

with inference parameters

$$\lambda_k^i = C[y_{ik} - \tilde{P}_k(\mathbf{x}_i)].$$

The Lagrange parameter function $\beta_k(\mathbf{x})$ in Equation (3.13) needs to ensure that the probability scores $P_k(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}$ according to (3.3), and the Lagrange parameter function $z(\mathbf{x})$ needs to ensure normalized probabilities $\sum_{k=1}^{M} P_k(\mathbf{x}) = 1$ according to (3.4).

[1] $K(\mathbf{x}, \mathbf{v}) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{v})$. There is no need to explicitly compute the map $\Phi(\cdot)$ since it only appears in inner-product form.

The procedure for obtaining the set of inference parameters $\Lambda = \{\lambda_k^i\}, i = 1, .., N, \quad k = 1, .., M$ entails solving (3.1) over the set of training data $\mathcal{T}$. Expressing the quadratic distance $D_Q(Y, \tilde{P})$ in equation (5.12) in terms of the inference parameters λ_k^i and substituting back in the cost function (3.1) along with the agnostic distance $D_I(P, U)$ (3.8) leads to a dual formulation for the *Gini*SVM cost function

$$H_g = \sum_{k=1}^{M} \left[\frac{1}{2C} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_k^i Q_{ij} \lambda_k^j + \frac{\gamma}{2} \sum_{i=1}^{N} (y_{ik} - \lambda_k^i / C)^2 \right] \tag{3.14}$$

where $Q_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ denote elements of the $k$ernel matrix $\mathbf{Q}$. The constant term ($= -N/2M$) in the agnostic distance $D_I(.,.)$ of equation(3.8) has been discarded since it has no effect on the minimization. As is the case for the primal (3.1), minimization of the dual H_d should also be performed while ensuring that the linear constraints (5.8)-(3.4) are satisfied when search for solutions. The constraints rewritten in terms of the inference parameters are as below

$$\begin{aligned} \sum_{k=1}^{M} \lambda_k^i &= 0, \quad i = 1, \ldots N, \\ \sum_{i=1}^{N} \lambda_k^i &= 0, \quad k = 1, \ldots M, \\ \lambda_k^i &\leq C y_{ik}. \end{aligned} \tag{3.15}$$

The solution to the *Gini* dual in equation (3.14) subject to the constraints (3.15) can now be using standard quadratic optimization techniques that are available in several packages [58–60].

3.1.2 Probabilistic Labeling of Respiratory Data

The *Gini* kernel machine framework described above estimates the probabilities of a discrete set of classes. As a result the respiratory signal that is to be estimated must be converted in probabilities. This can be accomplished by viewing the estimation of the breathing signal to be a two class problem with *class*-1 representing *expiration* (or exhalation) and *class*-2

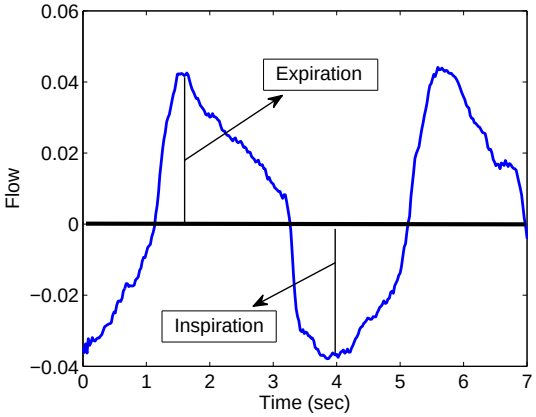
representing *inspiration* (or inhalation). Furthermore, the exact value of the respiratory (flow) curve obtained from the spirometer can be mapped to soft probability labels using the logistic transform. Consider for example the spirometer output shown in Figure 3.4 (a). Here positive values represent expiration (or *class*-1) and negative values represent inspiration (or *class*-2). The probability $y_{i1} = P(C_1|\mathbf{x}_i)$ can be obtained by

$$y_{i1}(t) = \frac{1}{1 + exp\,(-c\phi_i)} \tag{3.16}$$

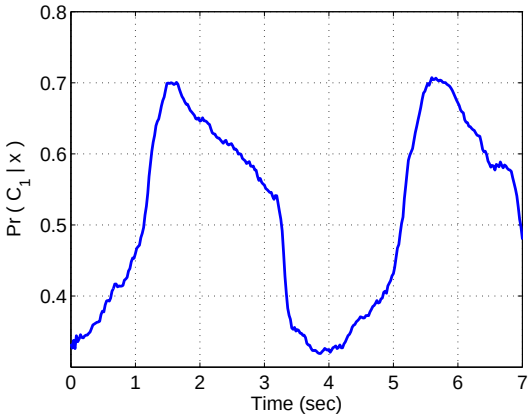
where, c is a constant controlling the shape (or saturation) of the logistic curve and ϕ_i is the value of the spirometer signal at the i-th time sample. Since there are only two classes the probability $y_{2i} = 1 - y_{1i}$. Probability labels obtained by application of the logistic transform of equation (3.16), with c=20, to the spirometer signal in Figure 3.4 (a) are shown in Figures 3.4 (b) and (c).

3.1.3 Results Gini Kernel Machine

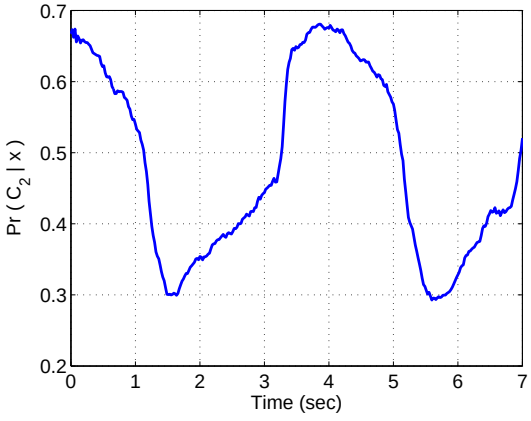
For evaluation of respiration rate error the the evaluation criteria is made more stringent. In chapter 2 a temporal window of 60 seconds with an overlap of 25 seconds was employed. This is now reduced to a window length to 10 seconds with an overlap of 5 seconds. A shorter windows implies that even very small duration errors will be taken in to consideration when evaluating the breathing rate error. The output of the *Gini*-SVR is denoised using the daubechies wavelet in order to eliminate noise. The window length employed for denoising is 20 seconds. As done in the previous chapter data sessions are grouped into two categories: (1) Artifact Sessions (or subsets)during which the subject is mobile but breathing at a normal rate; and (2) Accelerated Breathing and Apnea Sessions during which the subject is stationary and breathing at an accelerated rate or holding breath; these sessions do not contain motion artifacts. The average breathing rate error for all subjects is presented in Table 3.1. Also presented in Table 3.1 is the breathing rate error obtained when employing on a single impedance-plethysmographic electrode sensor. Electrode-4 is selected for comparison



(a)



(b)



(c)

Figure 3.4: Probabilistic transformation of respiratory signal (a) Reference Spirometer output, positive values of flow indicate expiration, negative values indicate inspiration (b) Plot of $y_{i1} = P(C_1|\mathbf{x}_i)$ (or expiration probability) versus time (expiration) (c) Probability of $y_{i2} = P(C_2|\mathbf{x}_i)$ (or inspiration probability) versus time.

Subject	Artifact Sessions			Accelerated Breathing & Apnea Sessions		
	Elec-4	SEC	Gini	Elec-4	SEC	Gini
1	4.54	2.29	**1.81**	11.67	**8.67**	15.29
2	5.65	2.04	**1.93**	12.41	**5.78**	15.74
3	3.07	19.06	**1.90**	**10.52**	13.11	12.25
4	7.93	5.76	**3.60**	16.52	**2.34**	8.74
5	7.23	3.66	**3.13**	15.99	4.65	**4.13**
6	5.06	5.10	**3.23**	**4.93**	5.18	12.94
7	8.41	8.33	**2.54**	28.58	32.85	**16.03**
8	12.08	4.63	**3.75**	12.49	**3.90**	16.49
9	**2.73**	2.77	2.88	3.70	**3.50**	5.85
10	4.77	3.77	**2.57**	12.15	**6.59**	15.34
11	6.26	2.75	**2.27**	9.95	**9.64**	15.56
12	6.19	2.83	**2.47**	**9.52**	9.68	14.15
13	8.86	3.38	**2.48**	8.42	**7.77**	10.10
14	3.14	2.94	**2.72**	**7.93**	11.30	13.61
15	3.39	3.18	**2.80**	5.10	7.42	**6.29**
16	3.29	2.69	**2.54**	10.14	**9.17**	15.56
17	5.64	4.42	**3.49**	8.92	**3.32**	16.74
18	6.11	3.82	**2.95**	7.06	**3.54**	7.79
19	4.60	4.79	**4.20**	6.37	**2.77**	11.01
Mean	5.73	4.62	**2.80**	10.65	**7.96**	12.29

Table 3.1: Average Respiration Rate Error (RR_{err}) in BPM for different human subjects. Errors are computed over 10 second windows with 5 second overlaps.

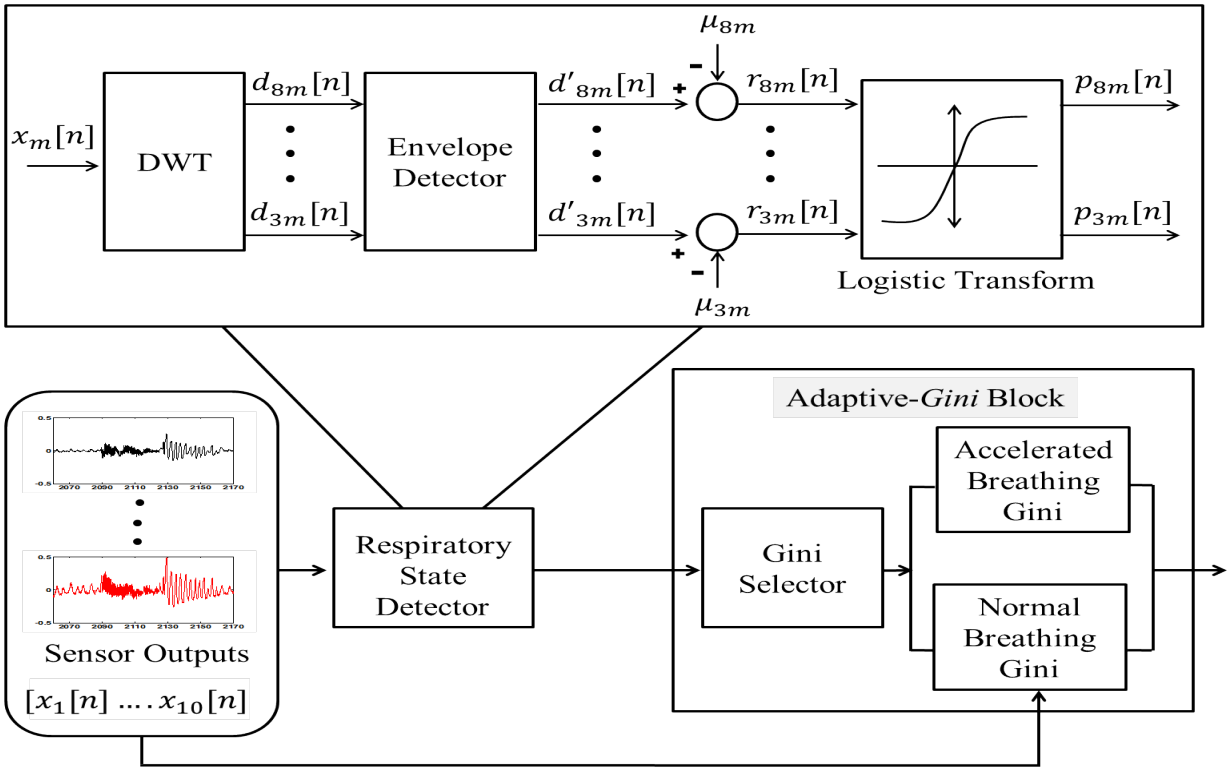


Figure 3.5: WA-Gini block diagram. The top plot illustrates the main steps of the respiratory state detector (see equations (3.17) to (3.19)).

because it gives the breathing rate error amongst all the 10 individual sensors (as illustrated in Figure 2.2).

Results demonstrate that use of the *Gini* kernel machine results in significant performance improvements in artifact sessions where the average reduction in breathing rate error is almost equal to 2 BPM when compared to the SEC. The *Gini* kernel machine outperforms both electrode-4 and SEC in artifact sessions for all subjects except one (subject-9). Unfortunately, it seems that performance degrades in accelerated breathing and apnea sessions where the mean error increases from 7.96 BPM to 12.29 BPM when compared to the SEC. The problems that cause this performance degradation are discussed and remedied in the next section.

3.2 "WA-Gini" Wavelet Adaptive Gini Kernel Machines

Performance degradation of *Gini* kernel machine in accelerated breathing and apnea can be attributed to two factors: (1) Training data is slightly unbalanced because there are relatively larger instances of each subject breathing normally than there are of accelerated breathing and apnea. As a result, the learning algorithm is biased towards normal breathing rates. (2) Denoising suppresses or eliminates high-frequencies in regions containing accelerated breathing. To overcome these problems an additional respiratory state detection block is added on top of the *Gini* kernel machine algorithm. Two enhancements are made to the *Gini* kernel machine regression framework discussed in the preceding section. First, wavelet filtering is employed to accurately identify accelerated-breathing and apnea regions. Second, a separate *Gini* kernel machine trained just on accelerated breathing data is also added. This means that the framework now has two *Gini* kernel machines; one for normal breathing and one for accelerated breathing. Upon detection of an accelerated region, the algorithm switches to the accelerated breathing *Gini*. This algorithm is titled the Wavelet-Adaptive-*Gini* (or WAGini) and is illustrated in Figure 3.5. The respiratory state detector employs wavelet filtering for detecting the presence of accelerated breathing. The *Gini-selector* block switches between the accelerated and normal-breathing *Ginis* based on the output of the respiratory state detector. For comparison purposes results for a much simpler respiratory state detector, which employs the Discrete Cosine Transform (DCT), are also presented in section 3.2.4.

3.2.1 Respiratory State Detection using Wavelets Filters

Knowledge of the subject's respiratory state can enable further enhancement of the performance of the learning algorithm. For example, if the learning algorithm knows with high confidence that the subject is breathing normally then it knows that the high-frequencies in the electrodes are caused by noise or interference and should be attenuated. Similarly,

knowledge that subject is breathing at an accelerated rate should trigger the inverse process resulting in amplification of high frequencies and suppression of lower frequencies. The SEC rate estimation does utilize rate respiratory state detection to differentiate between hyperventilation and apnea regions however, it makes a hard decision. The respiratory state in a frame is classified as either accelerated breathing or apnea depending on the energy in level in the high frequency bands. There are a number of disadvantages to using this approach; the primary being constant resolution in the spectral and temporal domains. Use of fixed length time windows implies that a wrong decision will impact the entire frame. The Single Gini approach operates at the other end of the spectrum it has much higher resolution since it predicts the value of each individual sample and although it has much better temporal resolution, a few noisy samples in its output can increase the rate estimation error. Another problem arises from the fact that the physiological breathing rates are located well below the sampling rate (< 0.05 f$_s$) making it difficult to design traditional filters that can differentiate between the various frequency bands spanning the physiological frequencies.

Fortunately, tools like wavelet based filtering provide an efficient way to achieve high frequency resolution at lower frequency bands. The discrete wavelet transformation (DWT) of a signal of length N can be obtain efficiently via multi-resolution analysis (MRA) proposed by Mallat [61]. The MRA based DWT for three decomposition levels is illustrated in Figure 3.6. MRA can be considered to have a tree like structure where at each level the input is filtered using either a low-pass filter (LPF), $h[n]$, or a high-pass filter (HPF), $g[n]$. The impulse response of each filter depends on mother wavelet being employed. The LPF $h[n]$ at each level retains the lower half of the input frequency band and discards the upper half whereas; the HPF $g[n]$ does the opposite and retains only the upper half of the frequency band. For example, at level-1 the highest frequency in the input signal is π rad/sec (corresponding to half of the sampling frequency in Hz). The output of the level-1 LPF $h[n]$ therefore contains the lower half of the input frequency band $[0 - \pi/2]$ whereas, output of the corresponding HPF, $g[n]$, covers the upper half; the $[\pi/2 - \pi]$ frequency band. Since

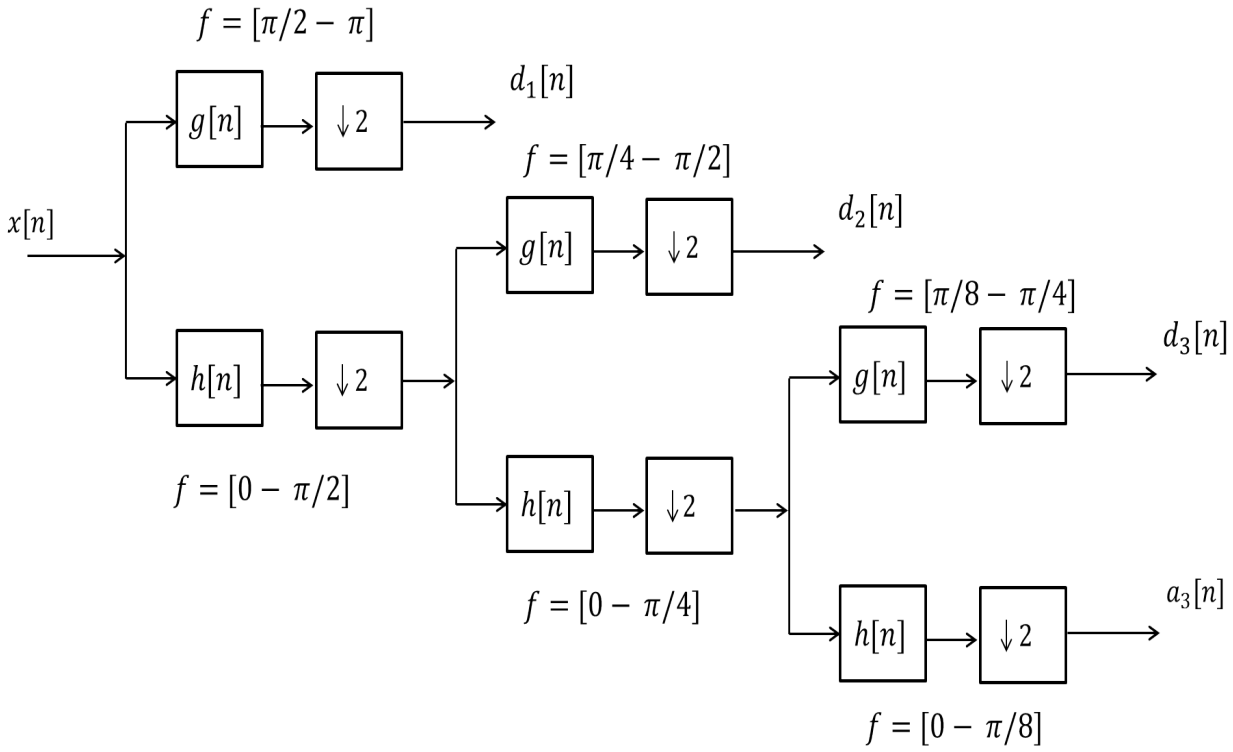


Figure 3.6: Wavelet decomposition using Multi-resolution analysis.

the output of filters contains half of the original frequency half of the time samples in the output become redundant and therefore can be discarded by down-sampling by a factor of 2. As more levels are added to the tree the frequency resolution starts to increase, doubling at each stage.

The enhanced frequency resolution achieved by the application of DWT can be very useful for differentiating between the various respiratory states which are located very close to each other in the frequency domain. An illustration of this is presented in Figure 3.7 where the top plot contains the reference spirometer signal of a subject in different respiratory states. Decomposition levels d_3, d_5, d_6 and d_8 are shown in the lower plots. A lower wavelet decomposition levels corresponds to higher frequency bands wheres as high decomposition levels represent lower frequencies. It can be seen that d_3 coefficients have the highest value in shallow high breathing (or hyperventilation) regions and are almost zero in normal and slow breathing regions. This means that decomposition coefficients d_3 may be employed

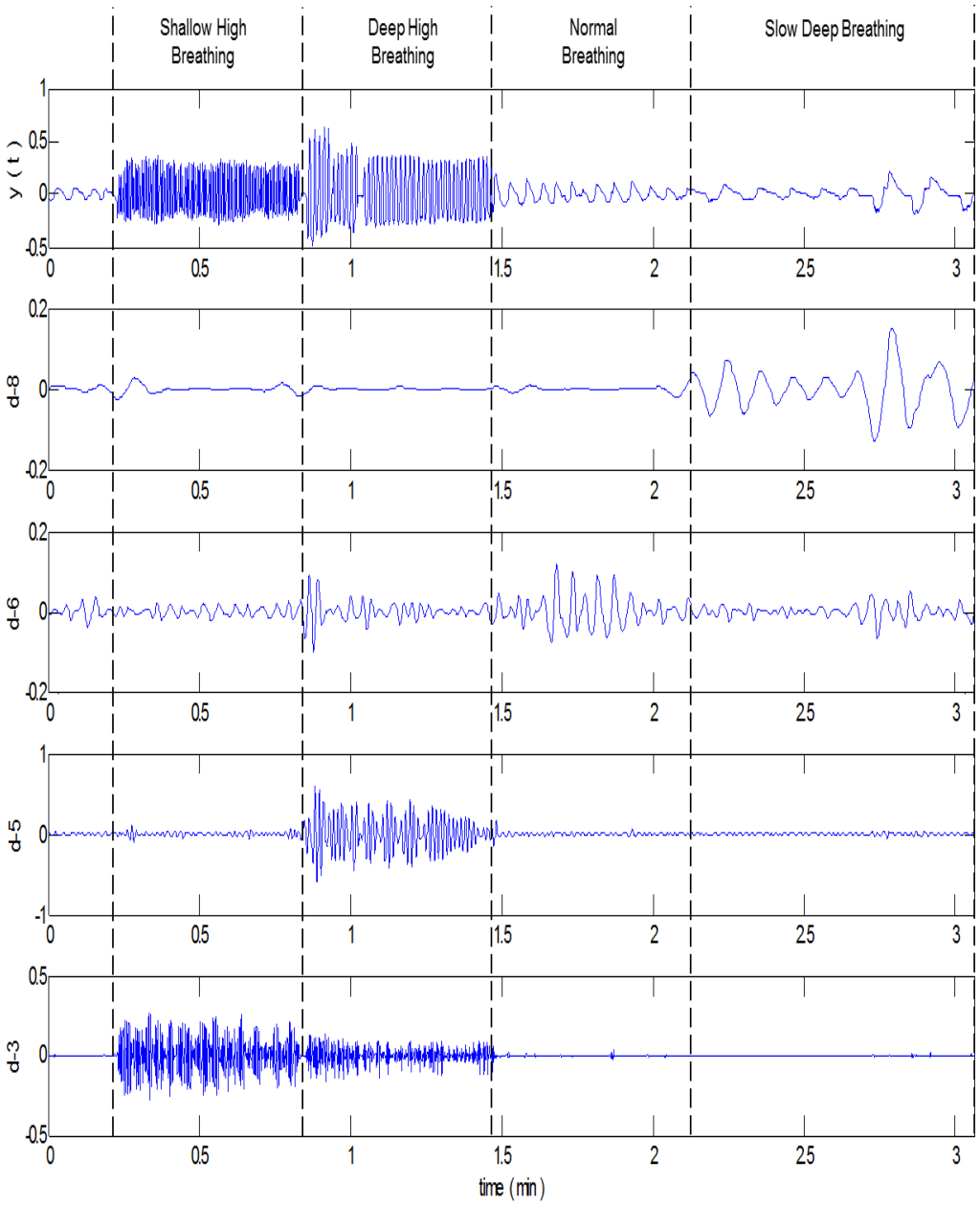


Figure 3.7: Reference spirometer signal $y(t)$ and its corresponding Daubechies-Wavelet [62] details at different levels.

to differentiate hyperventilation regions from normal and slow breathing regions. Similarly, decomposition d_5 seems to accurately indicate the presence of Deep High breathing regions whereas decompositions d_6 and d_8 seem to presence of normal shallow breathing regions or respiratory states. The next section describes how probabilistic curves can be computed from the multiple electrode signals to identify different respiratory states.

3.2.1.1 Region Score Computation

Ideally the identification of any respiratory state or event should be based on a probabilistic measure as it would enable making soft decisions making the overall framework flexible. The advantages of such a setup will become clear as its details are examined in the forthcoming discussion. Figure 3.7 demonstrates that different respiratory states are easy to identify at distinct decomposition levels for instance, hyperventilation is most easily identifiable at level-3 whereas, slow-deep-breathing lies at level-8. Therefore, a measure for detection of a certain respiratory state can be constructed from the values of detail coefficients of the wavelet decomposition level at which it occurs. However, the wavelet decompositions in Figure 3.7 are derived from the spirometer signal; the actual measures will need to be based on the outputs of multiple electrode channels. The top plot Figure 3.8 displays the output of electrode-1 corresponding to the spirometer signal shown in Figure 3.7. Note the high-level of noise in d_3 shown in the bottom plot. This explains the reason why it has been so challenging to extract the breathing rate during hyperventilation so far. There exists a significant amount of noise in the frequency band containing hyperventilation rates. The problem is further aggravated by the fact that the lung volume changes are very small in this state making the energy of the desirable frequencies quite low. Fortunately it seems that the energy is large enough to be detectable since the wavelet details are higher in the hyperventilation region than they are in the other states. In addition to this, use of multiple electrode channels can also salvage the situation.

We now derive two probabilistic scores based on the wavelet detail curves that enable us to

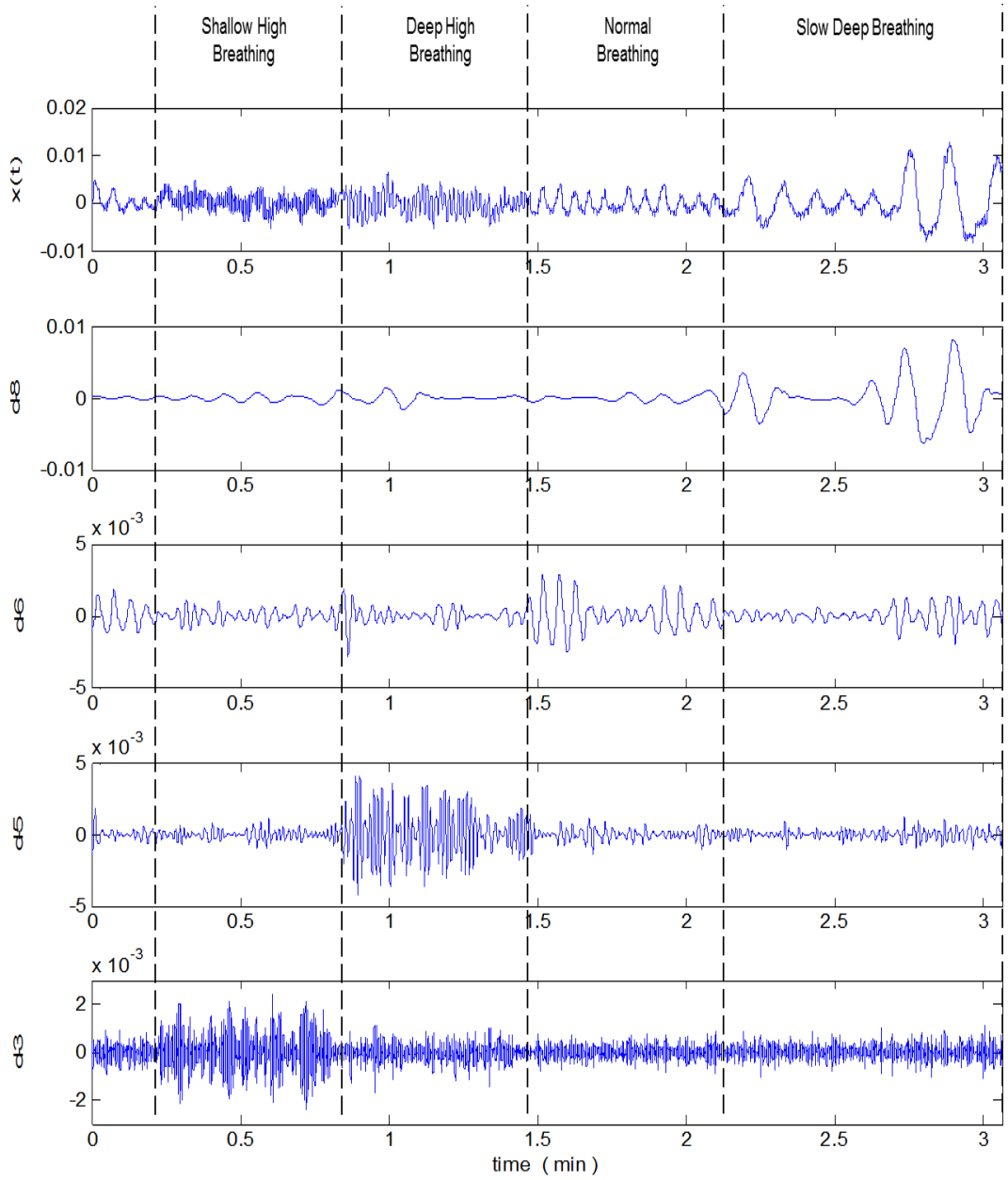


Figure 3.8: Electrode-1 output $x_1(t)$ and its corresponding Daubechies-Wavelet [62] details at different levels.

make decisions as to whether the underlying respiratory state is normal (or low) breathing, hyperventilation or apnea. A pleasantly surprising, bonus, outcome of using these score curves is that in addition to respiratory state detection they also enable detection of motion artifacts with very high temporal resolution and accuracy. Given $d_{lm}(t)$ the wavelet detail curve corresponding to level-l of electrode-m we first extract its envelope:

$$d'_{lm}(t) = \text{Envelope}[d_{lm}(t)] \tag{3.17}$$

The envelope is employed here because the intent is to capture the slow temporal variations of the wavelet curve. For a given test signal we obtain a distance measure $r_{lm}(t)$ as below:

$$r_{lm}(t) = d'_{lm}(t) - \text{E}_{trn}[d'_{lm}(t)] \tag{3.18}$$

Here $\text{E}_{trn}[d'_{lm}(t)]$ represents the expected value of $d'_{lm}(t)$ in the training data. $r_{lm}(t)$ is a simple distance measure that indicates deviation of the wavelet detail-l (of electrode-m) above or below it mean value in the training data. $r_{lm}(t)$ can be converted into a probabilistic measure using the logistic function:

$$p_{lm}(t) = \frac{1}{1 + exp\left(-cr_{lm}(t)\right)} \tag{3.19}$$

where, c is a constant controlling the steepness of the curve[2]. Positive values of $r_{lm}(t)$ are transformed to $p_{lm}(t) \in (0.5, 1]$ whereas negative values of $r_{lm}(t)$ result in $p_{lm}(t) \in [0, 0.5)$. $p_{lm}(t) = 0.5$ when $r_{lm}(t) = 0$. As illustrated in Figures 3.7 and 3.8 the high or accelerated breathing rates are generally located in levels 3 and 4. Therefore, the probability that a sample, at time t, from electrode-m belongs to an accelerated respiratory state is given by:

$$p'_m(t) = \frac{1}{2}\left(p_{3m}(t) + p_{4m}(t)\right) \tag{3.20}$$

Simiarly, the probability that a sample from electrode-m belongs to a normal (or low) breathing state is obtained by merging the probabilities obtained from levels 5 to 8:

$$\bar{p}_m(t) = \frac{1}{4}\sum_{l=5}^{8} p_{lm}(t) \tag{3.21}$$

[2] $p_{lm}(t)$ represents the value of probability score p_{lm} at time t.

The overall probability that a vector consisting of single sample from all 10 electrodes belongs to accelerated respiratory state is obtained by combining the probability scores of the individual electrodes:

$$p'(t) = \sum_{m=1}^{10} \omega_m p'_m(t) \tag{3.22}$$

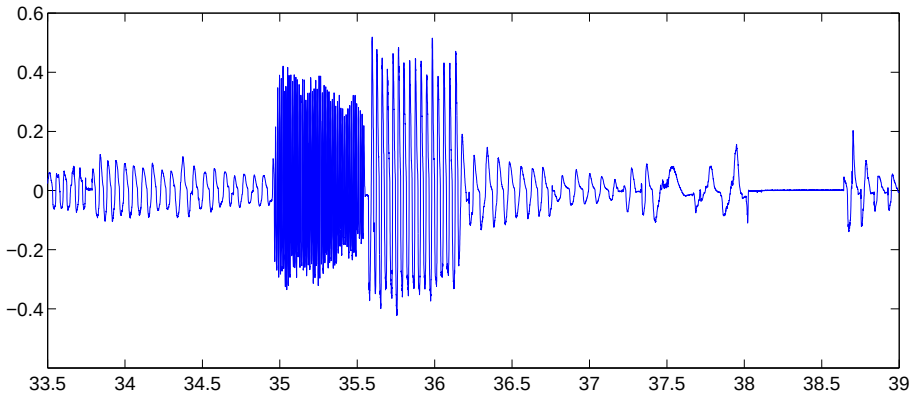
Where ω_m represents the weight assigned to electrode-m. Furthermore the electrode weights must sum to 1 in order to ensure a valid probability score.

$$\sum_{m=1}^{10} \omega_m = 1 \tag{3.23}$$

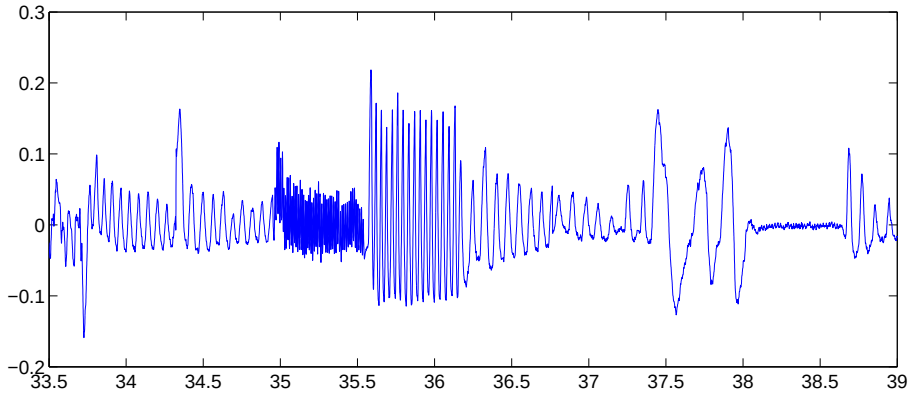
The results presented here assign equal weights ($\omega_m = 1/10,\ m = [1 \ldots 10]$) to all electrode however, it is also possible to assign electrode weights based on some signal quality indicators (SQI). In such a setup electrodes that are considered to be noisy, based on the value of the SQI, should be assigned a lower weight whereas electrodes with low noise should be assigned higher weights. The combined probability of a sample belonging to normal or low breathing respiratory state is given by:

$$\bar{p}(t) = \sum_{m=1}^{10} \omega_m \bar{p}_m(t) \tag{3.24}$$

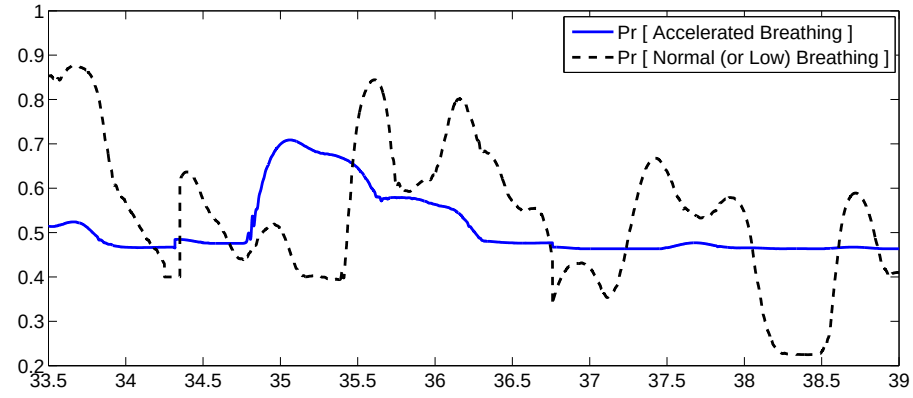
The normalization constrain in equation (3.23) applies in this case as well. An illustration of how the probability scores $p'(t)$ and $\bar{p}(t)$ enable classification of different respiratory states is presented in Figure 3.9. It is apparent that the probability scores vary according to the underlying respiratory state. For instance the value of $p'(t)$ is close to 0.5 in normal, low breathing and apnea regions. It rises sharply in the hyperventilation region (between the 35 min to 35.5 min mark) and then begins to drop. Similarly, a normal or low breathing state is indicated by the dominance of the low probability curve $\bar{p}(t)$. An apnea region can be is characterized by $p'(t) \approx 0.5$ and extremely low values of $\bar{p}(t)$. Ideally, $p'(t)$ should also approach 0 in an apnea region however, practically this does not happen because of the presence of noise in the electrode signals. Note that the probabilities $p'(t)$ and $\bar{p}(t)$, at a time sample t, do not sum to 1 this is because almost all respiratory states have some contribution from both high and low frequencies. The above discussion indicates that comparison of the



(a)



(b)



(c)

Figure 3.9: Respiratory state detection (a) Spirometer output (b) Mean of all 10 electrodes (c) Probability curves: solid line represents probability of accelerated breathing, $p'(t)$; dotted line represent probability of normal (or Low) breathing, $\bar{p}(t)$.

relative value of both probabilities allows us to make a correct decision about the underlying respiratory state. The highest priority in this work is assigned to differentiation between three respiratory states namely, hyperventilation (or accelerated breathing regions), apnea (or regions with zero breathing rates) and normal breathing. Although it is possible to differentiate between more respiratory states this thesis differentiates between only three. The decision making process about whether a sample at time t belongs to a hyperventilation, apnea or normal breathing region is described below.

Hyperventilation: A sample at time t belongs to hyperventilation respiratory state if its accelerated breathing probability score $p'(t)$ is greater than 0.5 and the normal breathing probability score $\bar{p}(t)$ is less than $p'(t)$. This can be represented in terms of an indicator function, $\mathbb{1}_H(t)$ as below:

$$\mathbb{1}_H(t) = \begin{cases} 1, & p'(t) > \eta_w, \text{ and } \bar{p}(t) < p'(t) \\ 0, & \text{otherwise} \end{cases} \tag{3.25}$$

The threshold η_w is fixed at 0.5 for the majority of subjects. Data for three subjects contains significant high frequency noise, in these case η_w was assigned values slightly less than 0.5.

Apnea: An apnea region contains is characterized by low values of both accelerated and normal breathing scores because it contains no genuine breathing cycles. This can be represented in terms of an indicator function, $\mathbb{1}_A(t)$ as below:

$$\mathbb{1}_A(t) = \begin{cases} 1, & p'(t) \leq 0.5, \text{ and } \bar{p}(t) < \xi_w \\ 0, & \text{otherwise} \end{cases} \tag{3.26}$$

The threshold ξ_w is fixed at 0.3 for all subjects.

Normal Breathing: All samples that do not belong to either an apnea or hyperventilation state are considered to be generated from a normal breathing state. The indicator function $\mathbb{1}_N(t)$ representing the locations of normal breathing samples can be obtained by applying the exclusive-OR operation to the indicator functions $\mathbb{1}_H(t)$ and $\mathbb{1}_A(t)$ and negating

the answer:

$$\mathbb{1}_N(t) = \neg\left[\mathbb{1}_H(t) \oplus \mathbb{1}_A(t)\right] \tag{3.27}$$

Where the symbols $\oplus$ and $\neg$ represent the exclusive-OR and negation operations respectively[3]. Since $\mathbb{1}_H(t)$ and $\mathbb{1}_A(t)$ are mutually exclusive therefore, the exclusive-OR operation gives the locations of the time samples where the respiratory state is either hyperventilation or apnea. Negation of the exclusive-OR output gives the location of all non-apnea and non-hyperventilation samples.

3.2.2 Respiratory State Detection using DCT Filters

In theory wavelet filtering should enable high-accuracy classification of different respiratory states however, in order to validate this claim results for respiratory state detection using DCT filtering are also presented. The mean sensor signal, $\bar{x}(t)$, is first divided into non-overlapping time frames of 5 seconds each. The i^{th} frame $\mathbf{x}_i$ consisting of $M(=5\times f_s)$ time samples[4] of $\bar{x}(t)$ is transformed to obtain the DCT coefficient vector $\mathbf{c}_i$:

$$\mathbf{c}_i = \mathbf{T}\mathbf{x}_i \tag{3.28}$$

Here, $\mathbf{T}$ denotes the $(M \times M)$ DCT matrix. Frame-i is considered to belong to an *apnea* region if the norm of its DCT coefficients $|\mathbf{c}_i|$ is significantly below a threshold ξ_D. In order to be classified as *accelerated-breathing*, frame-i must satisfy the following two conditions: **(1)** It must contain significant energy at higher frequency components. In other words, the ratio of the norm of its high-frequency DCT coefficients to the norm of all its DCT coefficients must exceed a threshold η_D. **(2)** It must not be an apnea frame. Most apnea frames also contain significant energy at higher spectral components due to high-frequency noise in the sensor output. Therefore, condition-(2) is imposed to avoid misclassifying an apnea frame as an accelerated-breathing frame. Frames that do not fall into apnea or accelerated breathing are

[3] All logical operations are modulo-2

[4] f_s represents the sampling frequency in Hz

classified as *normal-breathing* frames. The thresholds η_D and ξ_D control the decision function and are determined from their receiver-operating-characteristic (ROC) curves. That is, the DCT apnea threshold, ξ_D, is varied over a wide range to determine the apnea detection ROC curve which is a plot of the false-positive-rate (FPR) versus the true-positive rate (FPR). The value of ξ_D which results in a FPR of 1% is selected and used for respiratory state detection. Similarly, the value of η_D resulting in 1% FPR for accelerated breathing frames is selected from its corresponding accelerated breathing ROC curve. Note, in section 3.2.1 the wavelet thresholds η_w and ξ_w were fixed to 0.5 and 0.3 respectively. We would like to point out that these values also correspond to a 1% FPR for apnea and accelerated breathing regions. Limiting both the wavelet and DCT respiratory state detectors to the same FPR ensures that the comparison is fair.

3.2.3 Rate Estimation Using Adaptive Gini Kernel Machines

In order to improve estimation of breathing rate in accelerated breathing regions the WA-*Gini* algorithm employs a *Gini* kernel machine trained only on accelerated breathing samples . Upon detection of an accelerated breathing region the algorithm switches from the normal breathing *Gini* to the accelerated breathing kernel machine. For apnea regions, one possibility is to set the output to zero. However, this approach may result in large rate estimation errors in case of false alarms. Therefore, a soft-decision approach is employed i.e.; the normal-breathing *Gini* is used in apnea regions as well. The main issue with apnea regions is high-frequency noise which is removed by the wavelet denoising applied at the output of the normal breathing *Gini*; this results in zero, or minimal, breathing rate estimation error for the vast majority of apnea regions. Results indicate that this approach performs better than explicitly setting apnea regions to zero. In order to evaluate which of the two respiratory state detectors performs better results for both configurations. The framework that employs the wavelet based respiratory state detection is titled "Wavelet Adaptive *Gini*" or WAGini. Whereas, the DCT based framework is titled the "DCT Adaptive *Gini*" or DAGini. Both

approaches are identical in all other respects with the only difference being the respiratory state detector.

3.2.4 Results

The breathing rate errors obtained using a number of different techniques for 19 different human subjects in the normal breathing artifact sessions are presented in Table 3.2. The results for accelerated breathing and apnea sessions are contained in Table 3.3. The results for DAGini, in accelerated-breathing and apnea sessions, demonstrate noticeable performance improvement when compared to the single-*Gini* and the single sensor approaches. However, when compared to the SEC the performance improvement in accelerated breathing sessions is very small. In artifact sessions, the DAGini results in more than 1 BPM improvement when compared to the SEC however, it performs worse than the single-*Gini* approach. This degradation is due to false detections of accelerated breathing frames in normal breathing sessions of some subjects. Consider for example, the single-*Gini* results in artifact sessions of subject-1 and subject-2. For subject-1 the error rates for single-*Gini* and DAGini are identical because there are no false detections of accelerated frames. For subject-2 however, the single-*Gini* error rate is lower than that for DAGini because of high noise in certain normal breathing sessions. This results in false detection of accelerated breathing regions by the DCT based respiratory state detector causing an increase in the overall breathing rate error.

The results for WAGini demonstrate that wavelet filtering coupled with multiple kernel machines produce the best estimate of breathing rate in accelerated-breathing and apnea sessions. When compared to the SEC and DAGini we obtain more than 2 BPM improvement in the average error estimate. In artifact sessions the single-*Gini* still remains the best performing approach however, the difference with WAGini is not very significant. As was case in DAGini the errors in this case may also be attributed to false detection of accelerated breathing regions since the wavelet approach is not perfect. However, the overall rate error

Subject	Artifact Sessions				
	Elec-4	SEC	Gini	DAGini	WAGini
1	4.54	2.29	**1.81**	**1.81**	**1.81**
2	5.65	2.04	**1.93**	2.77	3.58
3	3.07	19.06	**1.90**	10.07	4.68
4	7.93	5.76	**3.60**	**3.60**	4.00
5	7.23	3.66	**3.13**	**3.13**	**3.13**
6	5.06	5.10	**3.23**	5.33	**3.23**
7	8.41	8.33	**2.54**	5.49	**2.54**
8	12.08	4.63	**3.75**	**3.75**	**3.75**
9	**2.73**	2.77	2.88	2.88	2.88
10	4.77	3.77	**2.57**	**2.57**	**2.57**
11	6.26	2.75	**2.27**	**2.27**	**2.27**
12	6.19	2.83	**2.47**	**2.47**	**2.47**
13	8.86	3.38	**2.48**	**2.48**	**2.48**
14	3.14	2.94	**2.72**	**2.72**	3.26
15	3.39	3.18	**2.80**	**2.80**	**2.80**
16	3.29	2.69	**2.54**	**2.54**	**2.54**
17	5.64	4.42	**3.49**	**3.49**	**3.49**
18	6.11	3.82	**2.95**	**2.95**	**2.95**
19	4.60	4.79	**4.20**	**4.20**	4.27
Mean	5.73	4.62	**2.80**	3.54	3.09

Table 3.2: Average Respiration Rate Error (RR_{err}) in BPM for different human subjects in *artifact sessions*. Errors are computed over 10 second windows with 5 second overlaps.

Subject	Accelerated Breathing & Apnea Sessions				
	Sensor-4	SEC	Gini	DAGini	WAGini
1	11.67	**8.67**	15.29	10.27	11.90
2	12.41	**5.78**	15.74	7.91	6.21
3	**10.52**	13.11	12.25	17.62	10.73
4	16.52	**2.34**	8.74	5.12	3.04
5	15.99	4.65	4.13	1.38	**1.33**
6	4.93	5.18	12.94	8.99	**3.08**
7	28.58	32.85	16.03	15.89	**12.74**
8	12.49	**3.90**	16.49	8.17	4.78
9	3.70	3.50	5.85	**3.09**	3.48
10	12.15	**6.59**	15.34	7.64	9.84
11	9.95	9.64	15.56	10.99	**7.84**
12	9.52	9.68	14.15	15.38	**4.09**
13	8.42	7.77	10.10	8.16	**7.27**
14	7.93	11.30	13.61	5.30	**3.92**
15	5.10	7.42	6.29	4.43	**2.46**
16	10.14	9.17	15.56	4.06	**3.67**
17	8.92	3.32	16.74	3.48	**2.66**
18	7.06	**3.54**	7.79	4.20	3.98
19	6.37	**2.77**	11.01	6.64	7.44
Mean	10.65	7.96	12.29	7.83	**5.81**

Table 3.3: Average Respiration Rate Error (RR_{err}) in BPM for different human subjects in *accelerated breathing & apnea sessions*. Errors are computed over 10 second windows with 5 second overlaps.

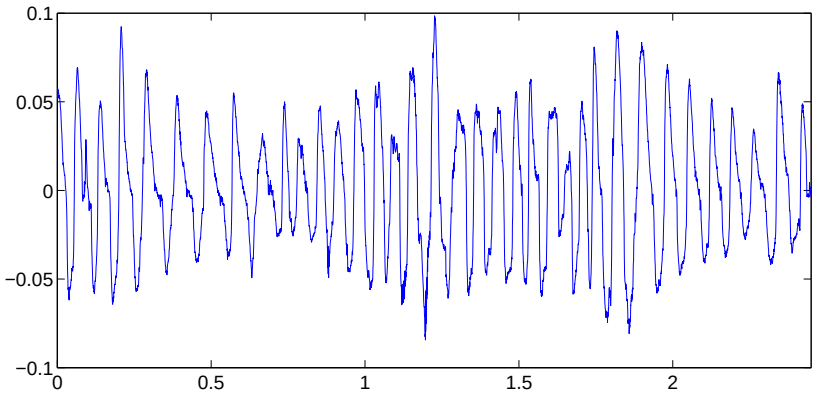
is still much less than that obtained via the DAGini. Overall it seems that the WAGini enables us to achieve a balance between rate estimation in artifact and accelerated breathing sessions.

3.2.5 Wavelet Based Artifact Detection

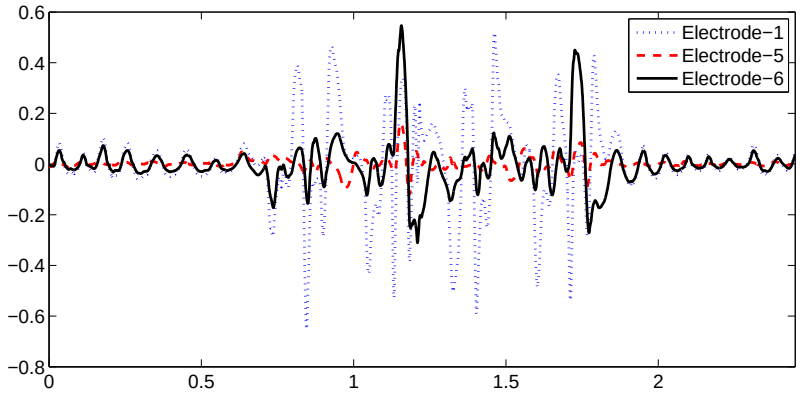
An added advantage of using the wavelet based probability curves is that they also enable very accurate detection of regions containing motion artifacts. Even though these metrics were not constructed primarily for this purpose they seem to be very good at identifying the presence of motion artifacts. Furthermore, it seems that they may also enable identification of the underlying physical state/ posture of the human subject. This section briefly discusses the possibility of using the probability scores of section 3.2.1.1 for artifact detection. The discussion is kept brief on purpose but it paves the way for any possible future work. A number of different motion artifact introduce low-frequency high-amplitude distortions in the electrode signals. For example consider the plot of electrode outputs shown in Figure 3.10 (b). In this case the subject is reaching for an object between the 0.5 to the 2 minute mark. It can be observed that the artifact signals are severely distorted. A brief glance at the value of the probability curves indicates that low frequency curve $p'(t)$ is significantly above the average level for the duration of the artifact. The high-frequency curve is also above the average value most of the artifact duration. Therefore, it seems that very high values of the $p'(t)$ curve may enable very accurate identification of artifact regions.

In addition to artifact detection it seems that we may also be able to identify the underlying physical state of the subject as well. For example, comparison of the plots in Figure 3.11 which correspond to a reaching activity; and Figure 3.12 which correspond to walking indicate that during walking the the difference between the high-frequency and low-frequency curves is smaller than difference during reaching. Similarly, the plots in Figure 3.13 correspond to the case when the subjects are on a bed rolling from left to right. Therefore, it seems highly likely that the exact shape of the score curves will allow differentiation between

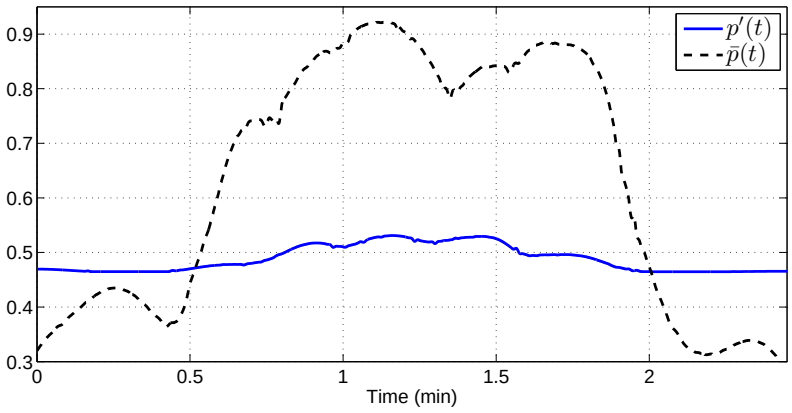
the underlying physical state. This type of knowledge may enable further improvement in the breathing rate estimate obtained via algorithms such as the WA-*Gini* however, they have been left for future work and have not been investigated further in this thesis.



(a)



(b)



(c)

Figure 3.10: Impact of motion-artifact on electrodes and probability curves; subject is reaching for object between the 0.5 to 2 min mark. Plots indicate: (a) Spirometer output; (b) Outputs of three electrodes and; (c) Probability curves.

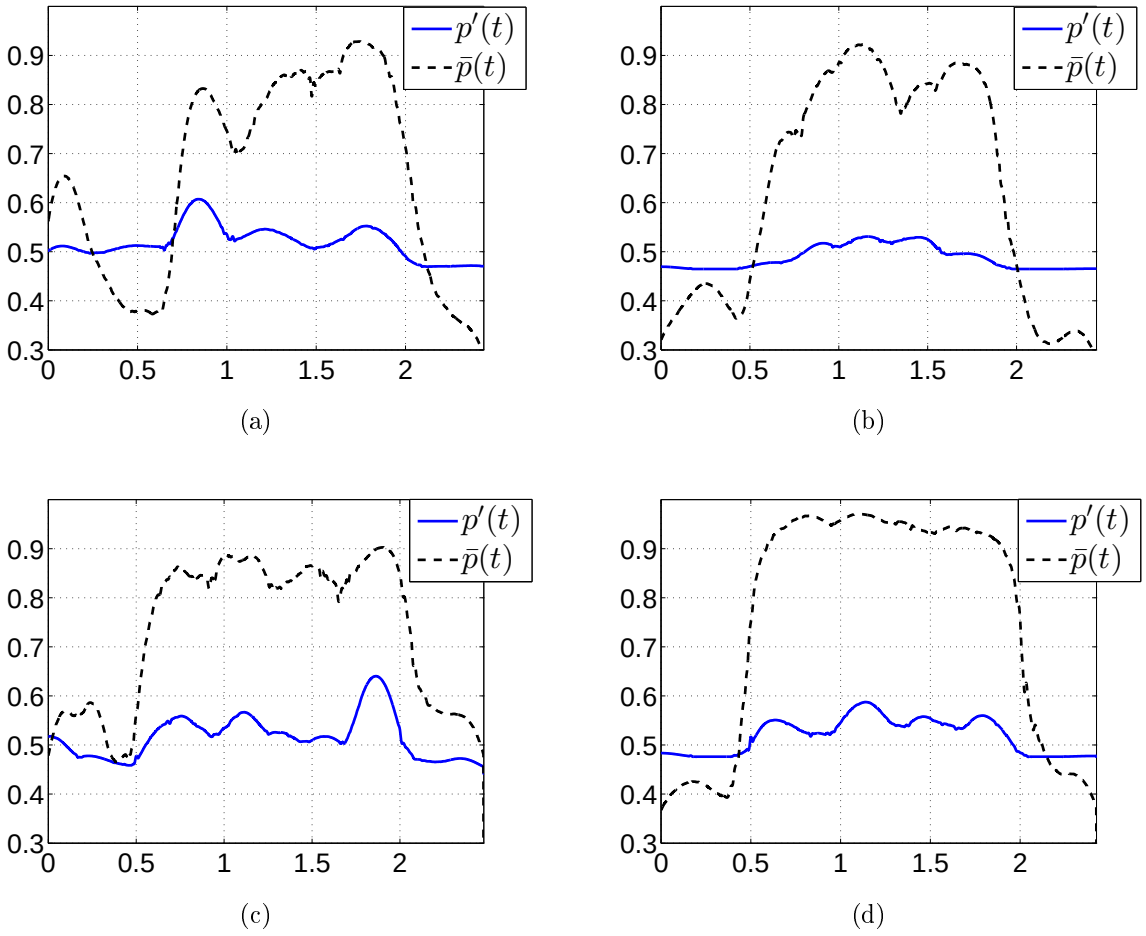


Figure 3.11: Probability curves of four different subjects when reaching for object.

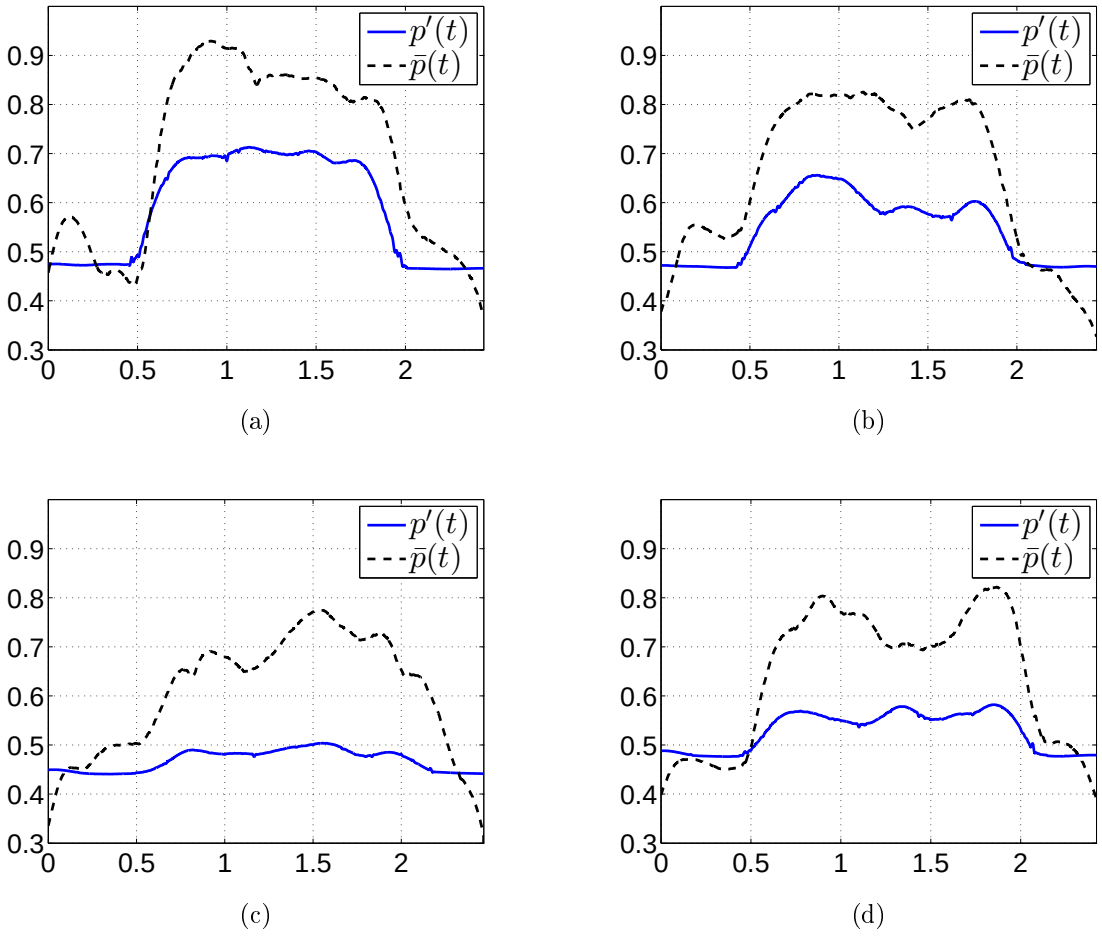


Figure 3.12: Probability curves of four different subjects walking at a normal pace.

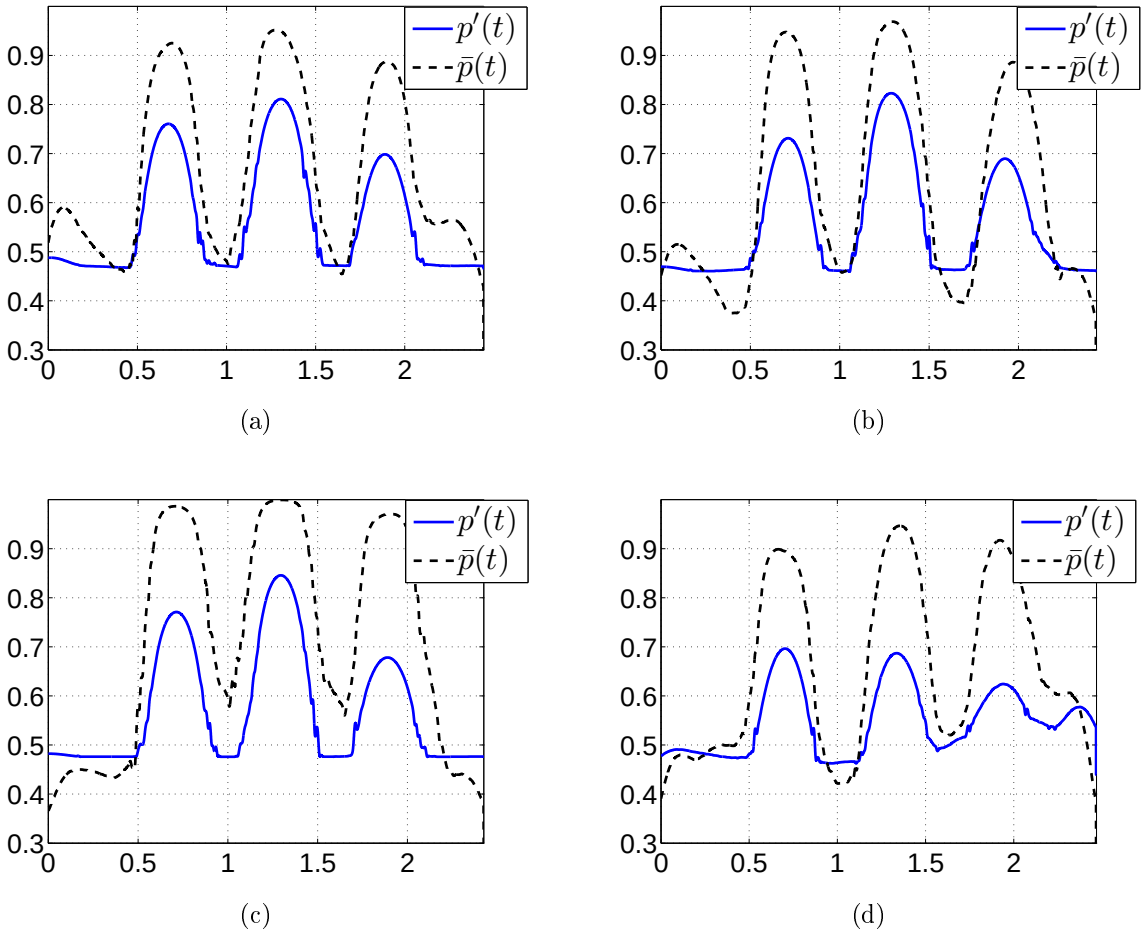


Figure 3.13: Probability curves of four different subjects when rolling left and right on bed.

CHAPTER 4

PROTEOMIC CHANNEL CAPACITY

This chapter presents a detailed strategy for evaluating the channel capacity of a proteomic channel. In order to be of any practical use the channel capacity calculations must be based on realistic channel conditions. This requires development of models that incorporate noise irregularities that may degrade protein detection performance. This chapter is organized as follows: Section 4.1 introduces the basic components of the protein receptor channel. Section 4.1.1 presents a model of the diffusion process that is relevant for sensing applications such as protein arrays. A *two-compartment* approach is employed by sub-dividing the diffusion process into two stages: (1) Large-scale, deterministic, diffusion from transmitter to receiver probe (2) Small-scale, stochastic, diffusion in a small volume around the receptor. Section 4.1.2 describes the response of combinatorial and specific probes and highlights in the impact of different parameters on detection performance. Section 4.2 introduces the conditional distribution of the protein array channel and discusses the impact of receptor parameters on the noise variance. Capacity is computed as a function of receptor parameters in Section 4.3.

4.1 Proteomic Channel Models

A typical affinity-based sensing application begins with the addition of a test sample to the array reaction chamber which contains a number of different probes immobilized in micrometer or nanometer sized spots throughout its entire volume. In the absence of any drift the protein particles present in the test sample follow a Brownian type motion and over time distribute evenly over the reaction chamber. The concentration of a protein is measured by using receptors that capture particles in their vicinity. A single *Receptor* consists of a large number of probes /recognition elements that are uniformly distributed over its surface.

The physical representation of a simplified dual protein assay with a single combinatorial receptor that detects both input proteins is shown in Figure 4.1. At the start of the test, a sample volume containing particles to be analyzed is injected into the reaction chamber. The change in particle concentration rate at the spot of the target receptor depends on the *Diffusion* properties of the medium (determined by factors such as the viscosity and the physical dimensions of the target proteins). The receptor samples the concentration information in its encompassing volume and depending on whether the probe is specific or combinatorial produces an output signal proportional to the concentration of one or more proteins. In the following sub-sections we present a mathematical formulation that models the diffusion process and the receptor binding process.

4.1.1 Protein Diffusion Model

While at a macroscopic level diffusion can be viewed as a deterministic process, at the microscopic level the perpetual Brownian motion of particles causes random variations in the transport of particles resulting in so-called "Diffusion noise". Although mass-transport limited biochemical systems can be modeled using stochastic differential equations such as the Langevin equation [63], the calculus of multivariate stochastic differential equations becomes cumbersome except for some special cases. In order to keep the model mathematically tractable we use a deterministic transport model based on the Fick's second law of diffusion. The variation due to the constant random motion of particles is modeled only within a small sub-volume around the receiver probe. For a protein array which performs multiplexed detection of N types of proteins, we denote the input of the system by a vector $\mathbf{x} \in \mathbb{Z}_{\geq 0}^{N}$. Each component of $\mathbf{x}$ is a non-negative integer random variable denoted by x_n and represents the total number of particles of protein-n in the test sample. Particles are injected into the reaction chamber using a device such as a micro-pipette and injection location will taken to be the origin $\Theta_I = (0, 0, 0)$ in the three dimensional space. Fick's second law can be used to predict how the concentration of protein-n changes over time at a receptor located at the

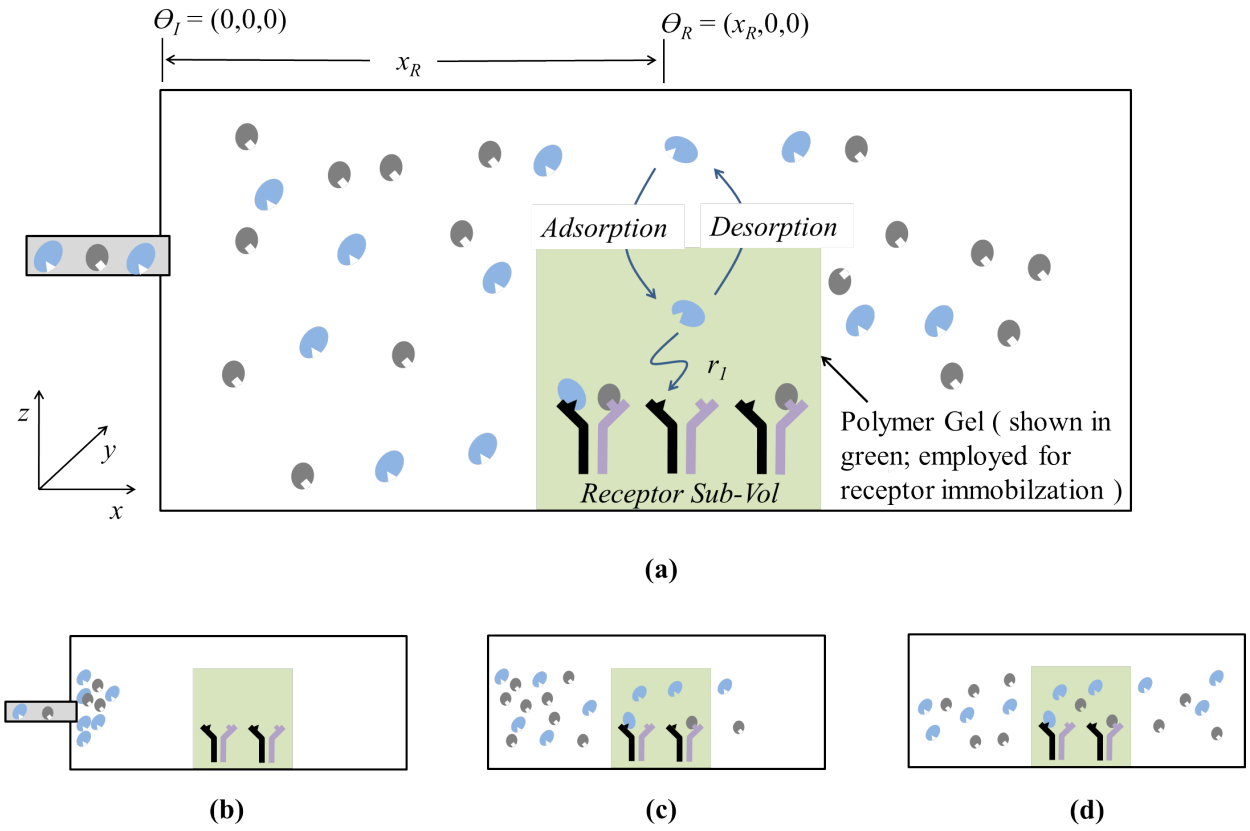


Figure 4.1: (a) Cross-sectional view of diffusion in a multi-protein array. Different states of the channel: (b) $t = 0$: X_n particles injected at origin $\Theta_I = (0, 0, 0)$; (c) $t > 0$: concentration, $\Lambda_n(\Theta_R, t)$, of particles in the receptor sub-volume is given by (5); (d) $t \to \infty$: (Steady-State) concentration, , of particles in the receptor sub-volume is given by (6).

coordinate $\Theta_R = (x_R, 0, 0)$ [64] and can be expressed as :

$$\frac{\partial \Lambda_n(\Theta_R, t)}{\partial t} = D_n \nabla^2 \Lambda_n(\Theta_R, t). \tag{4.1}$$

$\Lambda_n(\Theta_R, t)$ is the concentration of particles at the receptor location Θ_R at time t. D_n is the diffusion coefficient of the protein-n molecule. However, the model in equation (4.1) does not account for the interaction between the protein particles with the epitopes (binding sites) on the receptor and hence needs to be modified accordingly. The concentration of the protein-n within a small sub-volume where the receptors are immobilized (here onward referred to as the "Receptor Sub-Volume", V_R) is primarily effected by two processes:

1. *Sorption* which refers to the adsorption and desorption of protein particles to the

receptor surface or surface epitopes. We assume that the sorption rate is high compared to the transport rate of particles in the test sample, therefore, it is fair to assume that a local equilibrium exists for the adsorption and desorption processes. Under these conditions the sorption rate changes in proportion to the concentration [65]; this enables us to model sorption as a sink located at Θ_R:

$$\frac{\partial \Lambda_n(\Theta_R, t)}{\partial t} = D_n \nabla^2 \Lambda_n(\Theta_R, t) - K_d \frac{\partial \Lambda_n(\Theta_R, t)}{\partial t} \tag{4.2}$$

where, K_d is the equilibrium-partitioning coefficient between the fluid and the sorption to the receptor surface.

2. *Binding* of adsorbed particles to epitopes on the receptor. Binding between the adsorbed particles and the receptor epitopes can be modeled using an additional "sink" factor to equation (4.2) according to:

$$\begin{aligned} \frac{\partial \Lambda_n(\Theta_R, t)}{\partial t} &= D_n \nabla^2 \Lambda_n(\Theta_R, t) - K_d \frac{\partial \Lambda_n(\Theta_R, t)}{\partial t} \\ &\quad - r_n \Lambda_n(\Theta_R, t) \end{aligned} \tag{4.3}$$

where, r_n is the reaction rate at which protein-n targets bind with the epitopes on the receptor.

Equation (4.3) can rearranged to give:

$$\frac{\partial \Lambda_n(\Theta_R, t)}{\partial t} = \frac{D_n}{R_s} \nabla^2 \Lambda_n(\Theta_R, t) - \frac{r_n}{R_s} \Lambda_n(\Theta_R, t) \tag{4.4}$$

where, $R_s = (1 + K_d)$. Equation (4.4) indicates that both the diffusive transport and probe binding process are inhibited due to the equilibrium adsorption and desorption [65]. We assume that the size of the receptor sub-volume is significantly small in comparison to the overall size of the reaction chamber. We will approximate the input to the diffusion channel to be a volumetric point source that injects x_n protein particles during an infinitesimally small time interval (compared to the diffusion time-scales). This can be modeled using an

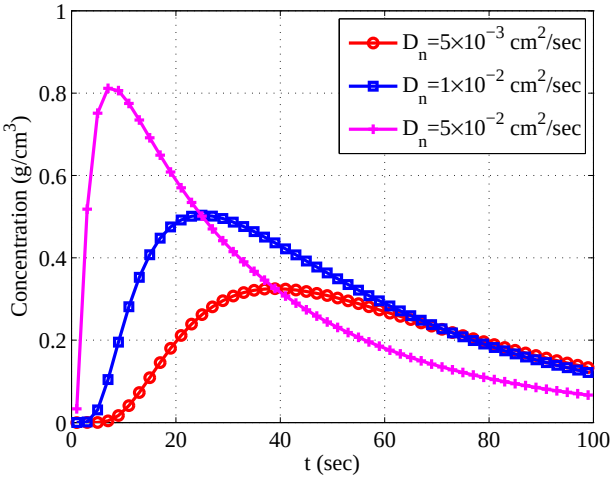


Figure 4.2: Concentration as a function time inside receptor sub-volume for different values of D_n. (Total input concentration $\Lambda_n(\Theta_I, 0) = 4$ g/cm^3, $x_R = 1$ cm, $r_n = 0.02s^{-1}$, $R_s = 2$).

impulse which occurs at the start of the test ($t = 0$) and hence the impulse response based on the equation (4.4) leads to:

$$\Lambda_n(\Theta_R, t) = \frac{x_n}{\sqrt{4\pi D'_n t}} exp\left(\frac{-x_R^2}{4D'_n t} - r'_n t\right). \tag{4.5}$$

$D'_n = D_n/R_s$ and $r'_n = r_n/R_s$ and represent the diffusion and reaction rates adjusted for the inhibition caused by sorption. The diligent observer will note that a 1-D diffusion model has been employed to solve for equation (5). The is justified due to the following reasons: (1) Given a 3-dimensional reference axes shown in Figure 4.1(a), we can assume that the respective concentrations of the receptors and the protein particles in the cross-sectional plane (along the y-axis) are constant and can vary only along the x-axis. A similar assumption was used in [66] where the effect of diffusion on the kinetics of an evanescent wave bionsensor was investigated. The variations along the z-axis can also be ignored under the assumption that the depth of the reaction is shallow. (2) Although, higher-dimensional models could be employed they often do not yield an analytical solution and hence need to be solved using numerical techniques. A simple 1-D model is therefore tractable and preferable.

Figure 4.2 is a plot of equation (4.5) for different values of D_n; and it plots the concentration observed in the receptor sub-volume. For our analysis we are interested in concentration

of protein-n inside the receptor under steady-state conditions. This can be calculated by integrating equation (4.5) over time according to:

$$\begin{aligned}\Lambda_n(\Theta_R) &= \int_0^\infty \frac{x_n}{\sqrt{4\pi D'_n t}} exp\left(\frac{-x_R^2}{4D'_n t} - r'_n t\right) dt \\ &= \frac{x_n}{\sqrt{4D' r'_n}} exp\left(-\sqrt{\frac{r'_n x_R^2}{D'_n}}\right). \end{aligned} \tag{4.6}$$

Due to the Brownian dynamics of particle diffusion the true stead-state concentration Λ_n inside the receptor volume is a random variable whose average value is given by equation (4.6). Under the assumption that the probability of two particles occupying the exact same spatial location is negligible and that the motion of all individual particles inside the reaction chamber is independent of each other; it can be shown that the actual concentration $\Lambda_n(\Theta_R)$ inside V_R is a Poisson random variable, with an arrival rate equal to the average concentration given by $\Lambda_n(\Theta_R)$ [67], [68] :

$$\widetilde{\Lambda}_n(\Theta_R) \sim Poiss(\Lambda_n(\Theta_R)) \tag{4.7}$$

For the rest of the analysis we will not include Θ_R in our expressions with the understanding that Λ_n represents the average concentration of protein-n inside the receptor sub-volume located at Θ_R. We now have a model which we can use to characterize the random variation of concentration inside the receptor sub-volume, V_R, at steady-state.

4.1.2 Receptor Response Model

Deviations from ideal behavior at the receptor are critical for constructing realistic models of the system in question. In this context, a majority of the preliminary investigations into molecular communication systems have focused primarily on non-idealities resulting from the diffusion process, while assuming ideal receptor models (see for example [68], [39]). Only a limited number of studies have addressed this problem. For example, in [69] the impact of sensor cleanse time on the performance of a molecular communication system has

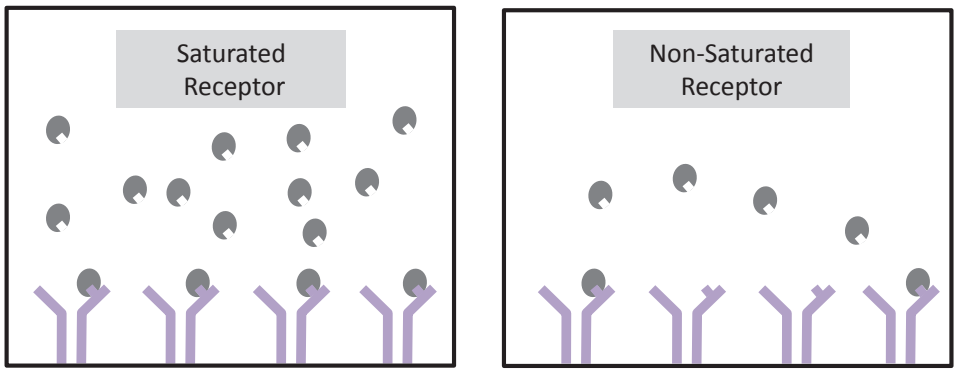


Figure 4.3: Illustration of receptor saturation due to unavailability of free probes.

been investigated. In this section a realistic model based on actual receptor prototypes, constructed in the lab, is presented.

The number of particles inside the receptor sub-volume can be measured using an electrical or an optical transducer that is also immobilized to the probes [35, 70] (for e.g. gold nanoparticles for optical detection or conductive polymer for electrical detection). Since there are only a limited number of epitopes on a receptor, at high concentrations not all protein particles inside V_R will be able to find a vacant epitope to bind with; as illustrated in Figure 4.3. Therefore, at low-to-medium concentrations a change in the transducer's output signal Y varies in direct proportion to a corresponding change in the average concentration of protein-n inside V_R. However, at high concentrations the receptor reaches saturation and the rate of change of Y decreases due to unavailability of free probes as shown in Figure 4.4. Furthermore, at ultra-large protein concentrations a large number of affinity based assays suffer from a phenomenon called the *Hook Effect* [71] which results in a drop in the output with increasing concentration. The Hook Effect occurs beyond the Saturation region and is thought to be the result of factors such as shadowing in which the high density of captured particles prevents the binding of detector probes resulting in a drop in the overall signal. Under the Hook Effect an assay will almost certainly give a faulty reading resulting in a capacity equal to zero; making capacity calculations irrelevant. As a result we assume that the

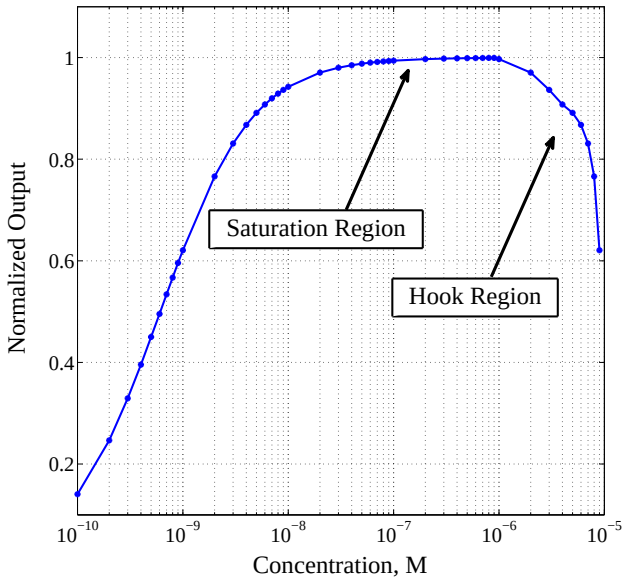


Figure 4.4: Output signal saturation in a typical affinity based array.

input concentration is upper-bounded by a value which is well-below the Hook concentration and do not affect our receptor model. We can now write the rate of change in the average transducer output with respect to a corresponding change in the average concentration inside V_R as [72]:

$$dy/d\Lambda_n = k\mathcal{F}(\Lambda_n) \tag{4.8}$$

where, k is proportionality constant, it models the sensitivity of the receptor to protein-n and is independent of Λ_n. The function $\mathcal{F}(\,)$ incorporates the saturation response; in general it should satisfy the following boundary conditions:

$$\mathcal{F}(\,0\,) = K < \infty \tag{4.9}$$

$$\mathcal{F}(\infty) = 0 \tag{4.10}$$

Furthermore, in order to ensure that the change in magnitude of y reduces as more recognition elements are occupied by incoming particles, the following conditions should also be satisfied:

$$\mathcal{F}(\Lambda) > 0 \tag{4.11}$$

$$\frac{d}{d\Lambda}\mathcal{F}(\Lambda) \leq 0 \tag{4.12}$$

One function satisfying conditions (4.9) to (4.12) is listed below [72]:

$$\mathcal{F}(\Lambda) = \frac{1}{\alpha + \Lambda} \tag{4.13}$$

where, $\alpha > 0$ is a constant and controls the saturation function under control conditions (when no particles are present). Equation (4.13) can now be written in differential form and integrated to give a model for the transducer output:

$$dy = k\frac{d\Lambda}{\alpha + \Lambda} \tag{4.14}$$

$$\int_{y_0}^{y} dy = k\int_{0}^{\Lambda_n} \frac{d\Lambda}{\alpha + \Lambda} \tag{4.15}$$

$$y(\Lambda_n) = y_0 + k\log\left(\frac{\alpha + \Lambda_n}{\alpha}\right) \tag{4.16}$$

Therefore, the output signal of the transducer is a log-linear function of the concentration inside the receptor sub-volume. For the dual-protein combinatorial probe which is constructed by immobilizing recognition elements specific to two different proteins (as shown in Figure 1.4(b)) the output signal will be a function of particles of both proteins. In this case the gradient of the output signal with respect to the concentration of protein-1 and protein-2 will be given by:

$$\frac{\partial y}{\partial \Lambda_1} = k_1\left(\frac{1}{\alpha + \Lambda_1}\right) + k_{12}\left(\frac{1}{\gamma + \Lambda_1 + \Lambda_2}\right) \tag{4.17}$$

$$\frac{\partial y}{\partial \Lambda_2} = k_2\left(\frac{1}{\beta + \Lambda_2}\right) + k_{12}\left(\frac{1}{\gamma + \Lambda_1 + \Lambda_2}\right) \tag{4.18}$$

where, (k_1, α) and (k_2, β) are the model parameters corresponding to protein-1 and protein-2 respectively. The effect of *Joint Hybridization* on the output signal is captured by the parameters (k_{12}, γ). Joint hybridization can be interpreted as the sensitivity of combinatorial probes to the concentrations of both input proteins. In combinatorial probes joint hybridization maybe introduced by design as shown in Figure 4.5. The solution to equations (4.17)

Receptor	Parameter	Description	Value
Specific (Non-Comb)	Y_{m0}	Control conductance (Mouse IgG)	160 μS
	Y_{r0}	Control conductance (Rabbit IgG)	120 μS
	α	Detection Limit (Mouse IgG)	190 μg/ml
	β	Detection Limit (Rabbit IgG)	195 μg/ml
	k_m	Sensitivity (Mouse IgG)	4.4×10^{-3} μS
	k_r	Sensitivity (Rabbit IgG)	4.4×10^{-3} μS
soft-OR	Y_{mr0}	Control conductance (Mouse + Rabbit IgG)	120 μS
	k_m	Sensitivity (Mouse IgG)	-3×10^{-4} μS
	k_r	Sensitivity (Rabbit IgG)	1×10^{-3} μS
	k_{mr}	Sensitivity (Mouse + Rabbit IgG)	5.3×10^{-3} μS
	α	Detection Limit (Mouse IgG)	57.4 μg/ml
	β	Detection Limit (Rabbit IgG)	1000 μg/ml
	γ	Detection Limit (Mouse + Rabbit IgG)	63 μg/ml
soft-AND	Y_{mr0}	Control conductance (Mouse + Rabbit IgG)	110 μS
	k_m	Sensitivity (Mouse IgG)	2.4×10^{-3} μS
	k_r	Sensitivity (Rabbit IgG)	2.7×10^{-3} μS
	k_{mr}	Sensitivity (Mouse + Rabbit IgG)	2.6×10^{-3} μS
	α	Detection Limit (Mouse IgG)	4.6 μg/ml
	β	Detection Limit (Rabbit IgG)	24 μg/ml
	γ	Detection Limit (Mouse + Rabbit IgG)	1200 μg/ml

Table 4.1: Behavioral model parameters for three different types of receptors with Mouse and Rabbit IgG as target analytes [35] [36]. Note: the letters 'm' and 'r', in the subscript, have been employed here (instead of the numerals '1' and '2' in equation (4.19)) to represent Mouse and Rabbit IgG respectively.

and (4.18) is given by:

$$\begin{aligned} y = g(\Lambda_1, \Lambda_2) = y_0 \; &+ \; k_1 \log\left(\frac{\alpha + \Lambda_1}{\alpha}\right) \\ &+ \; k_2 \log\left(\frac{\beta + \Lambda_2}{\beta}\right) \\ &+ \; k_{12} \log\left(\frac{\gamma + \Lambda_1 + \Lambda_2}{\gamma}\right) \end{aligned} \tag{4.19}$$

The log-linear model in equation (4.19) is consistent with the experimental results that have been previously reported [35] for soft-logic receptors (corresponding to rabbit and mouse IgG) shown in Figure 1.4(c). The response of the soft-logic functions remain consistent with the

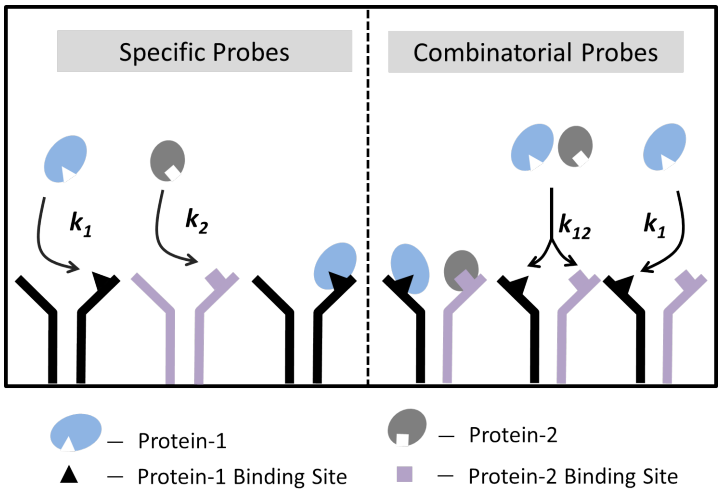


Figure 4.5: Specific and Combinatorial Probes.

presence and the absence of the IgG targets, whereas, the magnitude of the measured output (conductance in the case of [35]) scales log-linearly with the analyte concentration.

Equation (4.19) can also be generalized to more than two proteins and different types of combinatorial circuits such as: (1) “AND” gate which generates a signal when both input proteins are present. (2) “OR" gate which generates a signal when either protein is present. (3) “XOR" gate which generates a signal when one protein is present and the other is absent. Although, traditional protein assays suppress the joint hybridization effect it was shown to be helpful under certain conditions in [73] and [74]. A demonstration of the advantages of exploiting joint hybridization was presented in [35, 37] where FEC codes were constructed using combinatorial probes and in comparison to specific probes, an overall reduction in protein detection error rate was observed. However; up till now probe parameters have been selected experimentally; this is laborious and time consuming and consumes expensive lab material. For instance table 4.1 shows the experimentally determined parameters of the model (4.19) for the rabbit IgG and mouse IgG combinatorial probes (non-combinatorial, soft-AND and soft-OR functions). However, in this paper we are primarily interested in the impact of the different model parameters on the overall capacity, therefore, we vary the different model parameters instead of using fixed values as listed in table 4.1. In this context

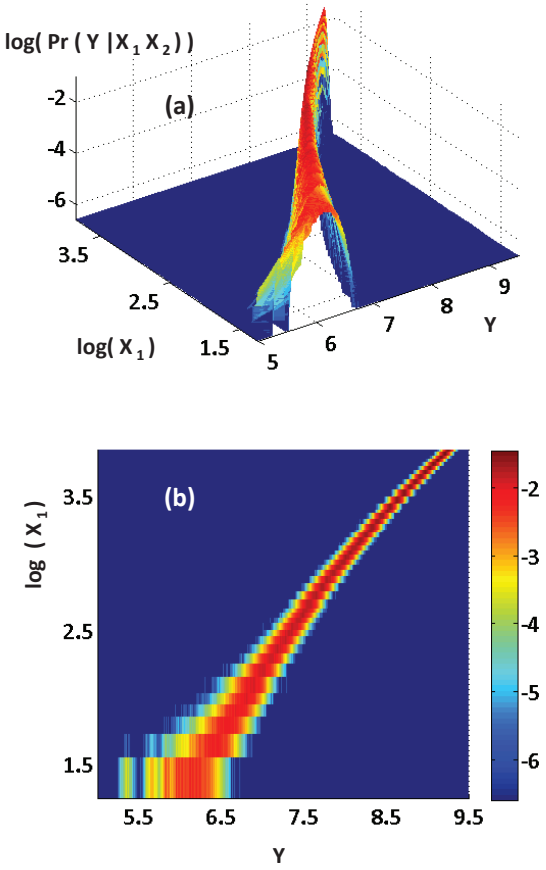


Figure 4.6: Conditional distribution of protein array channel; (a) 3D view (b) Top view. Receptor Parameters are fixed to $k_1 = 1, k_2 = 0.9, k_{12} = 0.9$; Diffusion parameters are as listed in Table 4.3; $x_2 = 1.765 \times 10^3$.

capacity estimation plays an important role and can be used as a tool for selecting model parameters. The optimal probe parameters should in principle correspond to the maximum capacity and therefore the capacity calculation should also provide key insights on designing synthetic probes with the desired hybridization parameters.

4.2 Conditional Distribution of Protein Array Channel

The next step towards determining the information capacity of the proteomic channel is to determine the conditional distribution $P_{Y|X}(y|\mathbf{x})$ where y is the output and $\mathbf{x} = [x_1, x_2]$ is the input vector to the channel. The conditional distribution $P_{Y|X}(y|\mathbf{x})$ of the proteomic channel

Parameter	Value/Range
Y_0	0.5
α	1
β	1
γ	1
k_1	[0,1]
k_2	[0,1]
k_{12}	[0,1]

Table 4.2: Values of Receptor Parameters

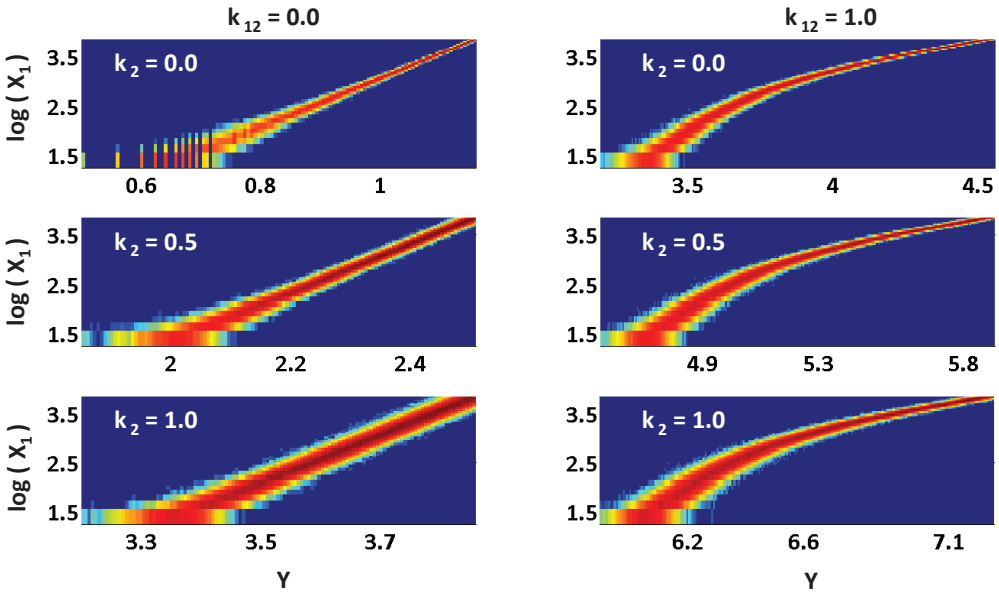


Figure 4.7: Conditional distribution of protein array channel for $k_1 = 0.2$ and $x_2 = 1.765 \times 10^3$. The values of k_2 and k_{12} vary row and column wise respectively.

will capture the effect of the diffusion noise as described in section 4.1.1 and the effect of the receptor response model as described in section 4.1.2. As is the case for any communication channel this conditional distribution can be empirically determined by observing the receptor outputs Y corresponding to a large number of inputs. Our empirical approach will be to use equation (4.6) to determine the steady-state concentration Λ_n corresponding to each protein-variant inside the receptor sub-volume, for different instances of the input particle concentration vector $\mathbf{x} = [x_1, x_2]$. Based on equation (4.7) the noisy values of the receptor concentration will be obtained by sampling a poisson random variable with rate equal to the steady-state concentration Λ_n. These noisy concentration samples are then used to obtain the corresponding values of the transducer output using equation (4.19) which will then be used to evaluate the conditional distribution $P_{Y|X}(y|\mathbf{x})$. The parameters of the receptor response model of equation (4.19) are listed in table-4.2. Since, we do not know the optimal set of parameters and therefore, k_1, k_2 and k_{12} are varied to determine their effect on the conditional distribution and eventually the array capacity. Although a practical array can

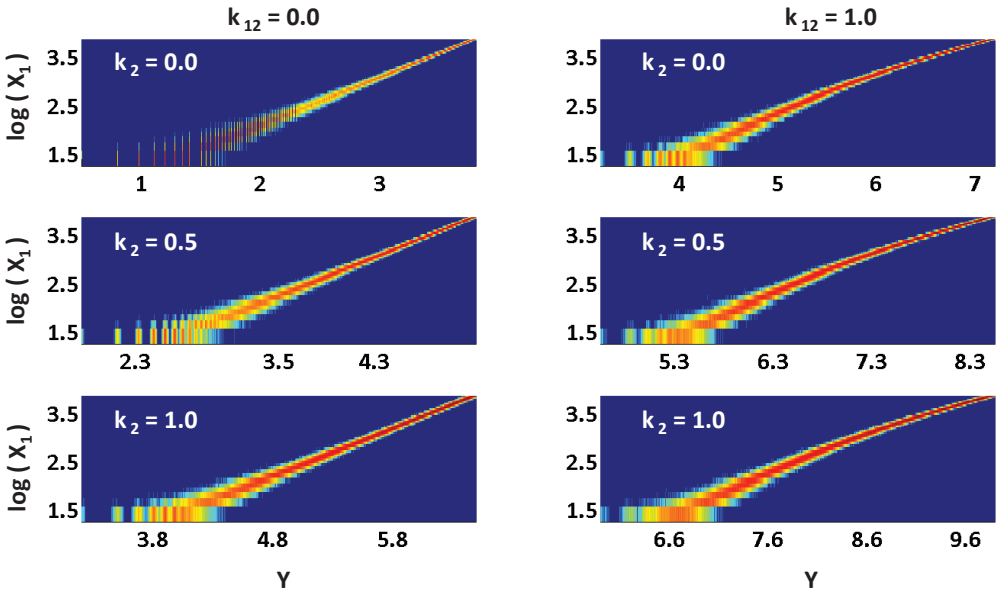


Figure 4.8: Conditional distribution of protein array channel for $k_1 = 1.0$ and $x_2 = 1.765 \times 10^3$. The values of k_2 and k_{12} vary row and column wise respectively.

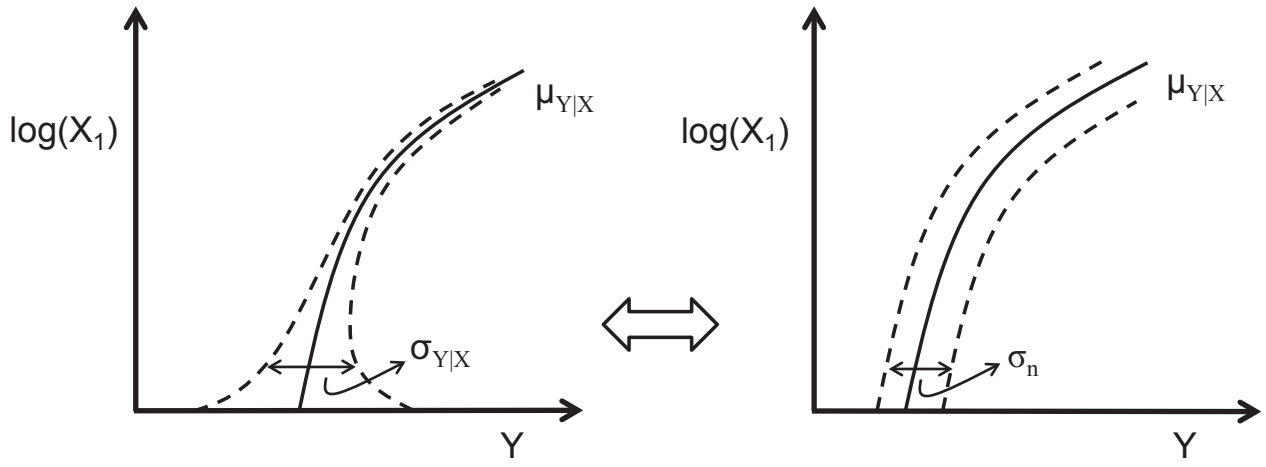


Figure 4.9: Cross sectional view of the conditional distribution $P_{Y|X}(y|\mathbf{x})$ for a fixed x_2 and varying x_1. Conditional variance $\sigma_{Y|X}(y|\mathbf{x})^2$ is approximated by it's average value σ_n^2.

Parameter	Value/Range
D'	$10^{-6}\text{cm}^2\text{s}^{-1}$
r'	$10\ \text{s}^{-1}$
x_R	$2 \times 10^{-3}\text{cm}$

Table 4.3: Diffusion and Reaction Parameters

have sensitivity parameters greater than 1, we limit parameter range between 0 and 1. The diffusion parameters for both the input proteins are assumed to be identical and are listed in table 4.3. Figure 4.6 displays the conditional distribution $P_{Y|X}(y|\mathbf{x})$ (for a fixed value of $x_2 = 1.765 \times 10^3$) from two different viewing angles; the receptor parameters are fixed to $k_1 = 1, k_2 = 0.9, k_{12} = 0.9$ and the diffusion-reaction parameters are as listed in table 4.3.

The effect of receptor parameters on the channel conditional distribution can be visually inspected for some instances of the conditional distribution and are displayed in Figure 4.7 and Figure 4.8. For a receptor with low sensitivity to protein-1 as shown in Figure 4.7 we observe a more pronounced effect on the conditional distribution. In contrast, for a

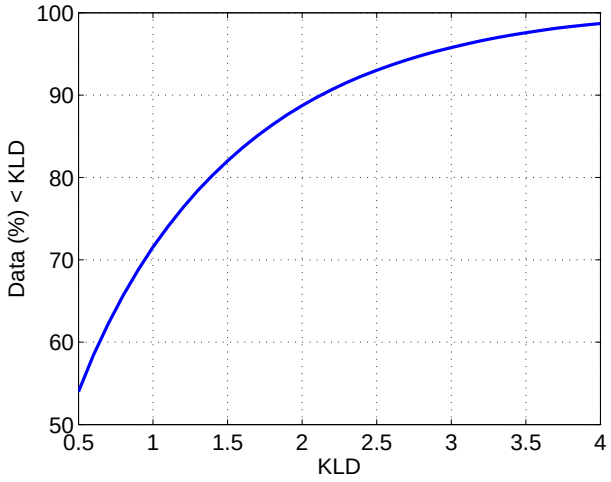


Figure 4.10: KL-Divergence between true and fixed variance distributions.

receptor with high sensitivity to protein-1 (Figure 4.8) the effect of the other two parameters (k_2 and k_{12}) is less pronounced. Plots of the conditional distribution indicate that the noise at the output of the protein array channel is signal dependent. Unfortunately closed form expressions for the capacity of channels with signal dependent noise are too complex to compute or cannot be computed in most scenarios. This problem is further complicated if the channel distribution has a non-stand form, as is the case for a proteomic channel. We therefore have to resort to a numerical approach or use some simplifying assumptions such that capacity expressions corresponding to standard channel distributions can be used. Because the objective of this paper is to determine approximate capacity expressions, we have opted for the analytical approximation based approach. Since we are interested in the impact of receptor parameters on the array capacity, we assume that the variance of the output signal is fixed for all input values and therefore, is independent of the value of the input particles. Thus, we approximate $P_{Y|X}(y|\mathbf{x})$ by a normal distribution with mean given by the true value of y of equation (4.19) and a constant variance σ_n^2 that is independent of the input. This is demonstrated in Figure 4.9 where the dotted line indicates the cross-sectional view of the actual conditional distribution $P_{Y|X}(y|\mathbf{x})$ whose variance $\sigma_{Y|X}(y|\mathbf{x})^2$ is signal dependent and decreases with increasing values of x_1. The highlighted area indicates

the approximate conditional distribution with constant variance σ_n^2 which is equal to the mean value of $\sigma_{Y|X}(y|\mathbf{x})$. In order to capture the dependence on the receptor parameters we compute the noise variance by transmitting a large number of inputs and observing the output y for a fixed set of receptor parameters. Since, the actual value of the variance depends on the value of the input $\mathbf{x}$ therefore the value of the variance corresponding to a fixed set of receptor parameters is obtained by averaging over the variance observed for different values of the input. This process is repeated until we obtain the variance values over the complete range of receptor parameters. The variance σ_n^2 is a function of the receptor parameters and is computed by regression on the average variance of the true conditional distribution that is observed for a range of different receptor parameters. Multivariate polynomial regression results in the following expression for σ_n^2:

$$\begin{aligned}\sigma_n^2 = \Big(& -0.946k_1^2 + 0.711k_2^2 + 27.518k_{12}^2 \\ & - 22.041k_1k_2 + 5.806k_1k_{12} + 4.684k_2k_{12} \\ & + 21.804k_1 + 19.352k_2 - 66.008k_{12} + 58.781) \times 10^{-3}\end{aligned} \tag{4.20}$$

The noise distribution can now be employed to evaluate the capacity of the protein array channel.

The degree of error incurred by applying the constant variance assumption can be gauged by comparing the actual, variable variance, probability distributions with the constant variance probability distribution employed for capacity calculation using a metric such as the Kullback-Leiber (KL) Divergence. Figure 4.10 plots, along the y-axis, the percentage of the observed (variable variance) output distributions whose KL-Divergence with the approximate, constant variance, distribution is less than the KL-Divergence listed along the x-axis. For example, given a KL-Divergence of 0.5 we can see that 55% of the observed distributions have a KL-Divergence of less than 0.5 when compared with the approximate distribution. Similarly, about 70% of the observed distributions are within a KL-Divergence of less than

1 from the constant variance distribution.

4.3 Proteomic Channel Capacity

The approximate conditional distribution can now be used to estimate the information capacity of the proteomic channel. The information capacity of any communication system is the maximum amount of information that can be successfully conveyed from a transmitter to a receiver in a single use [75]. Formally it is defined as the maximum mutual information between the transmitted and the received signal, with maximization performed over all probability distributions defined on the input alphabet. For the protein array communication system with input $\mathbf{x} = [x_1, x_2]$, output Y and given a set (k_1, k_2, k_{12}) of probe parameters we define capacity as:

$$C = \max I(\mathbf{x}; y)|_{(k_1,k_2,k_{12})} \tag{4.21}$$

$$= \max [H(\mathbf{x}) - H(\mathbf{x}|y)] \tag{4.22}$$

$$= \max [H(y) - H(y|\mathbf{x})] \tag{4.23}$$

$I(\mathbf{x}; y)$ represents the mutual information [76]. Capacity is obtained via maximization of the mutual information $I(\mathbf{x}; y)$ between the input and output signals over all possible probability distributions defined on the input alphabet. A combinatorial protein array channel can be viewed as a transform $\Psi : \mathbb{Z}^2_{\geq 0} \rightarrow \mathbb{R}$ that maps the input $\mathbf{x}$ to the output y. The transform Ψ is noisy and models the effect of different noise sources found in a proteomic channel. In section 4.2 it was assumed that the output noise distribution is identical for all possible noise-free (deterministic) outputs y_D therefore, we can replace the noisy map Ψ with a deterministic noise-free transform Ψ_D followed by an AWGN noise model. Equation 4.23 can now be rewritten as:

$$C = \max [H(y) - H(y|y_D)] \tag{4.24}$$

$$= \max I(y; y_D) \tag{4.25}$$

Since the output y is equal to the deterministic transducer output y_D plus gaussian noise therefore, we can employ the expression for the capacity of an additive white noise (AWGN) channel.

The validity of using a Gaussian distribution can be investigated using Goodness-of-Fit (GOF) measures. For this purpose three quantitative GOF metrics namely: (1) *Kolmogorov-Smirnov* (2) *Chi-Squared* and (3) *Anderson-Darling* Test metrics were employed. These metrics enable us to check whether observed samples are generated by a specific distribution (Gaussian in this case). To evaluate each metric we observe the noisy transducer outputs for different instances of the input particle concentration vector $\mathbf{x} = [x_1, x_2]$ in a manner similar to that outlined in section-III. The observed output sample vector, for a given parameter configuration $[k_1, k_2, k_{12}]$ and input concentration instance $\mathbf{x} = [x_1, x_2]$, is considered to have a Gaussian distribution if the GOF test accepts the Null-Hypothesis with a significance value of 1 %. The observed output sample vectors that pass the GOF test are then divided by the total transmitted sample vectors to compute the percentage of observed data that is considered to be Gaussian distributed. This value is averaged over all parameter configurations to obtain the mean value for each of the three metrics. It was observed that for all 3 test metrics, on average, more than 80 % of the observed outputs have a Gaussian distribution. To be specific, 88.5 % of the data passed the GOF test when using the Chi-Squared test; 97.47 % of the data passed using the Kalmogorov-Smirnov test and 81.6 % passed using the Aderson-Darlin test. Therefore, the assumption of the output having a Gaussian distribution is not an unfair one.

The capacity of an AWGN channel with noise variance σ_n^2 and an average input power less than equal to P is given by [75]:

$$C = \frac{1}{2} log \left(1 + \frac{P}{\sigma_n^2}\right) \tag{4.26}$$

For a proteomic channel the equivalent of a power constraint is an upper bound on the variance of the input particles. However, as described in section 4.1.2 the variance of the concentration of the input particles is not under our control. But it is reasonable to bound

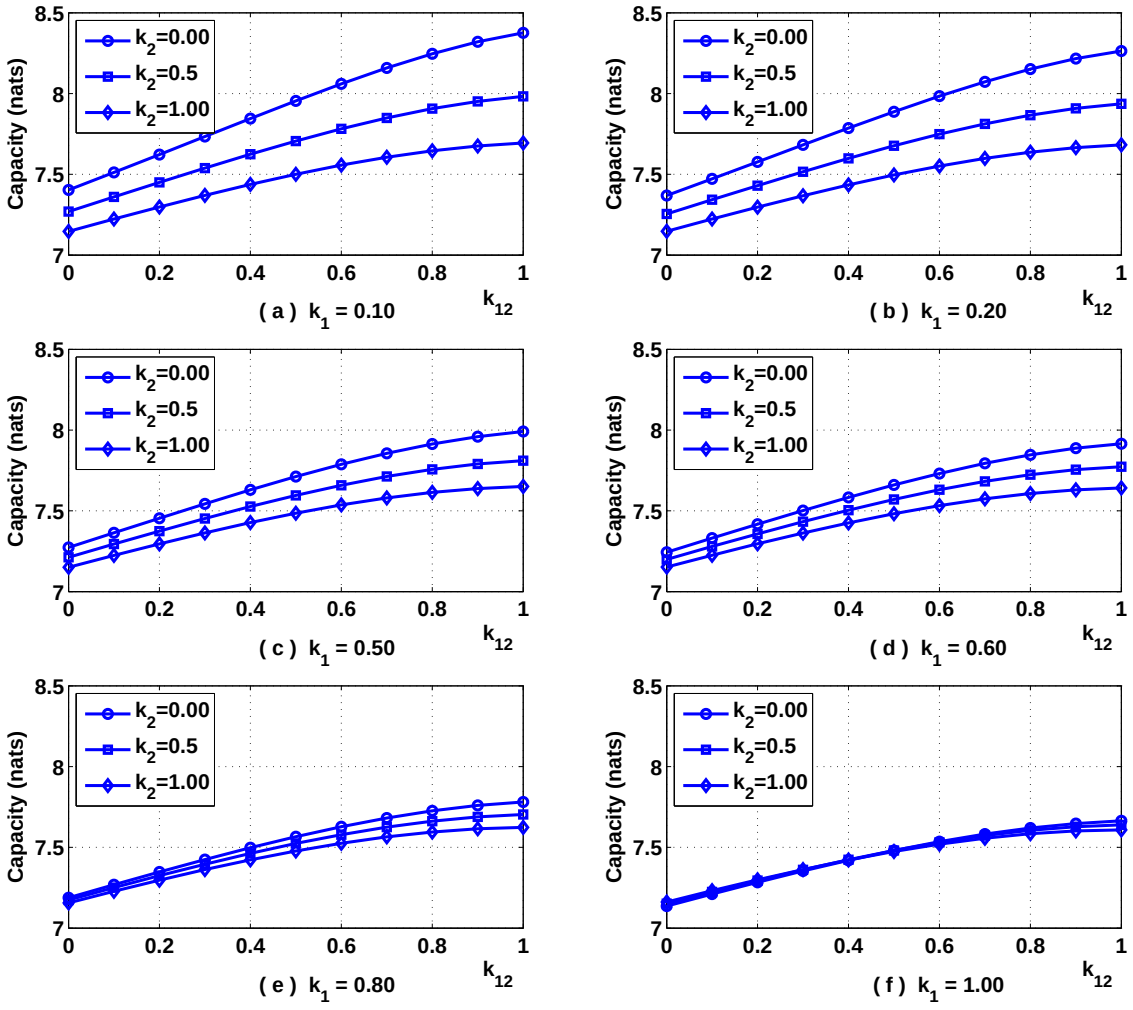


Figure 4.11: Capacity of protein array channel for different values of receptor parameters. Variance P of the input distribution is the same for all settings and is set equal to 10.

(from above) the concentration level of the input particles (by placing an upper bound on the variance of y_D). The capacity of a protein array as a function of receptor parameters can now be computed by substitution of equation (4.20) in to equation (4.26).

Figure 4.11 plots the estimated capacity of protein array channel for different values of receptor parameters. The concentration level (equivalent to power) of the input particle is kept fixed (at the same level) for all the experiments. Because three receptor parameters

were involved in the sweep, each figure in Figure 4.11 corresponds to a fixed value of k_1, which is the receptor sensitivity to protein-1. Due to limited space we only present only a small number of the capacity curves here however, it is highlighted that the trends observed in Figure 4.11 were observed over the entire range of the receptor parameters. It can be observed from the plots that for all parameter settings a higher capacity is achieved when we use a combinatorial receptor that can bind with both proteins simultaneously. Increasing the joint hybridization parameter k_{12} always improves the capacity. At low values of the protein-1 sensitivity parameter, k_1, increasing the sensitivity (k_2) to protein-2, generally results in a decrease in the capacity as can be seen in Figure 4.11 (a) to (c). At higher sensitivities to protein-1 however, the value of the sensitivity parameter of protein-2 does not have a significant impact on the overall capacity. Relative to a specific receptor the highest capacity gains are achieved at lower values of k_1. This can be attributed to the fact that the joint hybridization parameter k_{12} has a more significant impact on the variance at lower values of k_1. For example, by comparing the bottom two plots in Figure 4.7 and Figure 4.8; it is apparent that increasing k_{12} from 0 to 1 results in a much more significant reduction in the overall variance when $k_1 = 0.2$ (Figure 4.7) in comparison to the case where $k_1 = 1$ (Figure 4.8). As the value of k_1 increases the gain in capacity (relative to a specific probe) diminishes however, the loss is not very significant.

CHAPTER 5

KERNEL MACHINES FOR CAPACITY ESTIMATION

The capacity of a dual-protein proteomic channel was computed by approximating it is an additive white Gaussian noise (AWGN) channel in chapter 4. The proteomic channel is a non-linear channel with high-dimensional, continuous input alphabets. Capacity evaluation of such a channel is challenging and generally requires numerical solutions. Furthermore, even when using numerical techniques it often become very challenging to optimize Shannon's information measures such as the mutual information. This chapter presents a framework that employs *Gini* kernel machines to evaluate the (quadratic) mutual information of the proteomic channel. For this purpose a novel proteomic kernel is proposed which incorporates the bio-physics of the receptor and target protein interactions into the optimization problem. Furthermore, it enables array designers to identify the most important probes amongst a large number of candidates. In comparison to the capacity evaluation approach in chapter 4, the framework presented in this chapter considers a large number of input proteins, the approach employed in the previous chapter was limited to only 2 input proteins. Although it can be extended to larger number of proteins it becomes difficult since the number of cross terms (k_{ij}) in the transducer model of equation (4.19) increases significantly as the number of target proteins increase.

The organization of this chapter is as follows. Section 5.1 describes the diffusion model employed. The transducer model is presented in section 5.2. A framework for evaluation of the capacity using *Gini* kernel machines is presented in section 5.3. A novel kernel employed for capacity estimation is proposed in section 5.4.

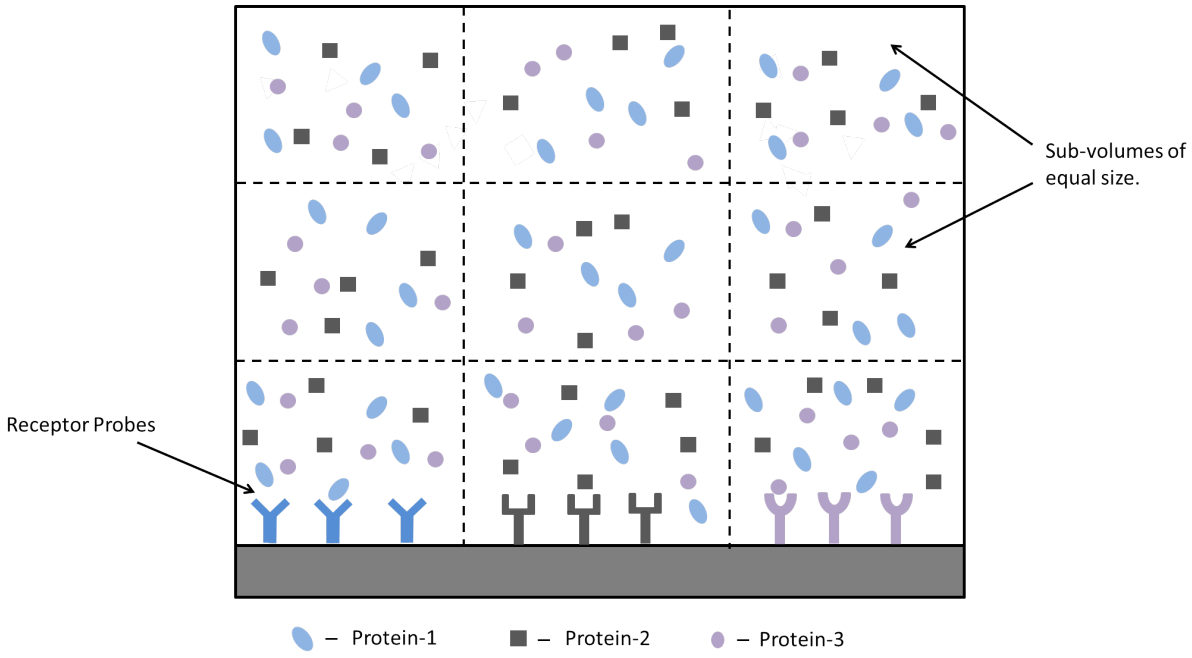


Figure 5.1: Cross-sectional view of diffusion in a multi-protein array.

5.1 Diffusion Model

Analytic solutions for diffusion models containing multiple protein targets in 3-dimensional volumes are often impossible to obtain. Keeping this in perspective this chapter employs a simple diffusion model which views the protein array reaction chamber at stead-state under well-mixed conditions. The physical representation of a simplified three protein assay with multiple receptors input proteins is shown in Figure 5.1. At the start of the test, a sample volume containing particles to be analyzed is injected into the reaction chamber. In order to keep the model mathematically tractable we assume that the reaction volume can be divided into smaller, cubic, sub-volumes of equal size as shown in Figure 5.1 and consider the random variation of particles only inside the sub-volumes containing the receptors located at the bottom of the reaction volume.

For a protein array which performs multiplexed detection of P types of proteins, we denote the input of the system by a vector $\mathbf{u} \in \mathbb{Z}_+^P$. Each component of $\mathbf{u}$ is a non-negative integer random variable denoted by u_n and represents the total number of particles of protein-n in

the test sample. Particles are injected into the reaction chamber using a device such as a micro-pipette. It is assumed that input particles get distributed uniformly throughout all the reaction volume and there are an equal number of protein particles, of each type, inside each sub-volume. The vector $\mathbf{x} \in \mathbb{Z}_+^P$ represents the average number of proteins of each type inside each sub-volume. Therefore, the average number of particles of protein-n inside each sub-volume is given by:

$$x_n = \frac{u_n}{M} \quad n = 1, \ldots P \tag{5.1}$$

where, M represents the total number of sub-volumes inside the reaction volume. Due to the Brownian dynamics of particle diffusion the true number of particles x_n inside sub-volume is a random variable whose average value is given by equation (5.1). Under the assumption that the probability of two particles occupying the exact same spatial location is negligible and that the motion of all individual particles inside the reaction volume is independent of each other; it can be shown that the actual number of particles $\widetilde{x}_n$, of protein-n, inside a sub-volume is a Poisson random variable, with an arrival rate equal to the average number of particles given by x_n [77], [78] :

$$\widetilde{x}_n \sim Poiss(x_n) \quad n = 1, \ldots P \tag{5.2}$$

This model can now be used to characterize the random variation of concentration inside the receptor sub-volumes.

5.2 Receptor Response Model

The receptor response model employed here is an extension of the joint model presented in section 4.1.2. The output of a receptor in a protein array with P different types of input proteins is given by:

$$y = g(\mathbf{x}) = y_0 \; + \; \sum_{i=0}^{P} k_i \log\left(\frac{\alpha + x_i}{\alpha}\right) + \; \sum_{i \neq j} k_{ij} \log\left(\frac{\gamma + x_i + x_j}{\gamma}\right) \tag{5.3}$$

Here, the sensitivity parameters k_i and $k_{ij} \in \mathbb{R}_{\geq 0}$. $k_i = 0$ implies that receptor does not contain probes that interact with particles of protein-i. Whereas, $k_{ij} = 0$ implies that the probes of type-i do not interact with protein-j.

5.3 Proteomic Channel Capacity Estimation

The information capacity of any communication system is defined as the maximum mutual information between the transmitted signal $\mathbf{x}$ and the received signal $\mathbf{y}$ [76]. The maximization is generally performed over all probability distributions defined on the input alphabet.

$$C = \max_{P(\mathbf{x})} I(\mathbf{x}; \mathbf{y}) \tag{5.4}$$

Channel capacity is generally a difficult metric to compute. Analytically its derivation for complex channels can be very challenging (if not impossible) to evaluate. Numerical solutions on the other hand are a more viable option however, classic numerical approaches for capacity evaluation such as the Arimoto-Blahut algorithm [79], [80] are limited to finite input and output alphabets. Although these algorithms have been extended to continuous input and/or output alphabets [81] the evaluation of capacity for continuous channels (such as the proteomic channel) with high-dimensional input-output alphabets is still an open problem. The primary challenge in capacity evaluation of the proteomic channel is accurate estimation of the conditional channel distribution $P(\mathbf{y}|\mathbf{x})$ which is difficult due to the high-dimensional and continuous nature of the input and output alphabets. To elaborate further, a traditional route can be use the empirical conditional channel distribution $\hat{P}(\mathbf{y}|\mathbf{x})$ however, an accurate estimate is difficult to obtain for high-dimensional, continuous channels. As a

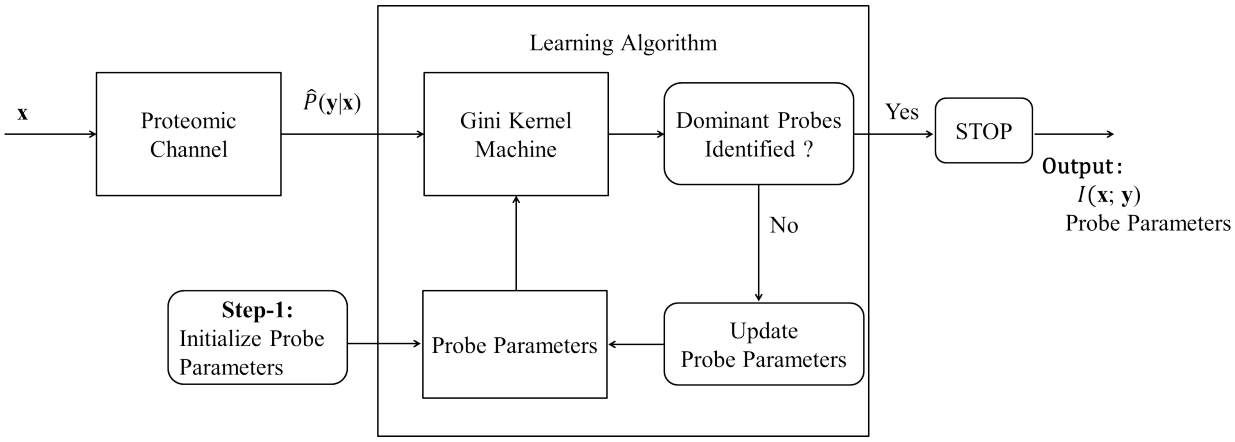


Figure 5.2: Block diagram illustrating the computation of capacity of the proteomic channel.

result we model proteomic channel capacity estimation as a supervised learning problem in which we employ regression to learn the channel conditional distribution, $\tilde{P}(\mathbf{y}|\mathbf{x})$ from a finite set of training examples and then utilize it to evaluate the mutual information which is then maximized to attain capacity. The capacity estimation of the proteomic channel is illustrated in Figure 5.2. It is highlighted that instead of Shannon's mutual information the framework in this chapter employs a quadratic measure of mutual information.

In the framework of supervised learning, the learner is supplied with a training set of feature vectors $\mathcal{T} \subset \mathcal{X} : \mathcal{T} = \{\mathbf{x}_i\}, i = 1, .., N$ drawn independently from a fixed distribution $P(\mathbf{x}), \mathbf{x} \in \mathcal{X}$. In the current scenario the input feature space $\mathcal{X} = \mathbb{Z}_+^P$ and corresponds to the number of particles of each protein type present inside the receptor sub-volumes. Also provided to the learner is a set conditional probability measures $y_{ik} = \hat{P}(y_k|\mathbf{x}_i)$ defined over the set of receptor spots y_k with $k = 1, \ldots S$. The labels therefore are normalized and satisfy $\sum_{k=1}^{S} y_{ik} = 1$. The task of the learner is to choose a set of regression functions $\tilde{P} = \{\tilde{P}(y_k|\mathbf{x})\}, k = 1, .., S$ that accurately predict the true conditional probabilities $P(y_k|\mathbf{x})$ for the receptor spots. This is accomplished by using a distance metric $D_Q : \mathbb{R}^S \times \mathbb{R}^S \to \mathbb{R}$ that embeds prior knowledge about the topology of the feature space. The capacity estimation of the proteomic channel can therefore, be formulated as the maximization of the mutual information subject to the constraint that the distance between the empirical distribution,

$\hat{P}(y_k|\mathbf{x})$ and the estimated distribution $\tilde{P}(y_k|\mathbf{x})$ is less than equal to an error threshold ε:

$$\begin{aligned} \max_{\boldsymbol{\beta}} \quad & I(\mathbf{x};\mathbf{y}) \\ \text{st:} \quad & D_Q\left(\hat{P}(y_k|\mathbf{x}), \tilde{P}(y_k|\mathbf{x})\right) \leq \varepsilon \end{aligned} \tag{5.5}$$

In contrast to the conventional capacity computation the maximization for the proteomic channel is performed over the channel parameters $\boldsymbol{\beta} = [\beta_1, \ldots \beta_P]$ which array designer can vary to embed the maximum amount of input information in the channel output. The parameters $\boldsymbol{\beta} \in \mathbb{R}^P_{\geq 0}$; and determine the importance assigned to each type of probe, they are discussed in detail in section 5.4. The optimization in (5.5) can also be rewritten as:

$$\begin{aligned} \max_{\boldsymbol{\beta}} \quad & D_I\left(\hat{P}\left(y_k|\mathbf{x}\right), P\left(y_k\right)\right) \\ \text{st:} \quad & D_Q\left(\hat{P}(y_k|\mathbf{x}), \tilde{P}(y_k|\mathbf{x})\right) \leq \varepsilon \end{aligned} \tag{5.6}$$

where $D_I : \mathbb{R}^S \times \mathbb{R}^S \rightarrow \mathbb{R}$ represents a distance metric. Although it is possible to employ a metric such as the Kullback-Leibler-Divergence (KLD) it makes the optimization problem difficult and therefore, a quadratic distance will be employed instead. For the proteomic channel it is reasonable to assume that the output distribution $P(y_k) = P_u = U\left[y_0, y_{max}\right]$ for $k = 1, \ldots S$. In other words $P(y_k)$ is uniformly distributed between the control output y_0 and the maximum transducer output under saturation conditions y_{max}. This enables us to reformulate the problem of mutual information estimation as a training procedure involving the minimization of joint distance metric over the probability functions $\tilde{P} = \left\{\tilde{P}(y_k|\mathbf{x})\right\}$

$$\min_{\tilde{P}} G(\tilde{P}) = \min_{\tilde{P}} \left[D_Q(\hat{P}, \tilde{P}) + \gamma D_I(\tilde{P}, P_u)\right] \tag{5.7}$$

In this setting the distance metric $D_I(.,.)$ can be viewed as an agnostic (non-informative) distance measure which assumes no knowledge of the training data. The hyper-parameter $\gamma > 0$ controls the trade-off between the agnostic and prior distance metrics. Minimization of the cost function in (5.7) yields the mutual-information $I(\mathbf{x};\mathbf{y})$ based on the distribution $\tilde{P}(\mathbf{y}|\mathbf{x})$ that lies between the prior distribution $\hat{P}(\mathbf{y}|\mathbf{x})$ and the non-informative (agnostic) distribution P_u. The minimization setup in (5.7) can be coupled with linear constraints

defined on the cumulative statistics of the training set. The first linear constraint expresses equivalence between the average estimated probabilities and empirical frequencies for each receptor over the training data:

$$\sum_{i=1}^{N} \tilde{P}\left(y_k|\mathbf{x}_i\right) = \sum_{i=1}^{N} \hat{P}\left(y_k|\mathbf{x}_i\right), \quad k = 1, \ldots S \tag{5.8}$$

This is based on the assumption that all features $\mathbf{x} \in \mathbb{Z}_+^P$ are equally likely. The normalization and boundary conditions for valid probability distributions are given by a second set of linear constraints:

$$\tilde{P}(y_k|\mathbf{x}) \geq 0, \quad k = 1, \ldots S, \tag{5.9}$$

$$\sum_{k=1}^{M} \tilde{P}(y_k|\mathbf{x}_i) = 1 \tag{5.10}$$

where the normalizing constraint (5.10) subsumes the additional inequality constraint $P_k(\mathbf{x}) \leq 1,\ k = 1, \ldots S$. Combining (5.5) and (5.7) the evaluation of the proteomic channel capacity becomes a *max-min* optimization problem:

$$C_Q = \max_{\boldsymbol{\beta}} \left[\min_{\tilde{P}} \left[D_Q(\hat{P}, \tilde{P}) + \gamma D_I(\tilde{P}, P_u) \right] \right] \tag{5.11}$$

subject to the constraints listed in (5.8), (5.9) and (5.10). Here, capacity is denoted by C_Q to highlight that we are talking about quadratic-capacity and also to differentiate it from the optimization-constant C used in the subsequent discussion.

The minimization step in (5.11) is the same as the optimization problem used section 3.1.1 therefore, the same process can be applied to obtain the quadratic distance D_Q between the conditional distributions $\hat{P}(y_k|\mathbf{x})$ and $\tilde{P}(y_k|\mathbf{x})$.

$$D_Q(\hat{P}, \tilde{P}) = \frac{C}{2} \sum_{k=1}^{S} \sum_{\mathbf{x},\mathbf{v} \in \mathcal{T}} K(\mathbf{x}, \mathbf{v}) \left[\hat{P}(y_k|\mathbf{x}) - \tilde{P}(y_k|\mathbf{x}) \right] \left[\hat{P}(y_k|\mathbf{v}) - \tilde{P}(y_k|\mathbf{v}) \right] \tag{5.12}$$

Here $K : \mathbb{R}^S \times \mathbb{R}^S \to \mathbb{R}$ represents a symmetric, positive definite kernel satisfying the *Mercer's criterion*. Although any standard off-the-shelf kernel such as a Gaussian radial

basis function or a polynomial spline [55, 57] can be employed for optimization we employ a kernel designed specifically for the proteomic channel. This is done because the purpose of the kernel $K(\mathbf{x}, \mathbf{v})$ is to quantify the topology of the metric space for points $\mathbf{x}, \mathbf{v} \in \mathcal{X}$ and therefore; it should, preferably, capture the underlying bio-physics of the proteomic channel.

Use of a *Gini* quadratic distance as D_I, the agnostic distance metric along with a uniform distribution $P_u = 1/(y_{max} - y_0)$ yields the following cost function

$$H_g = \sum_{k=1}^{S} \left[\frac{1}{2C} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_k^i Q_{ij} \lambda_k^j + \frac{\gamma}{2} \sum_{i=1}^{N} (\tilde{P}(y_k|\mathbf{x}_i) - \lambda_k^i / C)^2 \right] \tag{5.13}$$

Where, the inference parameters are defined as $\lambda_k^i = C\left[\hat{P}(y_k|\mathbf{x}_i) - \tilde{P}(y_k|\mathbf{x}_i)\right]$. As was the case in chapter 3 the *Gini* dual is subject to the following constraints

$$\begin{aligned} \sum_{k=1}^{S} \lambda_k^i &= 0, \quad i = 1, \ldots N, \\ \sum_{i=1}^{N} \lambda_k^i &= 0, \quad k = 1, \ldots S, \\ \lambda_k^i &\leq C\hat{P}(y_k|\mathbf{x}_i). \end{aligned} \tag{5.14}$$

The *Gini*-dual in (5.13) is a quadratic function and can be minimized using standard quadratic optimization libraries. The next section introduces the proteomic kernel which can be employed to optimize the receptor parameters.

5.4 Proteomic Kernel

The purpose of the kernel $K(\mathbf{x}, \mathbf{v})$ is to incorporate knowledge of the metric space $\mathcal{X}$ of the input vectors $\mathbf{x}$ and $\mathbf{v}$. In the current context this means that we require a similarity measure which takes into consideration the underlying bio-physics of the receptor and protein interactions. To elaborate further, a conventional *off-the-shelf* kernel returns a high value if its input vectors $\mathbf{x}$ and $\mathbf{v}$ contain similar values and lower value if they are dissimilar. Whereas

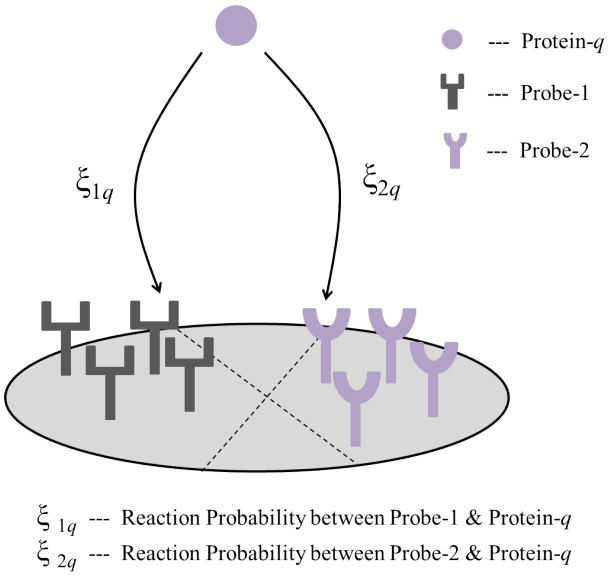


Figure 5.3: Illustration of interactions between protein of type 'i' and two different types of capturing probes .

for the proteomic sensing application we require a kernel which returns a high value if the receptor probes interact with two input concentration vectors in a similar manner.

EXAMPLE: Consider the simple case of an array with 3 protein targets with input vectors $\mathbf{x} = [100, 0, 0]$ and $\mathbf{v} = [100, 0, 50]$. Assume that the receptor under consideration (shown in Figure 5.3) contains only probes that interact with protein-1 and protein-2 but not with protein-3; as a result the output of the receptor will be unaffected by the concentration of protein-3 and will be identical for these values of $\mathbf{x}$ and $\mathbf{v}$. The proteomic kernel should be able to take this into consideration and return a high similarity value when comparing $\mathbf{x}$ and $\mathbf{v}$. If however, the scenario is slightly different and the concentration vector $\mathbf{x} = [100, 0, 0]$ and $\mathbf{v} = [100, 50, 0]$ then the receptor outputs resulting from $\mathbf{x}$ and $\mathbf{v}$ will be different. This should be reflected by a corresponding decrease in the similarity value returned by the proteomic kernel.

The proteomic kernel can be developed by using a product of two sub-kernels as will become clear below. Assuming that within the sub-volume of a combinatorial receptor a particle of protein-q reacts with a probe of type-p with a finite probability ξ_{pq} as illustrated in Figure 5.3. Since there are a maximum of P different types of target proteins, the number

of different types of probes which can be immobilized on a receptor is also $= P$. The total number of particles of protein-q within a receptor sub-volume $= x_q$ (the q-th element of the vector $\mathbf{x}$). Then the binding of the (x_q) particles to the different types of probes can be modeled by a multinomial distribution with P possible outcomes (under the assumption that binding of individual particles is independent of each other). The average number of particles, of protein-q that can be attached to probes of type-p is therefore, given by:

$$\omega_{pq} = x_q \xi_{pq} \tag{5.15}$$

However, the number of probes, immobilized at the receptor, is finite therefore, the ω_{pq} will be bounded from above by the maximum number of probes of type-p (denoted by L_p)

$$\omega_{pq} = min\left(L_p, x_q \xi_{pq}\right) \tag{5.16}$$

We now have a P-dimensional vector $\boldsymbol{\omega}_p$ which tell us the (average) number of particles of each protein type that can be accommodated by the probes of type-p for a given $\mathbf{x}$. The similarity between $\mathbf{x}$ and $\boldsymbol{\omega}_p$ can be computed using radial basis function kernel:

$$K_p(\mathbf{x}, \mathbf{z}_p) = exp\left(-\frac{|\mathbf{x} - \boldsymbol{\omega}_p|^2}{2\sigma^2}\right) \tag{5.17}$$

For a receptor containing P different types of probes we define the proteomic kernel $K(\mathbf{x}, \mathbf{v})$ as below:

$$K(\mathbf{x}, \mathbf{v}) = \sum_{p=1}^{P} \beta_p K_p(\mathbf{x}, \boldsymbol{\omega}_p) K_p(\mathbf{v}, \boldsymbol{\omega}_p) \tag{5.18}$$

where, the parameters β_p controls the importance assigned to probe of type-p. A higher value of β_p indicates that the need to immobilize a large number of probes of type-p at the receptor whereas, a value close to zero implies that probes of type-p should not be employed. In other words the values of β_p is indicative of how much importance should be assigned to probes of type-p whose reaction characteristics are given by the vector $\boldsymbol{\omega}_p$.

Notice that the number of β_p parameters in Equation (5.18) is equal to P. Therefore, the number of optimization variables is only equal to P. This is a significant advantage that

is achieved only due the to use of kernel methods. If the optimization problem had been formulated, without kernel methods, directly in terms of the probe parameters k_1, k_2, k_{12} etc then the number of optimization variables would have been prohibitively large. Consider for example the joint model in equation (5.3), if the optimization problem had been formulated directly in terms of the probe parameters then number of optimization variables would have been equal to $P + P(P-1) = P^2$ due to the large number of cross terms (k_{ij}). In the proteomic kernel however, the cross-terms are present inside the $\boldsymbol{\omega}_p$ and therefore do not need to optimized explicitly.

5.5 Optimization Algorithm

Evaluation of the quadratic capacity in equation (5.11) is performed using an alternating *max-min* procedure. In the first step, probe parameters $\boldsymbol{\beta}_p$ are initialized to have uniform values and minimization is performed using a process identical to the optimization approach employed in chapter 3. After the minimization step, the algorithm fixes the inference parameters λ_i^k and λ_j^k and performs maximization over the probe parameters. This section first incorporates the probe parameters inside the optimization function and then describes the algorithm which can be employed for maximization.

Substituting the proteomic kernel of equation (5.18 into the dual of (5.13) yields a new cost function

$$H_p = \sum_{p=1}^{P} \left[\frac{\beta_p}{2C} \sum_{k=1}^{S} \sum_{i,j=1}^{N} \lambda_k^i Q_{ip} Q_{jp} \lambda_k^j + \frac{\gamma}{2} \sum_{i=1}^{N} (\tilde{P}(y_k|\mathbf{x}_i) - \lambda_k^i/C)^2 \right] \tag{5.19}$$

Where, $Q_{ip} = K(\mathbf{x}_i, \boldsymbol{\omega}_p)$ and $Q_{jp} = K(\mathbf{x}_j, \boldsymbol{\omega}_p)$. Furthermore, in addition to the constraints in (5.14) the proteomic dual (5.19) is subject to the following constraints

$$\begin{aligned} \sum_{p=1}^{P} \beta_p &= 1 \\ \beta_p &\geq 0, \quad p = 1, \ldots P \end{aligned} \tag{5.20}$$

The cost function in (5.19) is a non-homogeneous polynomial and can have both positive and negative coefficients. In addition, the probability variables β_p in (5.19) are normalized $\forall p$. Such a function can be maximized directly by applying results from [82] and [83].

Theorem 2 ([83]) Let $H(\{\beta_p\})$ a polynomial of degree d in variables β_p in the domain $D : \beta_p \geq 0, \sum_{p=1}^{P} \beta_p = 1, p = 1, .., P$. Define an iterative map according to the following recursion

$$\widehat{\beta}_p \leftarrow \frac{\beta_p(\frac{\partial H}{\partial \beta_p}(\beta_p) + \Gamma)}{\sum_{p=1}^{P} \beta_p(\frac{\partial H}{\partial \beta_p}(\beta_p) + \Gamma)} \tag{5.21}$$

where $\Gamma \geq Md(P+1)^{d-1}$ with M being the smallest coefficient of the polynomial $H(\{\beta_p\})$. Then $\{\widehat{\beta}_p\} \in D$ and $H(\{\widehat{\beta}_p\}) > H(\{\beta_p\})$.

The polynomial dual corresponding to Equation (5.21) can be maximized using the result above. Assume that the kernel matrices are bounded such that $|Q_{ip}| \leq Q_{max}, \forall i, p$ and $|Q_{jp}| \leq Q_{max}, \forall j, p$. Furthermore, the initial value of the probability distribution $\beta_p^0 = 1/P$ $\forall p$. Denoting the value of the probability at m^{th} iteration by β_p^m the update at every step will be given by

$$\beta_p^{m+1} \leftarrow \beta_p^m \delta_p^m / \sum_{p=1}^{P} \beta_p^m \delta_p^m$$

where

$$\delta_p^m = \frac{1}{2C} \sum_{k=1}^{S} \sum_{i,j=1}^{N} \lambda_k^i Q_{ip} Q_{jp} \lambda_k^j + \Gamma$$

and $\Gamma = \frac{1}{2C}(P+1)Q_{max}^2$. The cost function in (5.19) increases at each iteration and the process is repeated until convergence. Some of the distribution variables β_p can never reach unity or zero; this is due to the multiplicative update procedure [51]. In practice however, they approach the limits within precision margins that are comparable to other implementations of training algorithms for SVMs. Furthermore, values of the distribution β_p close to unity or zero demonstrate almost no change this implies that caching and shrinking [84] can also be employed to improve the speed of large margin growth transformation.

The framework in this chapter can be employed for evaluating the information transfer across a mutiple-spot protein array channel. Here the information is measured in terms of

a quadratic distance. Validation of this framework using experimental data and numerical simulations shall be performed as part of future work.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Summary

The primary objective of this thesis is to examine the potential benefits that can be achieved via the application of the principles of kernel methods and signal processing in biosensing applications. For respiratory signal estimation it has been demonstrated that use of multiple non-invasive electrodes does enable a significant reduction in breathing rate estimation in comparison to using only one or two electrodes. Furthermore, it seems that use of well-designed learning algorithms results in more performance improvement. In this regard a number of algorithms were tested and it seems that a combination of both signal processing and kernel methods is the best approach. Furthermore, in terms of lung-volume estimation it seems that the SEC does enable effective estimation. Wavelet based features were developed for classification of subject's respiratory state; these features provide a simple yet accurate method for detecting the subject's respiratory state. Wavelet based respiratory state detection outperforms the much simpler DCT based respiratory state detection.

Capacity of the proteomic channel for a small-scale array was evaluated in the presence of diffusion noise and non-ideal receptors. For this array different probe parameters were investigated and it was demonstrated that combinatorial probes give higher capacity as compared to conventional receptor probes. A framework for evaluation of capacity using quadratic information measures was also presented. This framework can be employed for evaluating the capacity of arrays with a significantly higher number of target proteins with ease.

6.2 Future Directions

One of the most appealing future directions that has resulted from this thesis is use of the wavelet based probability curves for classifying not only the subject's respiratory state but also his physical state. As was demonstrated in section 3.2.5 these curves exhibit different characteristics based on the subject's physical activity and therefore, may enable identification of the subject physical state. This in turn may allow the adaptive algorithm to adjust its behavior accordingly e.g. assign lower weightage to chest electrodes and higher weightage to abdominal electrodes if arm motion is detected.

For proteomic channel capacity calculations we intend to validate the framework proposed in chapter 5 by employing numerical simulations and experimental prototypes. Another avenue for future research is to investigate the possibility of employing more complex models of diffusion and receptors when evaluating the channel capacity.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] M. R. Neuman, "Vital signs [tutorial]," *Pulse, IEEE*, vol. 2, no. 1, pp. 39–44, 2011.

[2] WHO. (2014, May) The top 10 causes of death. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs310/en/

[3] NIH. (2013) Explore copd. [Online]. Available: http://www.nhlbi.nih.gov/health/health-topics/topics/copd/

[4] ——. (2014) Explore heart failure. [Online]. Available: http://www.nhlbi.nih.gov/health/health-topics/topics/hf/

[5] ——. (2014, Jun.) Lung function tests. [Online]. Available: http://www.nhlbi.nih.gov/health//dci/Diseases/lft/lft_types.html

[6] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: the neglected vital sign," *Medical Journal of Australia*, vol. 188, no. 11, p. 657, 2008.

[7] T. Moore, "Respiratory assessment in adults," *Nursing standard*, vol. 21, no. 49, pp. 48–58, 2007.

[8] C. Butler-Williams, "Increasing staff awareness of respiratory rate significance," *Resuscitation*, vol. 62, no. 2, pp. 137–141.

[9] N. Shamim, M. Atul, C. Gari D *et al.*, "Data fusion for improved respiration rate estimation," *EURASIP journal on advances in signal processing*, vol. 2010, 2010.

[10] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ecgs," *Computers in cardiology*, vol. 12, pp. 113–116, 1985.

[11] J. A. Hirsch, B. Bishop *et al.*, "Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate," *Am J Physiol*, vol. 241, no. 4, pp. H620–H629, 1981.

[12] S.-B. Park, Y.-S. Noh, S.-J. Park, and H.-R. Yoon, "An improved algorithm for respiration signal extraction from electrocardiogram measured by conductive textile electrodes using instantaneous frequency estimation," *Medical & biological engineering & computing*, vol. 46, no. 2, pp. 147–158, 2008.

[13] C. Orphanidou, S. Fleming, S. Shah, and L. Tarassenko, "Data fusion for estimating respiratory rate from a single-lead ecg," *Biomedical Signal Processing and Control*, vol. 8, no. 1, pp. 98–105, 2013.

[14] C. Voscopoulos, D. Ladd, L. Campana, and E. George, "Non-invasive respiratory volume monitoring to detect apnea in post-operative patients: Case series," *Journal of clinical medicine research*, vol. 6, no. 3, p. 209, 2014.

[15] C. Voscopoulos, J. Brayanov, D. Ladd, M. Lalli, A. Panasyuk, and J. Freeman, "Evaluation of a novel noninvasive respiration monitor providing continuous measurement of minute ventilation in ambulatory subjects in a variety of clinical scenarios," *Anesthesia & Analgesia*, vol. 117, no. 1, pp. 91–100, 2013.

[16] AmericanMedicalAssociation, "Proteomics," Jun. 2013. [Online]. Available: http://www.ama-assn.org//ama/pub/physician-resources/medical-science/genetics-molecular-medicine/current-topics/proteomics.page

[17] R. Huang, B. Burkholder, V. Sloane Jones, W. Jiang, Y. Mao, Q. Chen, and Z. Shi, "Cytokine antibody arrays in biomarker discovery and validation," *Current Proteomics*, vol. 9, no. 1, pp. 55–70, 2012.

[18] H. Akiyama, S. Barger, S. Barnum, B. Bradt, J. Bauer, G. Cole, N. Cooper, P. Eikelenboom, M. Emmerling, B. Fiebich *et al.*, "Inflammation and alzheimerŠs disease," *Neurobiology of aging*, vol. 21, no. 3, pp. 383–421, 2000.

[19] E. Hirsch, S. Hunot *et al.*, "Neuroinflammation in parkinson's disease: a target for neuroprotection?" *Lancet neurology*, vol. 8, no. 4, pp. 382–397, 2009.

[20] A. Mantovani, P. Allavena, A. Sica, and F. Balkwill, "Cancer-related inflammation," *Nature*, vol. 454, no. 7203, pp. 436–444, 2008.

[21] A. Carlsson, C. Wingren, J. Ingvarsson, P. Ellmark, B. Baldertorp, M. Fernö, H. Olsson, and C. A. Borrebaeck, "Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays," *European Journal of Cancer*, vol. 44, no. 3, pp. 472–480, 2008.

[22] A. Vazquez-martin, R. Colomer, and J. A. Menendez, "Protein array technology to detect her2 (< i> erb</i> b-2)-induced Ścytokine signatureŠin breast cancer," *European Journal of Cancer*, vol. 43, no. 7, pp. 1117–1124, 2007.

[23] E. Gorelik, D. P. Landsittel, A. M. Marrangoni, F. Modugno, L. Velikokhatnaya, M. T. Winans, W. L. Bigbee, R. B. Herberman, and A. E. Lokshin, "Multiplexed immunobead-based cytokine profiling for early detection of ovarian cancer," *Cancer Epidemiology Biomarkers & Prevention*, vol. 14, no. 4, pp. 981–987, 2005.

[24] I. Visintin, Z. Feng, G. Longton, D. C. Ward, A. B. Alvero, Y. Lai, J. Tenthorey, A. Leiser, R. Flores-Saaib, H. Yu *et al.*, "Diagnostic markers for early detection of ovarian cancer," *Clinical Cancer Research*, vol. 14, no. 4, pp. 1065–1072, 2008.

[25] R.-P. Huang *et al.*, "Detection of multiple proteins in an antibody-based protein microarray system," *Journal of immunological methods*, vol. 255, no. 1, pp. 1–14, 2001.

[26] S. W. Tam, R. Wiese, S. Lee, J. Gilmore, and K. D. Kumble, "Simultaneous analysis of eight human th1/th2 cytokines using microarrays," *Journal of immunological methods*, vol. 261, no. 1, pp. 157–165, 2002.

[27] P. Oroszlan and M. Ehrat, "ZeptosensŠ protein microarrays: a novel high performance microarray platform for low abundance protein analysis," *Proteomics*, vol. 2, pp. 383–393, 2002.

[28] RayBiotech, "Human angiogenesis array g1 (8) code: Aah-ang-g1-8," http://www.raybiotech.com/g-series-human-angiogenesis-array-g1-8.html, 2004.

[29] ——, "Human angiogenesis array g2 (8) code: Aah-ang-g2-8," http://www.raybiotech.com/g-series-human-angiogenesis-array-g2-8.html, 2004.

[30] ——, "Human angiogenesis array g1000 (8) code: Aah-ang-g1000-8," http://www.raybiotech.com/g-series-human-angiogenesis-array-g1000-8.html, 2005.

[31] S. Sukhanov and P. Delafontaine, "Protein chip-based microarray profiling of oxidized low density lipoprotein-treated cells," *Proteomics*, vol. 5, no. 5, pp. 1274–1280, 2005.

[32] B. Huelseweh, R. Ehricht, and H.-J. Marschall, "A simple and rapid protein array based method for the simultaneous detection of biowarfare agents," *Proteomics*, vol. 6, no. 10, pp. 2972–2981, 2006.

[33] RayBiotech, "Human cytokine array c5 (4) code: Aah-cyt-5-4," https://www.raybiotech.com/c-series-human-cytokine-array-5-4.html, 2009.

[34] R. Huang, W. Jiang, J. Yang, Y. Q. Mao, Y. Zhang, W. Yang, D. Yang, B. Burkholder, R. F. Huang, and R.-P. Huang, "A biotin label-based antibody array for high-content profiling of protein expression," *Cancer Genomics-Proteomics*, vol. 7, no. 3, pp. 129–141, 2010.

[35] Y. Liu, M. Gu, E. Alocilja, and S. Chakrabartty, "Co-detection: Ultra-reliable nanoparticle-based electrical detection of biomolecules in the presence of large background interference," *Biosensors and Bioelectronics*, vol. 26, no. 3, pp. 1087–1092, 2010.

[36] Y. Liu, D. Zhang, E. C. Alocilja, and S. Chakrabartty, "Biomolecules detection using a silver-enhanced gold nanoparticle-based biochip," *Nanoscale research letters*, vol. 5, no. 3, pp. 533–538, 2010.

[37] Y. Liu and S. Chakrabartty, "Factor graph-based biomolecular circuit analysis for designing forward error correcting biosensors," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 3, no. 3, pp. 150–159, 2009.

[38] A. Hassibi, H. Vikalo, and A. Hajimiri, "On noise processes and limits of performance in biosensors," *Journal of applied physics*, vol. 102, no. 1, pp. 014909–014909, 2007.

[39] M. Pierobon and I. F. Akyildiz, "Capacity of a diffusion-based molecular communication system with channel memory and molecular noise," *Information Theory, IEEE Transactions on*, vol. 59, no. 2, pp. 942–954, 2013.

[40] I. Smith, J. Mackay, N. Fahrid, and D. Krucheck, "Respiratory rate measurement: a comparison of methods," *British Journal of Healthcare Assistants*, vol. 5, no. 1, p. 18, 2011.

[41] E. P. Scilingo, A. Lanata, and A. Tognetti, "Sensors for wearable systems," in *Wearable Monitoring Systems*. Springer, 2011, pp. 3–25.

[42] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

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

[44] I. Guyon, B. Boser, and V. Vapnik, "Automatic capacity tuning of very large vc-dimension classifiers," *Advances in neural information processing systems*, pp. 147–147, 1993.

[45] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[46] J.-L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," *Speech and audio processing, ieee transactions on*, vol. 2, no. 2, pp. 291–298, 1994.

[47] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital signal processing*, vol. 10, no. 1, pp. 19–41, 2000.

[48] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. springer New York, 2006, vol. 1.

[49] S. Lloyd, "Least squares quantization in pcm," *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 129–137, 1982.

[50] A. P. Dempster, N. M. Laird, D. B. Rubin *et al.*, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[51] S. Chakrabartty and G. Cauwenberghs, "Gini support vector machine: Quadratic entropy based robust multi-class probability regression," *The Journal of Machine Learning Research*, vol. 8, pp. 813–839, 2007.

[52] E. T. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, p. 620, 1957.

[53] T. Jebara, "Discriminative, generative and imitative learning," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[54] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.

[55] B. Schölkopf, C. Burges, and A. Smola, "Advances in kernel methodsŮsupport vector learningmit press," *Cambridge, MA*, 1999.

[56] M. I. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the em algorithm," *Neural computation*, vol. 6, no. 2, pp. 181–214, 1994.

[57] G. Wahba *et al.*, "Support vector machines, reproducing kernel hilbert spaces and the randomized gacv," *Advances in Kernel Methods-Support Vector Learning*, vol. 6, pp. 69–87, 1999.

[58] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on.* IEEE, 1997, pp. 130–136.

[59] J. Platt *et al.*, "Fast training of support vector machines using sequential minimal optimization," *Advances in kernel methodsŮsupport vector learning*, vol. 3, 1999.

[60] T. Poggio and G. Cauwenberghs, "Incremental and decremental support vector machine learning," *Advances in neural information processing systems*, vol. 13, p. 409, 2001.

[61] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, 1989.

[62] I. Daubechies *et al.*, *Ten lectures on wavelets.* SIAM, 1992, vol. 61.

[63] D. S. Lemons and A. Gythiel, "Paul langevinŠs 1908 paper Şon the theory of brownian motionŤ[Şsur la théorie du mouvement brownien,Ť cr acad. sci.(paris)[bold 146], 530–533 (1908)]," *American Journal of Physics*, vol. 65, p. 1079, 1997.

[64] P. F. Green, *Kinetics and Transport in Soft and Hard Materials.* Taylor & Francis Group, 2005.

[65] J. S. Gulliver, *Introduction to chemical transport in the environment.* Cambridge University Press, 2007.

[66] P. Schuck, "Kinetics of ligand binding to receptor immobilized in a polymer matrix, as detected with an evanescent wave biosensor. i. a computer simulation of the influence of mass transport," *Biophysical journal*, vol. 70, no. 3, pp. 1230–1249, 1996.

[67] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes.* Tata McGraw-Hill Education, 2002.

[68] M. Pierobon and I. F. Akyildiz, "Diffusion-based noise analysis for molecular communication in nanonetworks," *Signal Processing, IEEE Transactions on*, vol. 59, no. 6, pp. 2532–2547, 2011.

[69] S. Wang, W. Guo, S. Qiu, and M. D. McDonnell, "Performance of macro-scale molecular communications with sensor cleanse time," in *Telecommunications (ICT), 2014 21st International Conference on.* IEEE, 2014, pp. 363–368.

[70] Y. Liu, S. Chakrabartty, and E. Alocilja, "Fundamental building blocks for molecular biowire based forward error-correcting biosensors," *Nanotechnology*, vol. 18, no. 42, p. 424017, 2007.

[71] R. G. Ryall, C. J. Story, and D. R. Turner, "Reappraisal of the causes of the Ş̆hook effectŤ in two-site immunoradiometric assays," *Analytical Biochemistry*, vol. 127, no. 2, pp. 308–315, 1982.

[72] M. Gu and S. Chakrabartty, "Fast: A framework for simulation and analysis of large-scale protein-silicon biosensor circuits," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 7, no. 4, pp. 451–459, 2013.

[73] H. Vikalo, B. Hassibi, and A. Hassibi, "A statistical model for microarrays, optimal estimation algorithms, and limits of performance," *Signal Processing, IEEE Transactions on*, vol. 54, no. 6, pp. 2444–2455, 2006.

[74] ——, "On limits of performance of dna microarrays," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 2. IEEE, 2006, pp. II–II.

[75] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[76] T. Cover and J. Thomas, *Elements of information theory*. Wiley-interscience, 2006.

[77] S. Chandrasekhar, "Stochastic problems in physics and astronomy," *Reviews of modern physics*, vol. 15, no. 1, p. 1, 1943.

[78] M. Höller, "Advanced fluorescence fluctuation spectroscopy with pulsed interleaved excitation," Ph.D. dissertation, lmu, 2011.

[79] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *Information Theory, IEEE Transactions on*, vol. 18, no. 1, pp. 14–20, 1972.

[80] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *Information Theory, IEEE Transactions on*, vol. 18, no. 4, pp. 460–473, 1972.

[81] J. Dauwels, "Numerical computation of the capacity of continuous memoryless channels," in *Proceedings of the 26th Symposium on Information Theory in the BENELUX*. Citeseer, 2005, pp. 221–228.

[82] L. E. Baum, G. R. Sell *et al.*, "Growth transformations for functions on manifolds," *Pacific J. Math*, vol. 27, no. 2, pp. 211–227, 1968.

[83] P. Gopalakrishnan, D. Kanevsky, A. Nadas, and D. Nahamoo, "An inequality for rational functions with applications to some statistical estimation problems," *Information Theory, IEEE Transactions on*, vol. 37, no. 1, pp. 107–113, Jan 1991.

[84] T. Joachims, *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.