

1-1-2009

Analysis of electromyogram in rapid eye movement sleep

Mehrnaz Shokrollahi
Ryerson University

Follow this and additional works at: <http://digitalcommons.ryerson.ca/dissertations>



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Shokrollahi, Mehrnaz, "Analysis of electromyogram in rapid eye movement sleep" (2009). *Theses and dissertations*. Paper 946.

This Thesis is brought to you for free and open access by Digital Commons @ Ryerson. It has been accepted for inclusion in Theses and dissertations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact bcameron@ryerson.ca.

619589104
R
857
547
556
2009

ANALYSIS OF ELECTROMYOGRAM IN RAPID EYE MOVEMENT SLEEP

Mehrnaz Shokrollahi, B. Eng
Ryerson University, Toronto, 2007

A thesis
presented to Ryerson University
in partial fulfillment of the
requirements for the degree of
Master of Applied Science
in the program of
Electrical and Computer Engineering

Toronto, Ontario, Canada, 2009

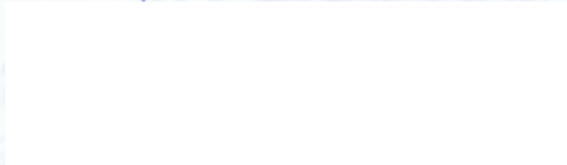
©Mehrnaz Shokrollahi, 2009

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature



I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature



Abstract

ANALYSIS OF ELECTROMYOGRAM IN RAPID EYE MOVEMENT SLEEP

©Mehrnaz Shokrollahi, 2009

**Masters of Applied Science
Electrical and Computer Engineering
Ryerson University**

The aim of this study is to analyze Electromyogram (EMG) signals in Rapid Eye Movement (REM) sleep using different techniques to detect the level of normality and abnormality of normal and abnormal (patients with a lack of REM sleep atonia) subjects and predict the development of Parkinson's disease in abnormal subjects.

Quantitative electromyography (EMG) signal analysis in the frequency domain using classical power spectrum analysis techniques have been well documented over the past decade. Yet none of these work have been done on EMG during Rapid Eye Movement (REM) Stage of sleep.

In this work three techniques for classifying chin movement via EMG signals during sleep is presented. Three methods (Autoregressive modeling, Cepstrum Analysis and Wavelet Analysis) for extracting features from EMG signal during sleep and a classification algorithm (Linear Discriminant Analysis (LDA)) were analyzed and compared.

EMG data are used to detect and describe different disease processes affecting sleep. Rapid Eye Movement Behavior Disorder (RBD) is an example of EMG abnormality in which patients lose their muscle control while in REM stage of sleep resulting in physically acting out their dreams. An adaptive segmentation based on Recursive Least Square (RLS) algorithm was analyzed. This algorithm was used to segment the non-stationary EMG signal into locally stationary components, which were then autoregressive modeled using the Burg-Lattice method. The cepstral measurements described was used and applied to modify the coefficients computed from the autoregressive (AR) model. Yet due to the nature of the EMG, frequency analysis cannot be used to approximate a signal whose properties change over time. To address this problem a time varying feature representation is necessary for analysis to extract useful information from the signal. As a consequence Wavelet coefficients were computed using discrete and continuous wavelet transforms.

Furthermore, the classification performance of the above three feature sets were investigated for the two classes (Normal and Abnormal). Results showed wavelet analysis compared to AR modeling and cepstrum analysis is a better assessment in finding EMG abnormalities during sleep. However, these methods may be useful in distinguishing EMG patterns that predict the emergence of Parkinson disease in humans.

Acknowledgment

I would like to acknowledge my supervisor Dr. Sridhar Krishnan who has been a strong source of inspiration throughout my project work. I have benefited greatly from his invaluable guidance and motivation. His guidance and moral support has helped me to succeed beyond my wildest dreams. I am forever grateful to him for this.

I would also like to thank my parents and family for their continual support and encouragement, which has given me the ability to pursue my goals with confidence and full determination. I would also like to express my special thanks to my sister, Elnaz Shokrollahi, who has also been my long term friend as well as a colleague throughout my student life at Ryerson University for both Undergraduate and Graduate studies. Her motivation, support and encouragement gave me the strength to peruse my goals and dreams. I am forever delighted and thankful for having her in my life. I also would like to thank my friend Saeed Gharnagh for all his support. His love and encouragement motivated me to continue my graduate studies in Ryerson.

I would also like to express my profound gratitude to Dr. Brain Murray for his guidance. He devoted considerable time and effort into explaining the clinical part of this project, providing relevant materials, and gave this project a definite practical applicability. I appreciate Mr. Dana Jewell's assistance in data collection and sharing his experience and knowledge with me.

I also would like to thank my fellow colleagues from the Signal Analysis Research (SAR) group, specially Peyman Shokrollahi, Dr. Karthi Umapathy and Nasim Shams for their enlightening and informative discussions they had with me about my research.

Last but not least, I would like to thank everyone at Ryerson University who provided me with an enhanced educational experience, in which I will never forget.

Dedication

To my parents Shohreh Kamali and Hadi Shokrollahi, my brother Peyman Shokrollahi, my sister Elnaz Shokrollahi, and Saeed Gharnagh for their love, support and encouragement.

Contents

1	Introduction	1
1.1	Biomedical Signals	3
1.1.1	EEG	3
1.1.2	EOG	4
1.1.3	EMG	4
1.2	Sleep	5
1.2.1	Sleep Disorder	8
1.2.2	EMG and RBD	10
1.2.3	Parkinson's Disease	11
1.2.4	Data Acquisition	12
1.2.5	Dataset Type	12
1.3	Signal Analysis	13
1.3.1	Signal Behavior	13
1.4	Motivation	15
1.5	Organization of the thesis	15
2	Signal Analysis	18
2.1	Literature Review	18
2.2	Adaptive Signal Processing	19
2.3	Recursive Least Square Algorithm	22
2.3.1	The concept of Adaptive Learning Curve	27
2.4	Linear Prediction	28
2.4.1	Time Series Modeling	30
2.4.2	AR Modeling	31
2.4.3	Cepstrum Analysis	32
2.5	Wavelet Transform	34
2.5.1	Continuous Wavelet Transform	38
2.5.2	Discrete Wavelet Transform	39
2.6	Classification	41
2.6.1	Linear Discriminant Function	41
2.6.2	Linear Discriminant Analysis	42
2.7	Motivation Behind this Study	44

3	Parametric Signal Analysis of EMG in Sleep	46
3.1	Introduction	46
3.2	Adaptive Signal Processing	47
3.2.1	Recursive Least Square Algorithm	48
3.3	AR Modeling	50
3.4	Cepstrum Analysis	56
3.5	Time-Series Analysis of Sleep EMG	57
3.5.1	Results	58
3.5.2	Classification Result	58
3.6	Discussion	65
4	Wavelet Analysis of EMG in Sleep	68
4.1	Introduction	68
4.2	Motivation Behind the Study	69
4.3	Discrete Wavelet Transform	70
4.3.1	Wavelet Decomposition	72
4.4	Continuous Wavelet Transform	72
4.4.1	Scaling Factor	74
4.4.2	Mother Wavelet	75
4.5	Wavelet Analysis of Sleep EMG	77
4.5.1	Algorithm	77
4.5.2	Results	82
4.5.3	Classification Result	83
4.6	Discussion	89
5	Conclusions	93
5.1	Sleep Related Problems	93
5.2	Parametric Signal Analysis of EMG in Sleep	94
5.3	Wavelet Analysis of EMG in Sleep	95
5.4	Future Work	96
	Bibliography	101

List of Figures

1.1	Annual Death in North America [3]	2
1.2	EEG waves [7]	4
1.3	Normal chin EMG signal in Sleep	5
1.4	Abnormal chin EMG signal in Sleep	6
1.5	Normal EMG signal in REM sleep	10
1.6	Abnormal EMG signal in REM sleep	11
1.7	Organization of the thesis	17
2.1	The Adaptive Learning Curve: Behavior of the performance function J with time	29
2.2	The Adaptive Learning Curve: Behavior of the performance function J with time	29
2.3	Fourier Basis Functions, time frequency tiles, and coverage of the time-frequency plane	37
2.4	Daubechies Wavelet Basis Functions, time frequency tiles, and coverage of the time-frequency plane	37
2.5	1 Level Decomposition Level for DWT	40
2.6	An example of between-class scatter over the within-class scatter in LDA . .	43
2.7	The Proposed Method	45
3.1	Block Diagram of the AR modeling and Cepstrum Analysis	47
3.2	Block Diagram of RLS Algorithm	51
3.3	Techniques for Parametric modeling, using a rational transfer function and algorithms for autoregressive parameter estimation	52
3.4	Finding the optimum model order, using the AIC criterion	55
3.5	Model spectrum of the Normal Signal	55
3.6	Model spectrum of the Abnormal Signal	56
3.7	The level of Normality in Normal Subjects using AR modeling	60
3.8	The level of Abnormality in Abnormal Subjects using AR modeling	60
3.9	The level of Normality in Normal Subjects using Cepstrum	61
3.10	The level of Abnormality in Abnormal Subjects using Cepstrum	61
3.11	The Separation of Dominant AR coefficients	63
3.12	The Separation of Dominant Cepstral coefficients	63

3.13 Receiver Operating Characteristic of the AR modeling	65
3.14 Receiver Operating Characteristic of the Cepstrum Analysis	66
4.1 Block Diagram of the Wavelet Analysis	70
4.2 1-Level Wavelet Decomposition	72
4.3 Wavelet Decomposition Tree	73
4.4 Scaling of Wavelet	75
4.5 Daubechies Wavelet	76
4.6 1 Level Decomposition EMG of Normal Segment	78
4.7 1 Level Decomposition EMG of Abnormal Segment	79
4.8 CWT of Normal Signal for Scale 15 to 100	79
4.9 CWT of Abnormal Signal for Scale 15 to 100	80
4.10 CWT of Normal Signal for Scale 150 to 300	81
4.11 CWT of Abnormal Signal for Scale 150 to 300	81
4.12 Distribution of Normality in Normal subjects	86
4.13 Distribution of Abnormality in Abnormal subjects.	86
4.14 Distribution of Normality in Normal subjects for Scale 150-300.	87
4.15 Distribution of Abnormality in Abnormal subjects for Scale 150-300.	87
4.16 The Separation of Dominant WT coefficients	88
4.17 ROC for classification of EMG in REM sleep using WT for Scale 15 -100 . .	90
4.18 ROC for classification of EMG in REM sleep using WT for Scale 150 -300 .	91
4.19 ROC for classification of EMG in REM Sleep	92

List of Tables

2.1	Related works in fixed and adaptive segmentation for biomedical signals . . .	21
3.1	Confusion matrix containing the number of correct classified EMG in Sleep as either normal or abnormal using Discriminant classification and AR modeling	62
3.2	Confusion matrix containing the number of correct classified EMG in Sleep as either normal or abnormal using Leave-one-out classification and AR modeling	62
3.3	Confusion matrix containing the number of correct classified EMG in Sleep as either normal or abnormal using Discriminant classification and Cepstrum Analysis	64
3.4	Confusion matrix containing the number of correct classified EMG in Sleep as either normal or abnormal using Leave-one-out classification and Cepstrum Analysis	64
4.1	Related works in Wavelet Transform	71
4.2	Confusion matrix: containing the number of correct classified EMG in Sleep, as either normal or abnormal, using Discriminant classification and WT for Scale 15 to 100.	83
4.3	Confusion matrix: containing the number of correct classified EMG in Sleep, as either normal or abnormal, using Leave-one-out method classification and WT for Scale 15 to 100.	84
4.4	Confusion matrix: containing the number of correct classified EMG in Sleep, as either normal or abnormal, using LDA classification and wavelet WT for Scale 150 to 300.	85
4.5	Confusion matrix: containing the number of correct classified EMG in Sleep, as either normal or abnormal, using Leave-one-out method classification and WT for Scale 150 to 300.	85

Chapter 1

Introduction

SLEEP and sleep-related problems play a role in a large number of human disorders, and affect every field of medicine. Clinicians and other healthcare professionals receive extensive training in order to be sufficiently qualified to detect and prevent diseases. Although the skills acquired by these medical facilitators are quite extensive, it is just as important for them to have access to an assortment of technologies, and to further improve their monitoring and treatment capabilities. In fact, these approaches may provide useful information to clinicians in the form of an easily applicable measure of disease treatment, which is sensitive to early neurodegenerations and treatment responses.

Electrodes and signal acquisition technology can be used to gather a variety of biomedical signals such as electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG), which records the electrical activities of the heart, brain, eyes, and muscles respectively. The analysis of these signals can show the physiological behavior of an organ or set of organs based on a quantity, which varies over time [1]. The information that these signals carry is significant in the development of better human understanding to further improve the healthcare and the quality of life of individuals.

The science of health in society has improved in the past decade, yet about 170,000 to 250,000 people die from a sudden cardiac arrest, 14.5 of every 1000 people die from severe sleep apnea, and the rate of cancer death in Canada in the year of 2008 was estimated to be 73,000 people. In addition, statistics concerning the 10 year mortality data from the Sydney

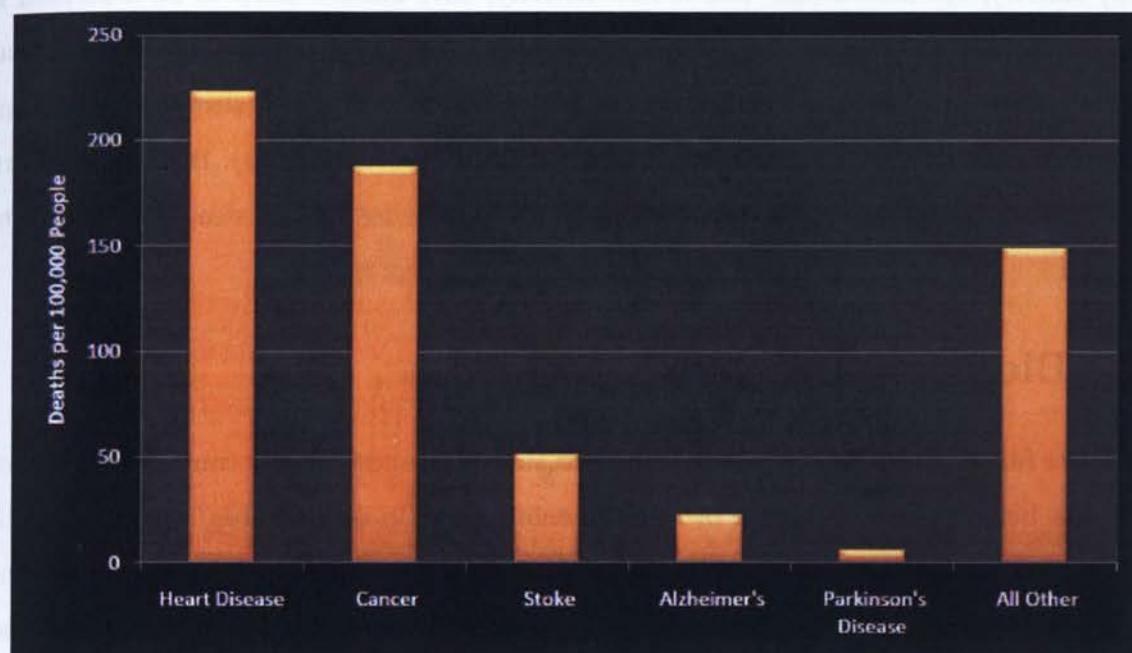


Figure 1.1: Annual Death in North America [3]

multicentric study of Parkinson's disease shows that patients with this condition still die at a rate in excess of their peers despite advances in therapeutics and surgery [2]. These diseases such as Severe Sleep Apnea or Parkinson's could cause traffic accidents, sudden cardiac arrest, or, reduce the quality of life. Therefore, society is still facing the most serious challenges of the world, in overcoming these fatal diseases. In this regard an extensive research is required for diagnostic and therapeutic of many diseases. Figure 1.1 shows the Annual Deaths in USA and Canada [3]. Sleep is, closely, related to every facet of the daily life. In this respect, disturbed sleep affects not only the health and wellbeing of individuals, but also their quality of life [4]. According to National Sleep Foundation (NSF), millions of people suffer from lack of proper sleep. NSF has shown that at least 40 million Americans suffer from over 70 different sleep disorders, which in turn, directly affects their quality of life [5]. The introduction of a number of new techniques, during the past few decades, including polysomnographic (sleep study), surface measurements of central nervous system

(CNS) activity, eye movements, and muscle activity, have allowed sleep to be described in electrophysiological terms [6]. Studying sleep behavior is very important, since it affects all humans. Therefore, this work is mainly dedicated to the analysis of sleep and to a type of disorder, called Rapid eye movement sleep Behavior Disorder (RBD). Before explaining the sleep macroarchitecture, the type of signals that are being used in sleep studies will be explained.

1.1 Biomedical Signals

Signals are functions of one or more independent variables, and typically contain information about the behavior or the nature of some phenomenon. In other words, systems usually respond to particular signals by producing other signals. Moreover, medical signals are very important and, are widely used in medicine to predict the disease, diagnosis of the disease, understand the effect of medicine, as well as response to therapy. As explained before, signals such as EEG, EOG, and EMG could be used to detect many sleep disorders, and in fact these signals are recorded in sleep laboratories from the subjects suffering from a sleep disorder, and further analyzed by sleep experts. The properties of EEG, EOG, and EMG signals help the sleep technicians score the various stages of sleep and develop a hypnogram – a graph of the sleep stages over time. The hypnogram reveals the macroarchitecture of sleep, by characterizing the alternation of different sleep stages. Each sleep staging decision is based on a 30-second window of the physiological signals, called an Epoch.

1.1.1 EEG

The EEG is a recording of the electrical activity of the brain, from the scalp. This recorded waveforms reflect the cortical electrical activity of the brain [1]. Figure 1.2 shows the main frequencies of the human EEG waves [7]. The different frequency types of the EEG wave has been used extensively in polysomnography to detect different sleep stages.

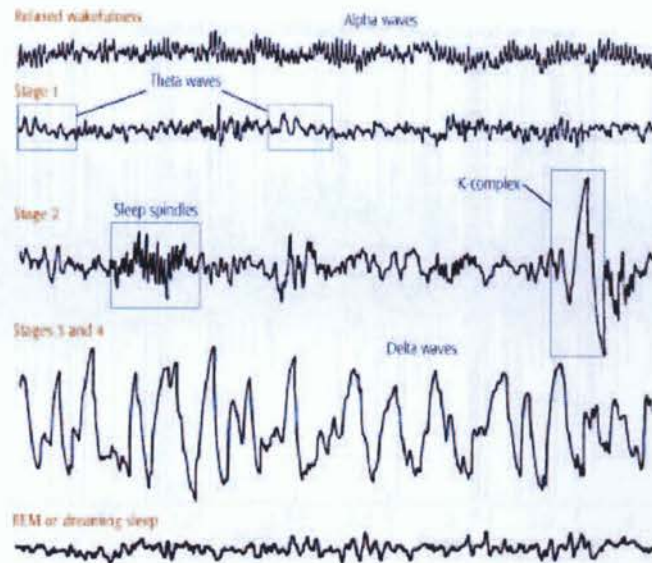


Figure 1.2: EEG waves [7]

1.1.2 EOG

The EOG signal has been frequently used in sleep recording, and it has been mostly used for recordings the eye movements in sleep research [8]. The eye movement (horizontal and vertical movements) aids doctors and sleep specialists to analyze the effect of medical drugs prescribed to a patient with different diseases, such as depression. Also, like EEG, EOG has been widely used in polysomnography to detect different sleep stages.

1.1.3 EMG

The electromyogram is used to record the electrical activity of muscles. When muscles are active they produce an electrical current which is proportional to the level of the activity. In polysomnography, the EMG consists of tonic (steady) and phasic (intermittently elevated) bursts in different stages of sleep. The phasic activity is defined as any burst of EMG activity lasting for $0.1 - 2s$, and has an amplitude of at least $50 \mu V$, while tonic constitutes the remainder [9]. Figures 1.3 and 1.4 show the EMG signals during sleep for both normal and abnormal subjects respectively. It has been shown that by studying the behavior of

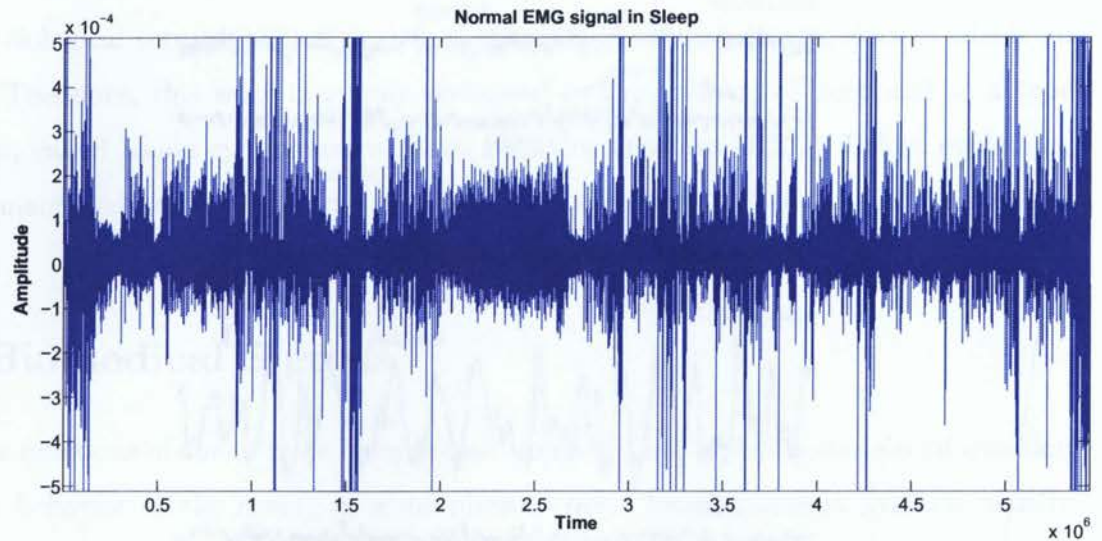


Figure 1.3: Normal chin EMG signal in Sleep

these different signals one can detect the sleep abnormalities, in different subjects [7].

1.2 Sleep

Sleep is a universal biological phenomenon, and in humans accounts for the way in which one spends a third of his/her life. Sleep is not an eventless process, on the contrary, many events occur in the body during this state: blood pressure falls, heartbeat slows down, muscles relax, and the body's metabolic rate decreases. Sleep in normal adults is accomplished when a number of changes in the CNS bring about a set of behavioral, physiological, and psychological changes. The sleep–wakefulness cycle can be characterized by the polysomnographic recording of three basic parameters: EEG, EOG, and EMG. From these three basic parameters three different states of sleep are distinguished: Waking state, Non–rapid Eye Movement (NREM) sleep, and Rapid Eye Movement (REM) sleep [10]. The following will show the characteristics of each state in details [6][11].

Wakefulness The signal has sinusoidal wave type characteristics, (alpha activity of 8 –

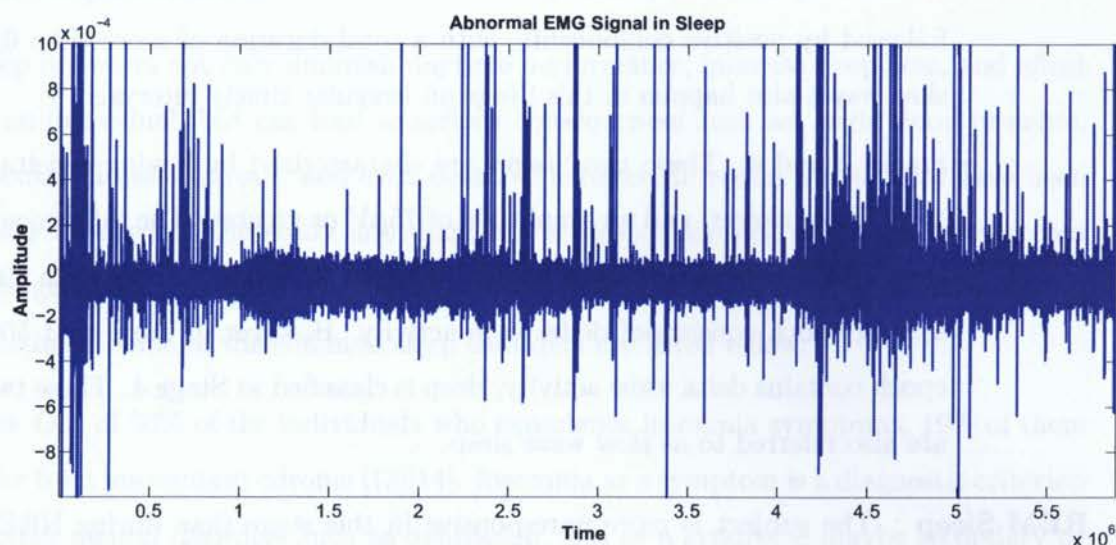


Figure 1.4: Abnormal chin EMG signal in Sleep

12 Hz), intermixed with lower amplitude irregular beta wave (13 – 35 Hz). The EOG activity maybe slow or rapid, usually recorded, as out of phase or in phase deflections. The EMG activity is relatively high and tonic, and there are movement artifacts.

Non-REM Sleep The subject is lying down motionless, his/her eyes are closed, and a given sensory input such as noise and light no longer induces behavioral responses. From the polysomnography point of view, this state of sleep consists of 4 different stages of sleep.

- Stage 1: The subject falls sleep and his/her muscles relax. The EOG has slow and predominantly horizontal eye movements. At the EEG level, the alpha activity has relatively low-voltage waves ($50 - 70\mu V$) and the theta range of 4 – 7 Hz. This Stage only last for a few minutes, if the subject is not disturbed.
- Stage 2: This Stage consists of sleep spindles and K-complexes. Sleep spindles are detected as brief bursts of rhythmic waves, with a frequency of 12 – 14 Hz and a duration of at least 0.5 s. On the other hand, K-complexes are defined to

be relatively high-amplitude potentials, as well as, having a negative sharp wave followed by positive components, with a total duration of more than 0.5s. The slow waves also happen in this Stage on irregular timely intervals.

- Stages 3 and 4: These two Stages are characterized by having a delta wave of $1 - 2\text{ Hz}$ or slower, and an amplitude of $75\mu\text{V}$ or greater. The difference between these two Stages is that of in Stage 3, at least 20% to not more than 50% of the scoring epoch consists of delta wave activity. However, if more than 50% of the epoch contains delta wave activity, sleep is classified as Stage 4. These two Stages are also referred to as slow wave sleep.

REM Sleep : The subject is more unresponsive in this stage than during NREM sleep. His/her eyes are periodically moving under the closed eyelids. This is the Stage in which the dreaming occurs. The polysomnography is characterized to have a low voltage EEG activity, which is closely similar to that of Stage 1. The theta activity of $4 - 7\text{ Hz}$ exists, which is often in conjunction with bursts of REM. In contrast, muscles are completely relaxed, yet, the flat EMG tracing is periodically interrupted by muscle twitches. The REM state of sleep is characterized by two different twitches. One is the discontinuous event such as muscle twitches and REMs, which is called the phasic events. The other is continuous processes such as desynchronizing EEG and muscle hypotonia, which is called tonic events.

Therefore, in mammals, the wake-sleep cycle have conjunction with prominent changes in their behavior. A young adult spends 20 – 28% of his/her sleep in the REM sleep; 4 – 5% in Stage 1; 46 – 50% in Stage 2; 6 – 8% in Stage 3; and 10 – 16% in Stage 4. As stated before at least 20% of the today's population suffer from a sleep disorder. When one does not sleep or is perceived not to sleep adequately, his/her quality of life is undeniably affected. The next section is devoted to a general overview of the different sleep disorders followed by an explanation of effect of the EMG signals in REM sleep, as well as a disease, called REM Behavior Disorder (RBD).

1.2.1 Sleep Disorder

Many sleep disorders not only diminish daytime performance, increase sleepiness, and affect mood of an individual, but can lead to serious consequences such as: high blood pressure, cardiovascular diseases, stroke, and even death. Therefore, in recent years there have been many works devoted to the study and diagnosis of sleep disorders, and in fact researches have made genuine progress in recognizing sleep disorders in the general population and in clinical settings. Some of the common sleep disorders are listed below:

Insomnia Out of 30% of the individuals who experience insomnia symptoms, 10% of them suffer from insomnia syndrome [13][14]. Insomnia as a symptom is a diagnostic criterion of other mental disorders such as depression, and as a syndrome maybe secondary or comorbid to another disease [15]. Insomnia is defined as a complaint of difficulty initiating, or maintaining sleep, or having a poor quality sleep for a period of at least a month. There are many consequences related to insomnia such as: fatigue, sleepiness, mood disruption, and effectively generating other problems including falling asleep while driving vehicles.

Sleep Apnea Obstructive sleep apnea, usually, involves when the breathing of the subject is disrupted in sleep. This disease usually creates loud snoring followed by 20 – 30s of silence. Because of a serious disturbance in the subject's normal sleep pattern, people with sleep apnea often experience extreme fatigue during the day, and their concentration and daytime performance suffer. It has been estimated that up to 50% of the sleep apnea patients have high blood pressure, which could result in a sudden cardiac death during the sleep and stroke [16].

Narcolepsy This disorder usually happens in young adults, and is usually illustrated by multiple refreshing naps during the day. The day time polysomnography shows signs of drowsiness on EEG signal, associated with REM periods. On the other hand, the night time polysomnography shows an increased number of awakenings, as well as an increased amount of Stage 1. Again because of sleepiness, individuals suffering from

narcolepsy are at higher risk of having accidents at home or at work, or even on the road. In addition, education, occupation, and social relations may be influenced which in turn will affect the quality of the life of the patient [17].

REM Sleep Behavior Disorder (RBD) RBD is characterized by elaborating movements correlated with dream during REM sleep [15][18]. In this condition, patients lose their normal muscle atonia, and in turn enact their dreams. REM sleep in mammals involves a highly energized state of brain activity, with tonic (i.e., continuous) and phasic (i.e., intermittent) activations, occurring across a spectrum of physiologic parameters [19][20]. Sleep neurophysiologists refer to REM sleep as active sleep, because of the high level of brain activity; and as paradoxical sleep, because there is a virtual absence of skeletal muscle activity despite a highly activated brain state. Generalized skeletal muscle atonia is one of the three defining features of mammalian REM sleep. Thus, the paradox of REM sleep resides in the absence of overt motor expression, during an active brain and mind (dream) state. The loss of this customary paradox in RBD, bears serious clinical consequences such as: paradox lost, which means loss of safe sleep. Serious injuries have the potential to occur as a result of dream acting behavior, leading to a complaint of “acting out my dream” [21]. The overnight polysomnography of the subjects also confirm the presence of muscle tone in REM sleep. In addition RBD can be an early warning for the emergence of Parkinsonism and other neurodegenerative conditions antedating the illness by many years [22].

Neuromuscular disorder in general may encompass many diseases that are via intrinsic muscle pathology, or via nerve pathology, which will in turn impair the functioning of the muscles. In other words, this type of disease could affect muscles and/or nervous control. Therefore, studying the neuromuscular behavior of the subjects could help doctors and neurologist to detect many abnormalities. The mentioned disorders, if not detected, could result into Parkinson's, diabetes, periodic limb movements, and many more. The next subsection will explain some of these diseases and their relationship with EMG in sleep in more detail.

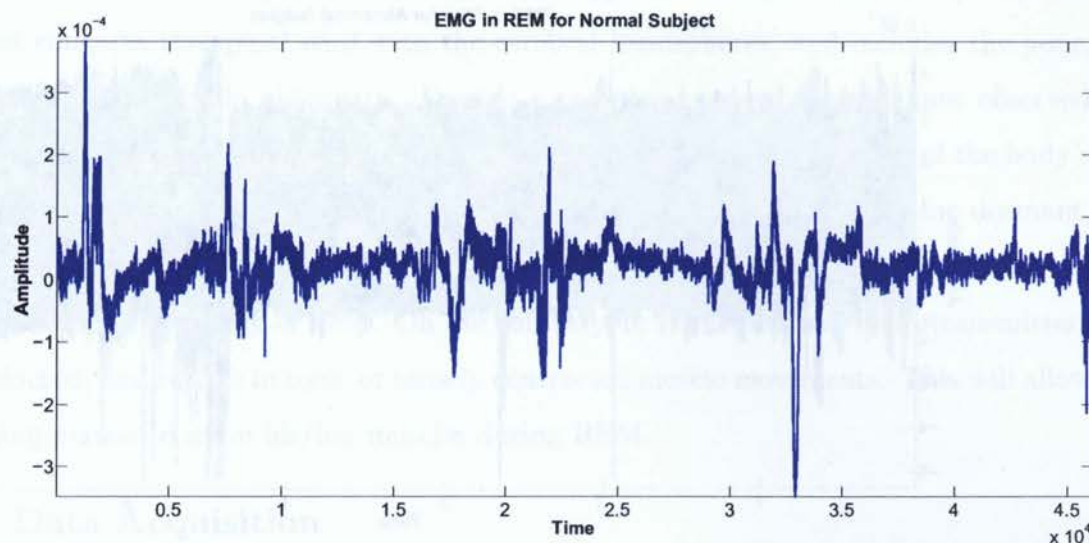


Figure 1.5: Normal EMG signal in REM sleep

1.2.2 EMG and RBD

For patients with RBD, tonic tone is higher than in normal subjects. The increase in tonic EMG activity might reflect disease progression [19]. It has been noted that RBD is a precursor to clinically evident Parkinson's disease, although medications may also affect this measure [20]. RBD is characterized by increased axial submental (under the chin) muscle tone. Therefore, chin EMG, which is routinely collected in sleep studies, can be used as a valuable signal in detecting early forms of neurodegenerative conditions. This may also provide a useful measure for assessing the response to neuroprotective drugs. Figures 1.5 and 1.6 show the normal and abnormal EMG in REM sleep.

It is worth noting that using the hypnogram for each signal, the EMG for different stages as well as REM sleep could be extracted. As a consequence Figures 1.5 and 1.6 are plotted using a hypnogram. As can be seen, the time axis of these two figures show different values, which is as a result of different REM periods of sleep for different subjects.

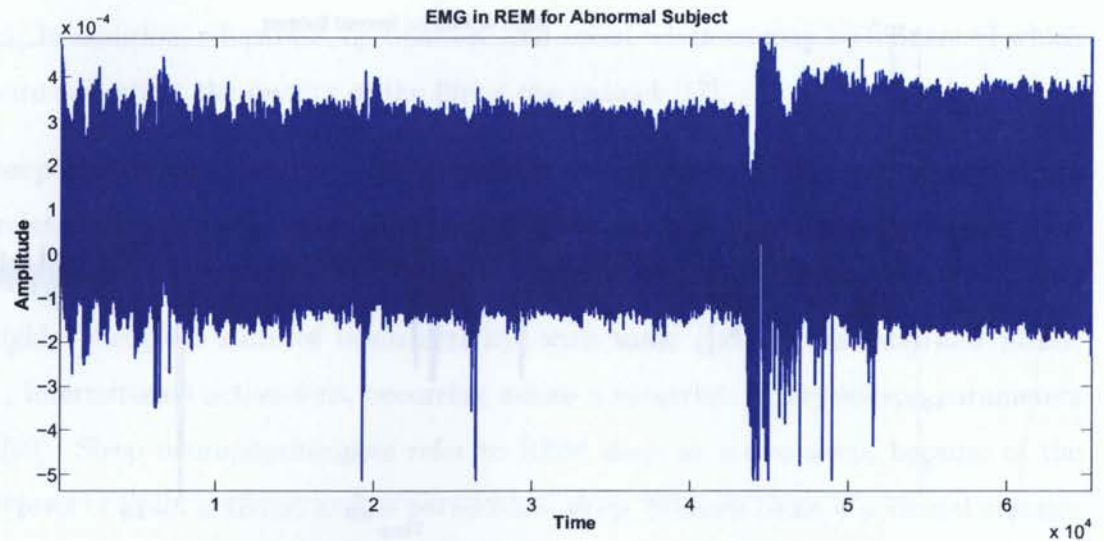


Figure 1.6: Abnormal EMG signal in REM sleep

1.2.3 Parkinson's Disease

Parkinson's disease (PD) and related movement disorders are common neurological diseases, particularly common in elderly people. Although the daytime phenomena of PD has been well recognized since the early 1800's, the frequent nocturnal symptoms, which occur in as many as 75% of patients, and the associated sleep abnormalities were not systematically studied until the 1960s. PD's characteristics, and pathological features have been known since the early 20th century, although its root cause is still uncertain [11].

Sleep medicine specialists have found that almost two-thirds of patients, with RBD, develop degenerative brain diseases, by approximately 11 years after diagnosis of RBD. That means, for every two RBD patients one will develop PD after 10 to 11 years. By the time a patient is diagnosed with PD about ninety percent of the patient's neurons are dead and medication becomes ineffective. However, if doctors could predict from RBD that a particular patient has the potential of developing a PD then treatment could start 10 years in advance, therefore increase recovery percentage of that patient.

The basic cause for sleep paralysis during REM happens in the brainstem, the part of the brain that connects the spinal cord with the cerebral hemispheres, and includes the pons, midbrain, and the medulla oblongata. Despite a complex process, doctors have observed that the brainstem undergoes changes in REM sleep, which results in paralysis of the body's voluntary muscles. Certain neurotransmitters, like acetylcholine (ACh), become dormant, and do not naturally transmit motor activity to ensure restful and, inactive sleep during the most electrically active stage of sleep. On the contrary, in RBD patients, neurotransmitters are not blocked, and results in tonic or tensely contracted muscle movements. This will allow the sleeping person to move his/her muscles during REM.

1.2.4 Data Acquisition

A traditional scoring system for sleep has been established [10], with the electrophysiological parameters of EEG, EOG, and EMG. The system used for recording chin EMG signals during sleep includes 3 relatively midline electrodes, one above the jaw line, one below the jaw line, and one back-up electrode. The EMG signal is freely triggered and bandpass filtered at 10 – 100Hz. The impedance of every electrode is less than $10K\Omega$ with a minimum digital resolution of 12 bits per sample. The sampling rate is 256Hz, and similar electrodes are used to record EEG and EOG amongst other physiological parameters. Data collection from humans were facilitated through a protocol, approved by the Local Research Ethics Board (LREB).

1.2.5 Dataset Type

In this work, a subject is defined as historically normal if there is no history of any violent behavior during the night sleep; otherwise it is considered as abnormal. 4 volunteers with normal behavior and 4 subjects with RBD scheduled were selected to undergo the sleep test, independent of this work. Each subject slept at the clinic during a night, and signals such as EEG, EMG, and EOG were recorded from these subjects.

1.3 Signal Analysis

Signal processing algorithms are needed to analyze and understand the signals. On the other hand, these techniques can exploit the intrinsic properties of the signal, which allows a specific function to be applied on the signal. Therefore, an analysis tool which could be tuned to characteristics of the signal is selected in such a way to extract the features of interest [23][24]. Thus, the primary goal of signal analysis is to extract useful information to understand the signal generation process, or extract features for signal classification purposes. Most of the methods in this area are treated under the disciplines of spectral estimation and signal modeling [25].

1.3.1 Signal Behavior

The mathematical analysis of a signal requires the availability of a mathematical description for the signal itself. The type of description, usually referred to as a signal model, determines the most appropriate mathematical approach for the analysis of the signal [26][27]. The term signal is used to refer to, either the signal itself, or its mathematical description, which is the signal model. The exact meaning will be apparent from the context. Clearly, this distinction is necessary if a signal can be described by more than one model. The most important classification of signal models are either deterministic or random.

Deterministic Signals

Deterministic signals are any signals which can be described by an explicit mathematical relationship. This characteristic of the signals allows for advanced prediction of signal quantities. In the case of continuous time signal, the mathematical relationship is a given function of time [28].

Nondeterministic or Random Signal

In contrast to deterministic signals, nondeterministic signals are defined to be signals that cannot be described, to any reasonable accuracy, by explicit mathematical relationships. In

other words, the lack of such an explicit relationship implies that the signal evolves in time in such an unpredictable manner. These signals are often called random.

Yet, although random signals are evolving in time, in an unpredictable manner, their average properties can often be assumed to be deterministic. In other words, they can be specified by explicit mathematical formulas. At this point one could say that complete knowledge of the physics of the signal could provide an explicit mathematical relationship, at least within the limits of the uncertainty principle. This concept is a key to the modeling of a random signal, as a stochastic process. Random signals are further divided into two groups:

Stationary : Stationary signal is a signal that its property does not change over time. In other words, it has a constant probability distribution for all time instants. This results in constant first and second order statistics such as mean and variance.

Nonstationary : Nonstationary signal on the other hand has a time-varying probability distribution. This causes other properties that depends on probability distribution function (PDF) to be time-varying.

It is a very well known fact, in the real physical world that there exists abundant kinds of signals. All those signals carry a lot of information that are of people's interests; people develop diverse techniques to analyze, interpret, manipulate, and process those signals. Biomedical signals are types of signals that are strongly related with human body or human organisms. It has been shown that the type of most of these signals are nondeterministic, and further nonstationary, or quasistationary. Yet, in order to be able to apply signal processing techniques on these types of signals, either the signal has to be divided into stationary components, or techniques that can be applied on nonstationary signals has to be used. The aim of this thesis is to analyze EMG signals by using these two different techniques.

1.4 Motivation

As stated before, the nonstationary properties of biomedical signals, lead researchers to use techniques that could be applied either without or with segmenting the signal into stationary components. This thesis aims to evaluate the usefulness of these algorithms for biomedical signal applications via EMG in REM sleep.

Adaptive signal processing is one of the techniques that has been used in this work to overcome the nonstationary problem. The method that has been used in this work is the Recursive Least Square (RLS) algorithm, which tries to “find and track” the optimum filter corresponding to the same signal operating environment, with complete knowledge of the required statistics. This will enable one to mark the stationary boundary, by studying the least square error of the signal. After dividing the signal into stationary components, Autoregressive (AR) modeling and Cepstrum analysis were applied to each stationary segment.

Wavelet Transform (WT) is another method to analyze the nonstationary signal. WT is a very powerful technique that has been widely used in biomedical signal processing literature. This technique is a joint time–frequency analysis, which analyzes both the time and frequency properties of the signal simultaneously. One of the main advantages of the WT is the use of a varying size window to access accurate view of the signal either in time, or in frequency. The time–frequency information extracted from the signal is very important, since it’s main objective is to analyze a time varying signal.

As of recently, none of these methods have been applied on EMG signal during sleep. Therefore, the use of an overnight EMG recording to detect RBD represents a unique methodology with a novel clinical dataset. This is the motivation behind the work reported in this thesis.

1.5 Organization of the thesis

As previously stated, this thesis investigates the behavior of EMG in normal and abnormal subjects. The abnormality of the EMG in REM sleep might lead to very neurodegenera-

tive disorders such as Parkinson disease (PD). If predicting that a subject is sensitive to generate PD, then the treatment could be started years prior to the diagnosis. As a result, this thesis focuses on signal processing techniques that will help understand and monitor neurodegenerative diseases. Figure 1.7 shows the organization of the thesis.

Chapter 2: Signal Analysis Methodologies

This Chapter discusses the signal analysis by the way of adaptive signal processing, AR modeling, Cepstrum analysis and Wavelet Transform. The methodology of each of these techniques, as well as, its applications are discussed in this Chapter.

Chapter 3: Parametric Signal Analysis of EMG in Sleep

This Chapter uses adaptive signal processing techniques to segment the signal into stationary components. The technique which is used is the RLS algorithm. The goals, as well as, the applications of these types of algorithms have been discussed. AR modeling and Cepstrum analysis have been described, and were applied on each stationary signals. A brief discussion on the classification scheme used is also included, along with experimental results.

Chapter 4: Wavelet Analysis of EMG in Sleep

An important application that have been widely used in the recent biomedical literature is being discussed in this Chapter. This application is specefic to Wavelet Transform. Both the Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT) are explained in this Chapter, followed by their application on the EMG. The overall performance and the advantages of using WT compared to the adaptive signal processing are discussed; followed by experimental results. The Receiver Operating Characteristic (ROC) curve is plotted at the end of this Chapter as well.

Chapter 5: Conclusion

Chapter 5 contains all the experimental discussion, a summary of the results, and directions for future work.

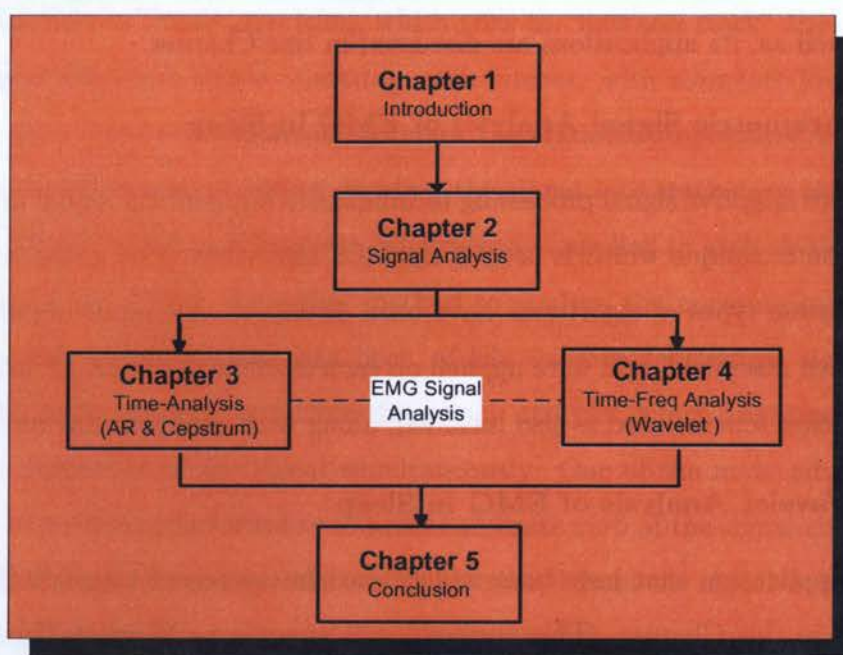


Figure 1.7: Organization of the thesis

Chapter 2

Signal Analysis

2.1 Literature Review

IN the previous Chapter it was shown that the fundamental characteristics of random signals were captured in the following statement: Although random signals are evolving in time, in an unpredictable manner, their average statistical properties exhibit considerable regularity. This provides the ground for the description of random signals, using statistical averages instead of explicit equations. Thus, in practice, random signals are being analyzed using the statistical techniques. Their instantaneous values are described mathematically using theory of probability, random variables, and stochastic processes, because of their unpredictable behavior [29]. Within this framework one can develop, at least in principle, theoretically optimum signal processing methods that can inspire the development, and can serve to evaluate the performance of practical and statistical signal processing techniques [28].

As stated before, random signals, in stochastic context, are divided into two categories: stationary and nonstationary signals. Stationarity of a signal is concerned with its probabilistic behavior. A signal is considered to be stationary, as long as some of its specified properties remain constant with time. These type of signals may allow for advanced prediction of the signal quantities, since the signal may be described by a mathematical function. However; from a practical point of view, stationarity has been understood as time invariance, only to the second order. As a consequence, given a stochastic signal $x(t)$, the first

order statistics such as mean value $m_x(t)$ and second order statistics such as variance σ are constants and the autocorrelation function R_x is a non-negative definite function. These are shown in Equations 2.1 to 2.4 [28]:

$$m_x(t) = E[x(t)] = c \quad (2.1)$$

$$\sigma^2 = E[X(t) - m_x(t)]^2 \quad (2.2)$$

$$R_x(t_1 : t_2) = E[x(t_1)x(t_2)] \quad (2.3)$$

$$R_x(t_1 - t_2) = R_x(\Phi) \quad (2.4)$$

In these equations c is constant, E is Expectation value or mean, and $\Phi = t_1 - t_2$.

The main objective of nonstationary signals are the statistical description, the dependence of the modeling, the exploitation of the values of one or more discrete-time signals, their application to theoretical, and practical problems [29].

A nonstationary signal has a time-varying probability distribution which causes other quantities that rely on the probability density function (PDF) to also be time-varying. In this case the mean, variance, and autocorrelation functions would change with time, since they are computed from the PDF of a signal. Based on the fact that the Fourier Transform of the autocorrelation function is equal to the power spectral density (PSD) of a signal, the PSD of a nonstationary signal is also time-varying. Consequently, a nonstationary signal has time-varying spectral content.

2.2 Adaptive Signal Processing

The area of adaptive signal processing involves the use of optimal and statistical signal processing techniques to design signal processing systems, that can modify their characteristics during normal operation (usually in real time). This may achieve a clearly predefined application-dependent objective [28], and may create signal processing tools which would be able to monitor time variations of statistical properties of the signals, and divide them into locally stationary components. One of these methods is modeling the nonstationary stochastic

time series which has been a continuing subject of research for many years. One important model for such time series, which has found widespread use, is the quasi-stationary autoregressive process, where it has been successfully applied to different areas such as speech processing, EEG, ECG, VAG (Vibroarthographic), and EMG [30][31]. In this model, statistical properties of the time series are described solely by sets of AR parameters where each set remains constant within a certain time interval or “segment” of arbitrary length (stationary segment), and changes abruptly to a new set of parameters on reaching the boundary of the segments. In order to establish a quasi-stationary model for a given time series, and to estimate the parameter values in each segment, it is necessary not only to detect, but also to localize the segment boundaries as well as possible. An algorithm performing this job is called a segmentation algorithm [32][33].

Segmentation can be done in two ways: 1) Fixed Segmentation which uses a fixed width window to divide the signal into constant-length segments, and 2) Adaptive Segmentation which tries to predict the behavior of the samples using past sample values. The drawback with the fixed segmentation technique is using constant-length reference window may cause problems of poor estimation of low-frequency components, and improper detection of sudden variations; another issue is defining an appropriate length for the window [31]. Adaptive Segmentation conversely, depending on the algorithm used, tries to predict the signal from the few past samples. This makes adaptive segmentation technique a powerful tool in signal processing.

It has been shown through literature that both fixed segmentation and adaptive segmentation have been widely dedicated to segment the nonstationary signal into locally stationary signals. As of recently, none of these methods have been applied on EMG signal in REM sleep. Therefore, the use of an overnight EMG recording to detect RBD represents a novel methodology.

To provide a comparative analysis of the techniques used in this thesis, other researches in the area of fixed signal processing and adaptive signal processing methods for non-stationary signals are included in Table 2.1. Michael et al. [34], directly calculated the boundaries

Table 2.1: Related works in fixed and adaptive segmentation for biomedical signals

Source	Method	Application	Drawback
Michael et al. [34]	The Autocorrelation Function	EEG	No modeling Fixed-size window
Appel et al. [32]	The Generalized Likelihood Ratio Method	Non-stationary Signal	Computationally Complex
Tavathia et al. [35]	The Spectral Error Measure (SEM)	VAG	Cannot Detect Sudden Variation
Moussavi et al. [31]	Recursive Least Square RLS	VAG	Threshold is a single dependent parameter
Krishnan et al. [36][37]	Recursive Least Square Lattice (RLSL)	VAG	Computationally complex

for each segments from the autocorrelation function (ACF) in which the ACF method of adaptive segmentation makes use of the values of the short-time autocorrelation function estimated from the signal. The problem with the ACF is that it does not use any explicit modeling techniques. Appel et al. [32] on the other hand uses an adaptive segmentation technique by means of Kalman–Bucy filtering and generalized ratio techniques to detect and to estimate abrupt variations in the given signal. The Kalman–Bucy is an efficient recursive filter that estimates the state of a linear dynamic system from a series of noisy measurements. Nevertheless the computational complexity of this method is the drawback with this technique. Tavathia et al. [35] uses an adaptive segmentation method, based on the spectral error measure (SEM) on VAG signal. In this method the samples were modeled, using parameters of a constant-width reference window. A limitation of this method was that if a rapid variation occurs in the middle of a test window, the segment boundary will be at the beginning of the window, and not at the exact position of the variation. In Krishnan et al. [36][37] an adaptive segmentation method was used based on the Recursive Least Square Lattice (RLSL). In this method a data sequence is replaced by an orthogonal set of variables which would increase the speed of adaptation, the ease of testing for the minimum phase condition, the convergence, and the tracking capability. However; this algorithm is computationally expensive since it involves updating forward prediction error

$f_m(n)$, backward prediction error $b_m(n)$, forward prediction error power $F_m(n)$, and backward prediction error power $B_m(n)$. The character m and n , are model order and time instants of the RLSL algorithm respectively. Moussavi et al. [31] uses the Recursive Least Square Algorithm (RLS) in which the sum of squared difference between the RLS filter tap-weight vectors of adjacent time samples were compared with a variable threshold value for detecting segment boundaries. These methods have been applied on many different signals as well as images such as MRI images. In this work the RLS algorithm was used to divide the EMG signal into locally-stationary segments, which will be discussed in more details in the next section [31].

2.3 Recursive Least Square Algorithm

The least square principle was first introduced by the German mathematician Carl Friedreich Gauss, who used it to determine the orbit of the asteroid Ceres in 1821 by formulating the estimation problem as an optimization problem [28]. These principles show that adaptive filters can improve their performance, during normal operation, by learning the statistical characteristics through processing current signal observations. The goal of any adaptive filter such as the RLS Algorithm is to “find and track” the optimum filter corresponding to the same signal operating environment, with the complete knowledge of the required statistics. The performance of such adaptive filters are evaluated, by using the concept of stability, speed of adaptation, quality of adaptation, and tracking capabilities. Therefore the distinguishing feature of adaptive filters is that they can modify their responses to improve their performance during operation, without any intervention from the user [38].

An important feature of the RLS Algorithm is that it utilizes information, contained in the input data, and extends it back to the instant of time when the algorithm was initiated [38]. That is, given the least squares estimates of the tap-weight vector of the filter at time $n - 1$, the updated estimate of this vector at time n can be computed upon the arrival of new data. RLS adaptive filters are designed so that the updating coefficients always attain the minimization of the total squared error from the time the filter initiated operation up to

the current time. Therefore the filter coefficients are chosen to minimize the cost function, where the cost function is shown in Equation 2.5 [28].

$$\xi(n) = \sum_{i=1}^n \lambda^{n-1} |e(i)|^2 = \sum_{i=0}^n \lambda^{n-1} |d(i) - w^T x(i)|^2 \quad (2.5)$$

where $e(i)$ is the instantaneous error, and the constant λ is the forgetting factor, $0 < \lambda \leq 1$. The value of the forgetting factor indicates how fast the algorithm forgets its past value, hence it ensures that data in the distant past are paid less attention. If the weight vectors are defined in Equation 2.6

$$\underline{w}(n) = [w_0(n), w_1(n), \dots, w_{M-1}(n)]^T \quad (2.6)$$

and the tap-input vector at each input time instant n be defined as an M dimensional vector to be of the form of Equation 2.7

$$\underline{x}(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T \quad (2.7)$$

then the estimation error is:

$$e(n) = d(n) - \underline{w}^T(n) \underline{x}(n) \quad (2.8)$$

The performance index or objective function to be minimized in the sense of least squares is defined in Equation 2.9

$$\xi(n) = \sum_{i=1}^n e^2(i) \quad (2.9)$$

This is the cost function with the assumption of $\lambda = 1$, where the main purpose of the algorithm is to minimize this value. Optimization of Equation 2.9 leads to the normal equation as defined below

$$\Phi(n) \hat{w}(n) = \Theta(n) \quad (2.10)$$

where $\hat{w}(n)$ from Equation 2.10 is the optimum value of the tap-weight vector for which the performance index is minimum, $\Phi(n)$ is an MXM time-averaged correlation matrix of the input, and it is defined as follows:

$$\Phi(n) = \sum_{i=1}^n \underline{x}(i) \underline{x}^T(i) \quad (2.11)$$

$\underline{\Theta}(n)$ is an $MX1$ time-averaged cross-correlation matrix between the desired response and the input shown below

$$\underline{\Theta}(n) = \sum_{i=1}^n \underline{x}(i)d(i) \quad (2.12)$$

Since it would be difficult to solve the normal equation for the optimum value of the tap-weight, recursive techniques need to be considered. In order to achieve that, the summation of Equation 2.12 has to be broken into two parts. The first part corresponds to sum of all the terms except the last term (i from 1 to $n-1$), and the second part corresponds to the last term ($i = n$). That is:

$$\underline{\Phi}(n) = [\sum_{i=1}^{n-1} \underline{x}(i)\underline{x}^T(i)] + \underline{x}(n)\underline{x}^T(n) \quad (2.13)$$

Comparing Equation 2.11 with Equation 2.13, the $[\sum_{i=1}^{n-1} \underline{x}(i)\underline{x}^T(i)]$ could be interpreted as $\underline{\Phi}(n-1)$. Therefore Equation 2.13 can be rewritten as a recursive expression, given by Equation 2.14,

$$\underline{\Phi}(n) = \underline{\Phi}(n-1) + \underline{x}(n)\underline{x}^T(n) \quad (2.14)$$

Similarly Equation 2.12 could be rewritten as the recursive form of Equation 2.15,

$$\underline{\Theta}(n) = \underline{\Theta}(n-1) + \underline{x}(n)d(n) \quad (2.15)$$

To compute the least squares estimate $\hat{\underline{w}}(n)$ for the tap-weight vector in accordance with Equation 2.10 the inverse correlation matrix $\underline{\Phi}(n)$ has to be calculated. However; this is time consuming, particularly, if the model order M is high. In order to reduce the time, a matrix inversion lemma called the "ABCD" lemma could be used. According to this lemma,

$$[\underline{A} + \underline{BCD}]^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{B}(\underline{DA}^{-1}\underline{B} + \underline{C}^{-1})^{-1}\underline{DA}^{-1} \quad (2.16)$$

where \underline{A} , \underline{C} , $\underline{A} + \underline{BCD}$ and $\underline{DA}^{-1}\underline{B} + \underline{C}^{-1}$ are all invertible functions. By applying Equation 2.16 to Equation 2.14 assuming $\underline{\Phi}(n)$ to be positive definite and therefore non-singular and also by assuming

$$\underline{A} = \underline{\Phi}(n-1)$$

$$\underline{B} = \underline{x}(n)$$

$$\underline{C} = 1$$

$$\underline{D} = \underline{x}^T(n)$$

the $\underline{\Phi}^{-1}(n)$ would be as follows:

$$\underline{\Phi}^{-1}(n) = \underline{\Phi}^{-1}(n-1) - \underline{\Phi}^{-1}(n-1)\underline{x}(n)[1 + \underline{x}^T(n)\underline{\Phi}^{-1}(n-1)\underline{x}(n)]^{-1}\underline{x}^T(n)\underline{\Phi}^{-1}(n-1) \quad (2.17)$$

However since the expression inside the bracket is constant therefore Equation 2.17 can be rewritten as follow:

$$\underline{\Phi}^{-1}(n) = \underline{\Phi}^{-1}(n-1) - \frac{\underline{\Phi}^{-1}(n-1)\underline{x}(n)\underline{x}^T(n)\underline{\Phi}^{-1}(n-1)}{1 + \underline{x}^T(n)\underline{\Phi}^{-1}(n-1)\underline{x}(n)} \quad (2.18)$$

For convenience of notation the $\underline{\Phi}^{-1}(n)$ has been redefined as

$$\underline{P}(n) = \underline{\Phi}^{-1}(n)$$

where $\underline{P}(0) = \delta^{-1}\underline{I}$, δ is a small constant and, \underline{I} is the identity matrix. The Kalman gain vector is defined as follows:

$$\underline{K}(n) = \frac{\underline{P}(n-1)\underline{x}(n)}{1 + \underline{x}^T(n)\underline{P}(n-1)\underline{x}(n)} \quad (2.19)$$

Equation 2.18 can be rewritten as:

$$\underline{P}(n) = \underline{P}(n-1) - \underline{K}(n)\underline{x}^T(n)\underline{P}(n-1) \quad (2.20)$$

By multiplying both sides of Equation 2.19 by the denominator on the right-hand side of the same Equation

$$\underline{K}(n)[1 + \underline{x}^T(n)\underline{P}(n-1)\underline{x}(n)] = \underline{P}(n-1)\underline{x}(n) \quad (2.21)$$

or:

$$\underline{K}(n) = [\underline{P}(n-1) - \underline{K}(n)\underline{x}^T(n)\underline{P}(n-1)]\underline{x}(n) \quad (2.22)$$

Comparing Equation 2.22 with Equation 2.20

$$\underline{K}(n) = \underline{P}(n)\underline{x}(n) \quad (2.23)$$

where $\underline{P}(n)$ and $\underline{K}(n)$ have the dimensions of $M \times M$ and $M \times 1$ respectively. By using Equation 2.10 and Equation 2.15 a recursive equation for updating the least square estimate $\hat{\underline{w}}(n)$ can be derived.

$$\begin{aligned}\hat{\underline{w}}(n) &= \underline{\Phi}^{-1}(n)\underline{\Theta}(n) \\ &= \underline{P}(n)\underline{\Theta}(n) \\ &= \underline{P}(n)\underline{\Theta}(n-1) + \underline{P}(n)\underline{x}(n)d(n)\end{aligned}\quad (2.24)$$

Substituting Equation 2.20 for $\underline{P}(n)$ in 2.25 leads to:

$$\begin{aligned}\hat{\underline{w}}(n) &= \underline{P}(n-1)\underline{\Theta}(n-1) - \underline{K}(n)\underline{x}^T(n)\underline{P}(n-1)\underline{\Theta}(n-1) + \underline{P}(n)\underline{x}(n)d(n) \\ &= \underline{\Phi}^{-1}(n-1)\underline{\Theta}(n-1) - \underline{K}(n)\underline{x}^T(n)\underline{\Phi}^{-1}(n-1)\underline{\Theta}(n-1) + \underline{P}(n)\underline{x}(n)d(n) \\ &= \hat{\underline{w}}(n-1) - \underline{K}(n)\underline{x}^T(n)\hat{\underline{w}}(n-1) + \underline{P}(n)\underline{x}(n)d(n) \\ &= \hat{\underline{w}}(n-1) + \underline{K}(n)[d(n) - \underline{x}^T(n)\hat{\underline{w}}(n-1)]\end{aligned}\quad (2.25)$$

By defining a new variable α in Equation 2.25, the equation for updating the least square estimates would be derived as follows:

$$\hat{\underline{w}}(n) = \hat{\underline{w}}(n-1) + \underline{K}(n)\alpha(n) \quad (2.26)$$

In which $\alpha(n)$ can be defined as

$$\begin{aligned}\alpha(n) &= d(n) - \underline{x}^T(n)\hat{\underline{w}}(n-1) \\ &= d(n) - \hat{\underline{w}}^T(n-1)\underline{x}(n)\end{aligned}\quad (2.27)$$

$\alpha(n)$ is often referred to as the *a priori* error, reflecting the fact that this is the error obtained using the old filter. This means that its value has been obtained prior updating the new data, since $\underline{P}(n) = \underline{\Phi}^{-1}(n)$. Therefore, the RLS algorithm uses the information contained in all the previous input data to estimate the inverse of the autocorrelation matrix of the input vector. It uses this estimate to properly adjust the tap weights of the filter. The behavior of the RLS algorithm showed that the filter coefficients are much more stable and they adhere much closer to the correct values.

In RLS algorithm, there are two variables involved in the recursion (those with time index $n-1$): $\hat{\underline{w}}(n-1)$ and $\underline{P}(n-1)$. In order to start the recursion the initial values of these parameters have to be provided.

- $\underline{w}(0)$

If some *a priori* information about the parameters $\hat{\underline{w}}$ is provided, this information will

be used to initialize the algorithm. Otherwise, the typical initialization is:

$$\underline{w}(0) = 0$$

- P(0) Recalling the significance of $\underline{P}(n)$

$$\underline{P}(n) = \underline{\Phi}^{-1}(n)$$

The exact initialization of the recursions uses a small initial segment of the data to compute its value. However, it is not a simple matter to select the length of data required for ensuring invertibility. The approximate initialization is, commonly, used, which is

$$P(0) = \delta^{-1}I$$

Thus $P(n)$ is proportional to the covariance matrix of the parameters $\underline{w}(n)$. Since the knowledge of these parameters at, $n = 0$, is vague, a high covariance matrix of the parameters is to be expected, and thus the value assigned to δ would be very small.

$$\delta^{-1} > 100/\sigma^2$$

For large data length, the initial values assigned at $n = 0$ are not important, since they are forgotten due to the exponential forgetting factor λ .

2.3.1 The concept of Adaptive Learning Curve

The concept of adaptive learning curve was developed early in the history of adaptive filtering and has proved to be very useful in both the analysis and the application of systems, employing adaptive algorithms. The ordinate of the curve is the performance function which is being minimized by the chosen adaptive algorithm. The iteration index l indicates how many iteration the algorithm has gone through. The learning curve begins at the upper left at the performance function's initial value, and then declines with iteration number l toward an asymptote. For a perfect adaptive algorithm, the descent of the curve would be abrupt, indicating quick convergence, and the asymptote would be J_{min} , the lowest value of

the performance function J attainable. Practical algorithms, however, is deviated from this model in three ways. Firstly, they would take some time to learn from the algorithm which will result in slower convergence. Secondly, the unpredictable behavior of the data might drive the performance function J upward for short intervals of time instead of proceeding monotonically downward. Thirdly, a problem might arise from the fact that some of the adaptive algorithms may not actually achieve the theoretical minimum [38]. One of the important examples would be the Least Mean Square (LMS) algorithm. This algorithm never converges, and instead “rattles around” the optimum coefficient choice. This results in an increase in the squared error, seen at the filter output. The presence of this extra squared error means that the learning curve for LMS adaptive filter reaches an asymptote which is always greater than the theoretical minimum of J . The shape of the learning curve and the performance function that has been chosen depends on many factors, such as: the choice of adaptive algorithm, the length of the filter, the choice of parameters used in the adaptation algorithm, and the nature of the input signal [40].

This learning curve would converge to a minimum value, J , if the signal is stationary. Figures 2.1 and 2.2 show the learning curve of both the LMS and RLS for a stationary signal respectively. From these two figures it can be seen that the RLS algorithm converges faster than the LMS algorithm. In case of a nonstationary signal, the learning curve will not converge to a single value. However, since a nonstationary signal consists of stationary or quasi-stationary segments, by using adaptive signal processing, one can segment the nonstationary signal into stationary parts. Therefore, time analysis of the signal can be performed on each stationary segment. One of these time-series analysis is Linear Prediction.

2.4 Linear Prediction

In linear prediction, a set of past samples of a signal $x(n)$ is used to predict the sample values of the process for some time in the future. The filter that performs the prediction is called a predictor. System identification, time series modeling, spectral estimation of speech, and biomedical signals are all applications of the Linear Prediction. The predictor output $\hat{x}(n)$

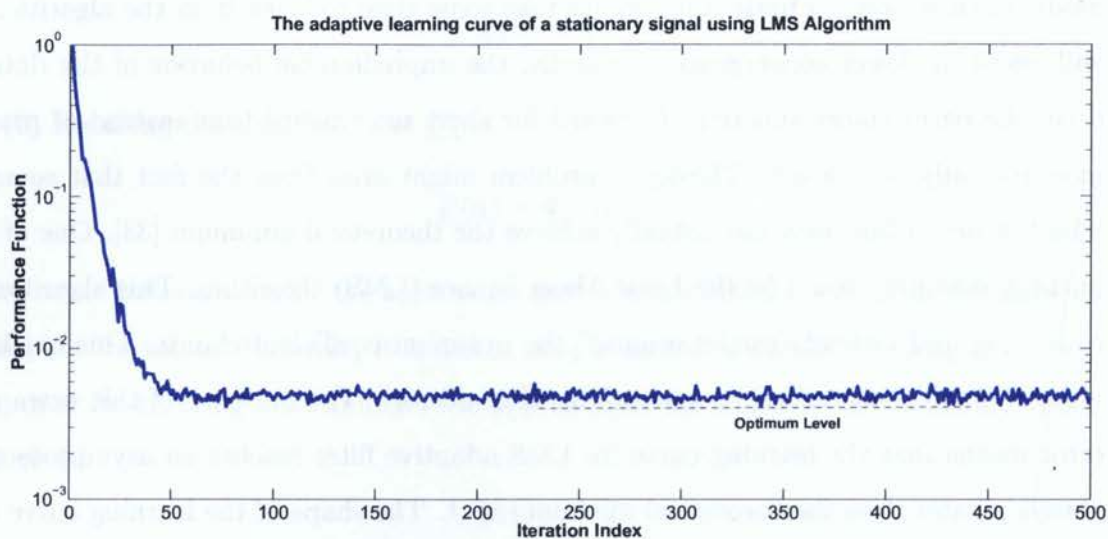


Figure 2.1: The Adaptive Learning Curve: Behavior of the performance function J with time

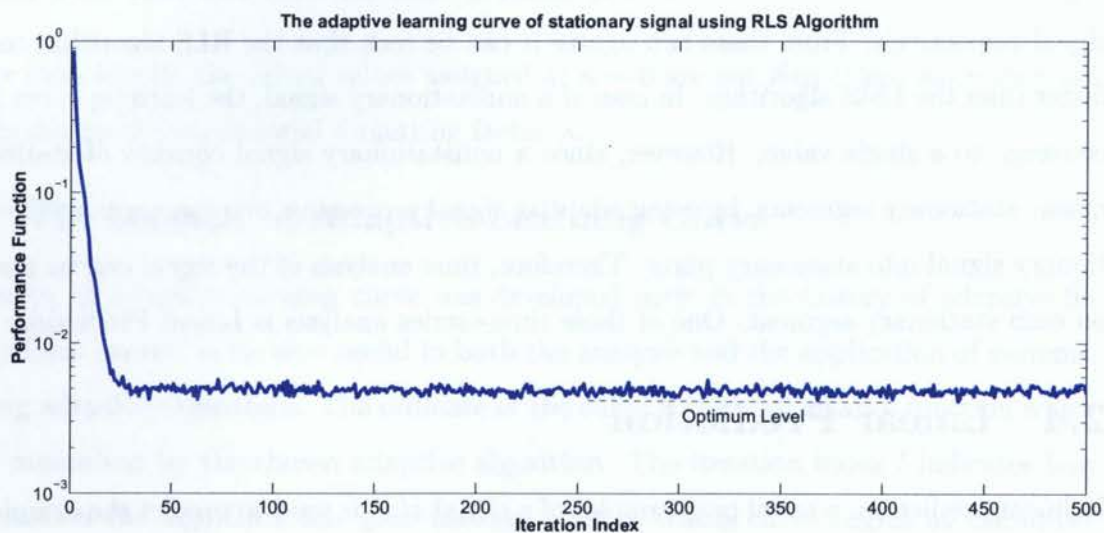


Figure 2.2: The Adaptive Learning Curve: Behavior of the performance function J with time

can also be expressed as a convolution of

$$\hat{x}(n) = \sum_{k=1}^P \alpha_k x(n-k) \quad (2.28)$$

where the tap weights α_k are also known as the predictor coefficients. One approach to optimize the predictor coefficients so that the predicted sample $\hat{x}(n)$ is being closed to the actual sample $x(n)$, is the method of least-squares. RLS which is a type of Least Square algorithm was explained in the previous Section. The error between $x(n)$ and the predicted value $\hat{x}(n)$ is

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^P \alpha_k x(n-k) \quad (2.29)$$

where α_k are obtained by minimization of the mean or total squared error, with respect to each of the coefficients. Time series modeling is a linear prediction technique that has been widely used in the area of biomedical signal processing. This has been explained in the next subsection.

2.4.1 Time Series Modeling

The signal $x(n)$ is modeled as a linear combination of its past values, and the present and past values of a “hypothetical” input, $u(n)$, to the system generating $x(n)$

$$x(n) = \sum_{k=1}^P a_k x(n-k) + G \sum_{l=1}^P b_l u(n-l) \quad (2.30)$$

where $b_0 = 0$ and a_k , b_l , and G are the parameters of the system. Equation 2.29 can be specified in the transform domain.

$$H(z) = \frac{X(Z)}{U(Z)} = G \frac{1 + \sum_{l=1}^Q b_l z^{-l}}{1 + \sum_{k=1}^P a_k z^{-k}} \quad (2.31)$$

There are three special cases of the model that are of interest:

1. All-zero moving average (MA) model : $a_k = 0$, $1 \leq k \leq P$
2. All-pole moving average (AR) model : $b_l = 0$, $1 \leq l \leq Q$

3. The pole–zero model given in Equation 2.30 is also known as the autoregressive moving average (ARMA) model.

In biomedical signal analysis, the all–pole model is the most widely used time series model. Therefore, in this work the AR modeling was applied on EMG signals.

2.4.2 AR Modeling

Modeling techniques such as autoregressive modeling (AR), also referred to as all–pole modeling, provide parameters which could potentially be correlated with the physiological system producing the signals. The AR model is a linear, second–moment stationary model [41]. Therefore the signal is modeled as a linear combination of its past values and some input:

$$x(n) = - \sum_{k=1}^P a_k x(n-k) + Gu(n) \quad (2.32)$$

where a_k are the AR coefficients and P is the model order. Model order indicates that for a given signal how many model parameters are sufficient to replicate the signal. Comparing the above Equation with Equation 2.29

$$x(n) = \sum_{k=1}^P \alpha_k x(n-k) + e(n) \quad (2.33)$$

It is evident from these two equations that AR coefficients are nothing but the predictor coefficients (i.e., $a_k = -\alpha_k$), and $Gu(n) = e(n)$. Therefore, the AR model of the current sample of the signal $x(n)$ is described as a linear combination of previous samples plus an error term, $e(n)$, which is independent of the past samples. P is the model order, where its value is very important.

The transfer function $H(z)$ of an AR process is given by $H(z) = \frac{X(z)}{U(z)}$, giving

$$H(z) = \frac{G}{1 + \sum_{k=1}^P a_k z^{-k}} \quad (2.34)$$

Then $H(z) = \frac{G}{A(z)}$, where $A(z) = 1 + \sum_{k=1}^P a_k z^{-k}$. The spectrum of the model can be estimated from Equation 2.35 as follows:

$$P_{AR}(\omega) = \frac{1}{\left| 1 + \sum_{k=1}^P a_k e^{j\omega k} \right|^2} \quad (2.35)$$

AR modeling techniques have been found to provide a sufficiently accurate representation for many different types of signals in many different applications. This is the reason for many researchers to use AR modeling as an approach of their signal processing. Another reason is that the structure of the all-pole model leads to fast and efficient algorithms for finding the all-pole parameters. However, AR modeling requires the fitted signal be stationary over a given interval, yet, adaptive segmentation techniques could possibly segment the signal into stationary components.

Another type of application that has been widely used in signal analysis, specially in speech analysis, is Cepstrum. Due to higher identification accuracy in speech recognition compared to AR, cepstrum analysis has been investigated in this work. The next Section will explain the cepstral analysis in detail [41][42].

2.4.3 Cepstrum Analysis

In this section, a class of nonlinear techniques referred to as *Cepstrum analysis* or *homomorphic deconvolution* will be explained. These methods have proved to be extremely effective and useful in certain applications such as speech processing [43]. In addition, they illustrate the considerable flexibility and sophistication, offered by discrete-time signal processing technologies [24][27].

In 1963, Bogert, Healy, and Tukey published a paper with an unusual title “The Frequency Analysis of Time Series for Echoes: Cepstrum, Pseudoautocovariance, Cross-Cepstrum, and Saphe Cracking” [24][44]. They observed that the logarithm of the power spectrum of a signal containing an echo has an additive periodic component, due to the echo. Thus, the Fourier transform of the logarithm of the power spectrum, should exhibit a peak at the echo delay. They called this function *Cepstrum*, interchanging the letters in the word spectrum because “In general, we find ourselves operating on the frequency side in ways customary on the time side and vice versa.” [24][44][45]. Another class of systems proposed by [24] was called homomorphic systems, which in the classical sense, they satisfy a generalization of the principle of superposition; i.e., input signals and their corresponding responses are

superimposed by an operation having the same algebraic properties as the addition. The concept of homomorphic filtering is very general and has been studied most extensively for the combining operations of multiplication and convolution because many signal models involve these operations. The transformation of a signal into its cepstrum is a homomorphic transformation, and the concept of the cepstrum is a fundamental part of the theory of homomorphic systems for processing signals that have been combined by convolution. The cepstrum has been proved as a valuable tool in speech coding and recognition applications, and has been extensively studied in the corresponding literature. The cepstrum of a signal is defined as the inverse Fourier Transform of the log power spectrum of the signal [46]. If all the poles of the $H(z)$ are inside the unit circle, the logarithmic transfer function, $\ln H(z)$, can be represented by the Laurent Expansion as follows [47]:

$$\ln H(z) = C(z) = \sum_{i=1}^{\infty} c_i z^{-i} \quad (2.36)$$

Differentiating both sides of the second equal sign, with respect to z^{-1} , and equating the coefficients of like powers with respect to z^{-1} , the following recursive relations will be derived [41][48].

$$\begin{aligned} c_1 &= -a_1 \\ c_n &= -a_n - \sum_{k=1}^{n-1} (1 - k/n) a_k c_{n-k} \quad \text{for } 1 < n \leq P \\ c_n &= - \sum_{k=1}^{n-1} (1 - k/n) a_k c_{n-k} \quad \text{for } n > P \end{aligned} \quad (2.37)$$

where a_n and c_n denote the n^{th} AR model and Cepstral coefficients respectively, P is the model order, and n is the samples.

Nevertheless, both AR modeling and Cepstrum analysis require the signal fitted be stationary over the given interval. But as explained in Chapter 1, biomedical signals such as EMG are nonstationary and these modeling can not be applied to them, unless the signals can be made stationary. Using the adaptive segmentation techniques explained in Section 2.2, the EMG signal, could be divided into stationary segments, and AR modeling and cepstral analysis can be applied to each stationary segment. These techniques as well as their result will be explained in more details in Chapter 3 [41][42]. Yet, applying techniques to the signal that do not need dividing the signal into stationary segments would be very useful. For

this reason the Wavelet Transform (WT) was also applied on the EMG signal. The result obtained from WT, compared to that of AR modeling and Cepstrum analysis made WT a better technique for classifying EMG into normal versus abnormal. Hence, WT will be explained briefly in Section 2.5 and in more detail in Chapter 4.

2.5 Wavelet Transform

In the previous Chapter, biomedical signals were shown to be nonstationary, and thus a nonstationary analysis tool must be identified to extract useful information from these signals. Yet these techniques may not necessarily be accurate and precise, in detecting the stationary boundaries, therefore techniques that could directly be applied to nonstationary signals are useful in a variety of applications. Therefore, Wavelet Transform has been used to address this issue in this work.

Wavelets are functions that satisfy certain mathematical requirements and are used to represent data or other functions. This idea is not new. Approximation using superposition of functions has existed since the early 1800 when Joseph Fourier discovered that he could superpose sines and cosines to represent other functions [49]. The fundamental idea behind wavelets is to analyze signals according to scale. Indeed, some researchers believe that using wavelets mean adopting a whole new mind-set or perspective in processing data [50][51]. To give the reader some back ground on wavelets, the Fourier Transform (FT) is compared with the WT [52].

The FT transforms the time domain signal to frequency domain by using sinusoidal basis function to approximate the original signal. There are many advantages for this kind of approximation, as signal can be analyzed for its frequency content. However, the FT representation has a major drawback due to using a sinusoidal basis function. Fourier sine and cosine functions are localized in frequency, but not in time. In other words, they stretch infinitely in time, in transforming to frequency domain, while the time information of the signal is lost. Therefore, with Fourier analysis, it is impossible to tell which frequencies appear at what time. As a result, FT can not be used to approximate a signal whose

properties change over time, i.e. nonstationary signal, since they result in poor estimation of sharp spikes.

Another way to see the time resolutional difference between the FT and WT is to look at the basis function coverage of the time–frequency plane [52]. In FT, the window plane is a fixed length window, which is simply a square wave, the square wave window truncates the sine and cosine basis function to fit a window of a particular width. Since a single window is used in the FT for all the frequencies, the resolution of the analysis is the same at all locations in time–frequency plane. Thus for many decades scientists were looking for more appropriate functions than the sines and cosines, which are the basis of the Fourier analysis, to approximate nonstationary signals.

To address this problem, a joint time–frequency representation is required. The Short Time Fourier Transform (STFT) is an intuitive modification of the Fourier transform to analyze nonstationary signals [53]. The basic idea behind the STFT is segmenting the signal by time–localized windowing, and performing Fourier transform to each segment at a time. The STFT maps a signal into a two dimensional time–frequency representation. The imprecision drawback comes from the fixed length time window, used to analyze the entire signal, regardless of frequency content of each segment. STFT is able to do wide–band frequency analysis using narrow window, or narrow–band frequency analysis using wide window, but not both simultaneously, once the window size is selected. STFT achieves some degree of compromise between time and frequency representation of signals. It provides information regarding, when and what frequencies occur in a signal with limited precision. The basis functions of the STFT provide a space–frequency resolution that is inadequate for capturing the wide variety of localized structures, and are common in biomedical signals. Additionally, this transformation is computationally complex and memory intensive, when computed for signals and images. Consequently, for biomedical signal analysis, a multiresolutional approach may be a good alternative to the STFT. Multiresolutional analysis (MRA) techniques can represent low and high frequency structures with different spatial resolutions, in a computationally efficient manner [53]. In order to isolate signal discontinuities, it is

ideal to have some very short basis functions. At the same time, in order to obtain detailed frequency analysis, one would like to have very long basis functions. To achieve this, one would like to have short high frequency basis functions and low frequency ones.

Wavelet algorithm process data at different scales or resolutions. If a large window is being used to analyze the signal, the gross features would be noticed, and if small window is being used small features would be extracted. "The result in wavelet analysis is to see both the forest and the trees, so to speak." [49]. This makes wavelet a very useful technique. Yet by using wavelet, one can use approximating functions that are contained nearly in finite domains making wavelet well-suited for approximating data with sharp discontinuities. Wavelet analysis techniques were independently researched and developed in the area of quantum physics, mathematics, electrical engineering and seismic geology [49]. Consequently, a series of applications which utilize these techniques have been created. These applications may include data compression, computer and human vision schemes, data denoising, feature extraction and many more [49].

The wavelet analysis procedure is to adopt a wavelet prototype function called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients.

Another issue with the FT is that, FT has a single basis functions which utilize sine and cosine, where WT has an infinite set of possible basis functions. Thus the wavelet analysis provides immediate access to information that can be obscured by other time-frequency methods such as Fourier Analysis [49].

While the FT maps a one dimensional series into a one dimensional sequence of coefficients, wavelet analysis maps it into a two dimensional array of coefficients. The extra dimension of information allows localizing signal in both time and frequency. But the two

dimensional time—scale representation is highly redundant, because of the non—orthogonal basis functions.

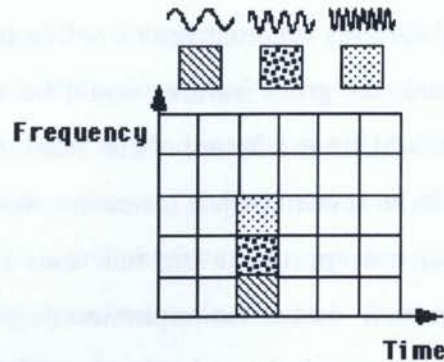


Figure 2.3: Fourier Basis Functions, time frequency tiles, and coverage of the time-frequency plane

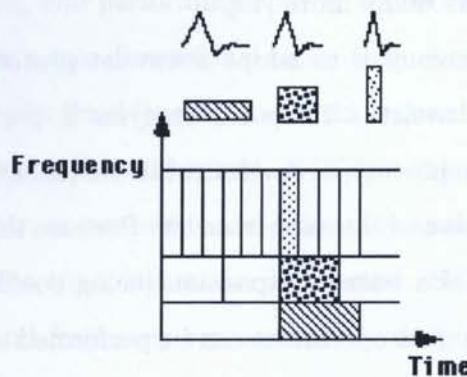


Figure 2.4: Daubechies Wavelet Basis Functions, time frequency tiles, and coverage of the time-frequency plane

The WT results in a series of coefficients in time, representing indirectly the instantaneous frequency of the signal in time [54]. When WT is computed restrictively in dyadic (power of two) scales (a) and positions (b) of the data series, the method is Discrete Wavelet Transform (DWT), on the other hand, if such restriction is not applied the WT is Continuous Wavelet Transform (CWT) [55]. The temporal resolution of the frequency and amplitude estimations

by the CWT are adequate to follow frequency and amplitude transients of the original signal in the time domain. Continuous analysis is often easier to be interpreted, since its redundancy tends to reinforce the traits and makes all information more visible, this is specially true of very subtle information. Thus, the analysis gains in readability, and in ease of interpretation what it loses in terms of saving space.

2.5.1 Continuous Wavelet Transform

In order to obtain the decomposition map (scalogram) of the signal $x(t)$, and to analyze the signal flexibly a size-variable window is needed to access accurate view of the signal, either in time or in frequency. The CWT gives exactly this achievement, which is shown below [56][57]:

$$W(a, b) = \frac{1}{\sqrt{a}} \int x(t) \psi^*\left(\frac{t-b}{a}\right) dt \quad (2.38)$$

where $a > 0$, and b are scale and translation parameters respectively, ψ denotes the basis function, often called mother wavelet, and $W(a, b)$ is the CWT of $x(t)$. Equation 2.38 can be interpreted as inner product of $x(t)$, the complex conjugate of the scaled, and translated version of the basis function (mother wavelet) ψ

$$W(a, b) = \int x(t) \psi_{(a,b)}^*(t) dt = \langle x(t), \psi_{(a,b)}^*(t) \rangle \quad (2.39)$$

where $\psi_{(a,b)} = \frac{1}{\sqrt{a}} \psi^*\left(\frac{t-b}{a}\right)$, $a > 0$ and $b \in \mathfrak{R}$ are real continuous variables. By increasing or decreasing a , the basis function will be fitted to a segment of $x(t)$; hence, a indirectly represents the frequency components of the signal. Fitting of CWT to the data results in calculation of a resemblance index between the signal $x(t)$, and the wavelet located at position b and scale a .

The mother wavelet used to generate all the basis functions is designed based on some desired characteristics associated with that function. The translation parameter b relates to the location of the wavelet function as it is shifted through the signal. Thus, it corresponds to the time information in the wavelet transform. On the other hand, the scale parameter, a , corresponds to the frequency information of the signal. This is useful in most of the

applications, since the high frequency (low scales) components of the signal do not last for long and mostly appears as short bursts, while low frequency (high scales), usually last for entire duration of the signal. The benefit of the algorithm is that the sampling rate could be changed, with accordance of the scale change, without violating the Nyquist Algorithm. Nyquist algorithm states that, the minimum sampling rate to reconstruct the original signal is 2 times the highest frequency of the signal. Therefore, as the scale goes higher (low frequencies), the sampling rate can be decreased thus reducing the number of computations.

2.5.2 Discrete Wavelet Transform

As explained in the last section, the transformation of the signal is another form of representing the signal, in which the information content of the signal would not change. The discrete wavelet transform are those wavelets that are discretely sampled, and is based on the sub-band coding, which yields to a fast computation of wavelet transform. Their implementation is easier than CWT, and is less time consuming [54].

One of the main differences between the CWT and the DWT is that, in CWT the signals are analyzed by using a set of basis functions, which are related to each other by simple scaling and translation. However, in DWT, the signal is passed through filters with different cutoff frequencies at different scales which is a time–scale representation of the signal. The two filters used in this technique are a lowpass filter and a highpass filter, often called filter banks [58].

DWT and Filter Banks

Filters are one of the most widely used signal processing functions. Wavelets can be realized by iteration of filters. The resolution of the signal is determined by the filtering operations. This is the measure of the amount of detailed information in the signal. The scaling however, is determined by upsampling and downsampling operations.

At each decomposition level, the half band filters produce signals spanning only half the frequency band, which results in doubling the frequency resolution since the uncertainty in frequency is reduced by half. Therefore, since the Nyquist theorem is 2 times the highest

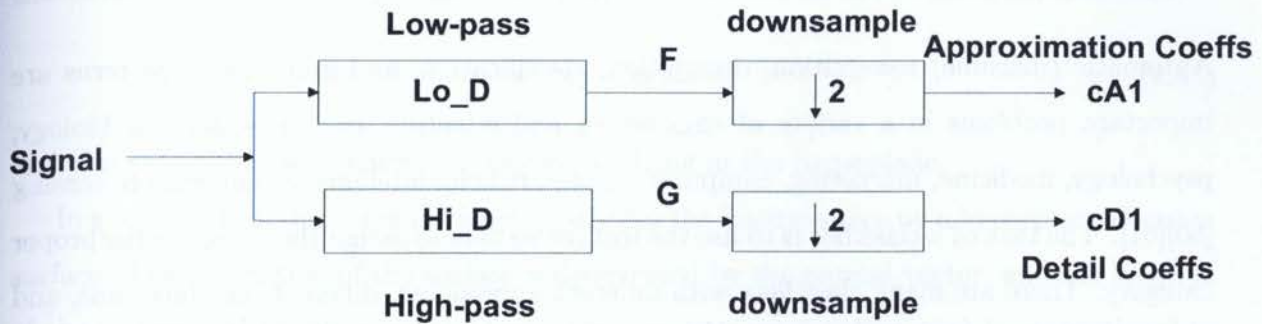


Figure 2.5: 1 Level Decomposition Level for DWT

frequency (ω) in the signal, the highest frequency is now half that value ($\omega/2$). This can be sampled at a frequency of ω radians thus discarding half the samples with no loss of information. This decimation by 2 halves the time resolution as the entire signal is now represented by only half the number of samples. Thus, while the half band low pass filtering removes half of the frequencies, and thus halves the resolution, the decimation by 2 doubles the scale. This approach will make a better time resolution at high frequencies, and a better frequency resolution at low frequencies. However most of the energy associated with a signal is located in the approximations. This allows us to roughly represent the signal by an approximation at any level, by omitting or reducing the detail values, without losing the information contained in the signal [54].

Inevitably the techniques mentioned (AR modeling, cepstrum analysis, and wavelet transform) need a scheme to validate their behaviors. As a consequence a very simple classifier has been used in this study to classify the signals into normal and abnormal, as well as to find the level of normality and abnormality in different subjects. The results of the classification will be explained in Chapters 3 and 4. The next Section will explain a type of supervised classifier (Linear Discriminant Analysis) in details.

2.6 Classification

Automatic (machine) recognition, description, classification, and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, and remote sensing [60][61]. The task of a classifier is to use the feature vectors to assign the object to the proper category. There are many classifiers with different approaches, different cost functions, and different algorithms. In order to choose the best classifier that best discriminates the different classes, the type of data and the amount of information existed in the data, is very important.

Classification can be categorized as: supervised and unsupervised. The supervised algorithm is an algorithm that a teacher provides with a category label, or cost of each pattern in the training set, and the purpose of this algorithm is to reduce the costs of these patterns [62]. Conversely in unsupervised learning or clustering, there is no explicit teacher, and grouping is being formed by the system [62][63]. Since the classes are defined in this thesis i.e. Normal/Abnormal, the supervised algorithm is being used. One of the supervised algorithm is the Linear Discriminant Analysis (LDA), which will be briefly explained in the next subsection and in more detail in Chapters 3 and 4.

2.6.1 Linear Discriminant Function

The main objective of Linear Discriminant Functions is to separate the classes as much as possible. A discriminant function that is a linear combination of the components of x can be written as

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 \quad (2.40)$$

where \mathbf{w} is the weight vector and w_0 is the bias. In case of two classes the linear classifier has the following decision rule: Decide w_1 if $g(\mathbf{x}) > 0$ and w_2 if $g(\mathbf{x}) < 0$. Thus \mathbf{x} is assigned to w_1 if the inner product $\mathbf{w}^T \mathbf{x}$ exceeds the threshold w_0 , and decide w_2 otherwise. if $g(\mathbf{x}) = 0$, \mathbf{x} can ordinarily be assigned to either classes. The equation $g(\mathbf{x}) = 0$ defines the decision surface that separates points assigned to w_1 from points assigned to w_2 . When $g(x)$ is linear,

this decision surface is a hyperplane. If \mathbf{x}_1 and \mathbf{x}_2 are both on the decision surface, then

$$\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2) = 0 \quad (2.41)$$

and this shows that \mathbf{w} is normal to any vector lying in the hyperplane.

In general, linear discriminant function divides the feature space by a hyperplane decision surface. The orientation of the surface is determined by the normal vector, and the location of the surface is determined by the bias. The discriminant function $g(\mathbf{x})$ is proportional to the signed distance from \mathbf{x} to the hyperplane, with $g(\mathbf{x}) > 0$ when \mathbf{x} is on the positive side, and $g(\mathbf{x}) < 0$ when \mathbf{x} is on the negative side.

In a two-category linearly separable problem, there is a set of n samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ some labeled w_1 and some labeled w_2 . This is being done to determine the weights \mathbf{w} in a linear discriminant function $g(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$.

A sampled \mathbf{x}_i is classified correctly if $\mathbf{w}\mathbf{x}_i > 0$ and \mathbf{x}_i is labeled w_1 , or if $\mathbf{w}\mathbf{x}_i < 0$ and \mathbf{x}_i is labeled w_2 . Therefore, from this it can be concluded that in the two-category case all samples of w_2 could be replaced by their negatives, and therefore the class label can be eliminated, in which $\mathbf{w}\mathbf{y}_i > 0$ for all of the samples. \mathbf{y}_i is the normalized sample. This is often called the solution or normalization sample. In order to find the solution to the set of inequalities, a criterion function has to be minimized [62].

2.6.2 Linear Discriminant Analysis

One of the methods that could be used for discriminating of two or more classes is the Linear Discriminant Analysis (LDA). LDA produces a discriminant function that maps the input into the classification space. LDA searches for those vectors in the underlying space that best discriminate among classes, rather than those that best describe the data [64]. The goal of LDA is to seek a transformation matrix W that in some sense maximizes the ratio of the between-class scatter to the within-class scatter. A within-class scatter matrix S_w is defined as:

$$S_w = \sum_{i=1}^c \sum_{x \in C_i} (x - m_i)(x - m_i)^t \quad (2.42)$$

where c is the number of classes, C_i is a set of data belonging to the i^{th} class, and m_i is the mean of the i^{th} class. The within-class scatter matrix represents the degree of scatter within classes as a summation of covariance matrices of all classes. A between-class scatter matrix S_B is defined as:

$$S_B = \sum_{i=1}^c n_i (m_i - m)(m_i - m)^t \quad (2.43)$$

The objective of LDA is to seek a transformation matrix W that in some sense maximize the ratio of the between-class scatter and the within-class scatter.

$$J(W) = \frac{|W^t S_B W|}{|W^t S_W W|} \quad (2.44)$$

LDA creates a linear combination of those which yields the largest mean differences of the desired classes [65][66].

Figure 2.6 shows an example of between class scatter and within class scatter. One of the

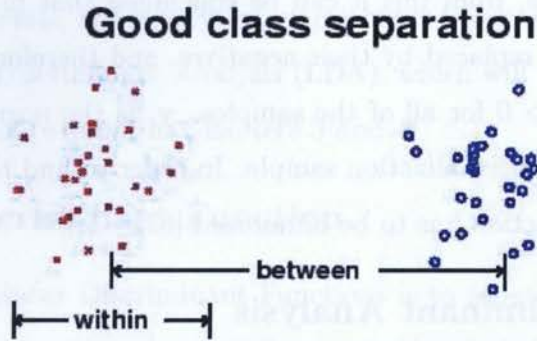


Figure 2.6: An example of between-class scatter over the within-class scatter in LDA

software packages that can be used for discriminant analysis is the Statistical Package for Social Science (SPSSTM) Software. The classification analysis of this thesis has been done with this software [67].

2.7 Motivation Behind this Study

The study of sleep recording is often frustrating due to the long length of the signal, as well as night time recording procedures. Recently, there has been considerable emphasis in the therapy of sleep abnormalities due to its effect on people's lives.

Computer aided signal processing saves time, standardizes the measurements, and enables the extraction of features which could not have been extracted, otherwise. Therefore, considerable amount of work has been carried out in the time and frequency domain for classification of EMG. However, none of these works have been done on EMG in REM sleep. As a consequence the use of sleep EMG represents a unique methodology, and can be used to detect RBD. This involves a novel clinical dataset. One of the techniques that will be explained is the quantitative signal analysis in the frequency domain, using classical power spectrum analysis, which has been well documented over the past decade. Another technique is the application of wavelet transform on EMG, in REM sleep, which will be explained in the following Chapter. Wavelet transform is a tool for the analysis of transient, nonstationary, or time-varying phenomena. The block diagram of the proposed method is shown in Figure 2.7.

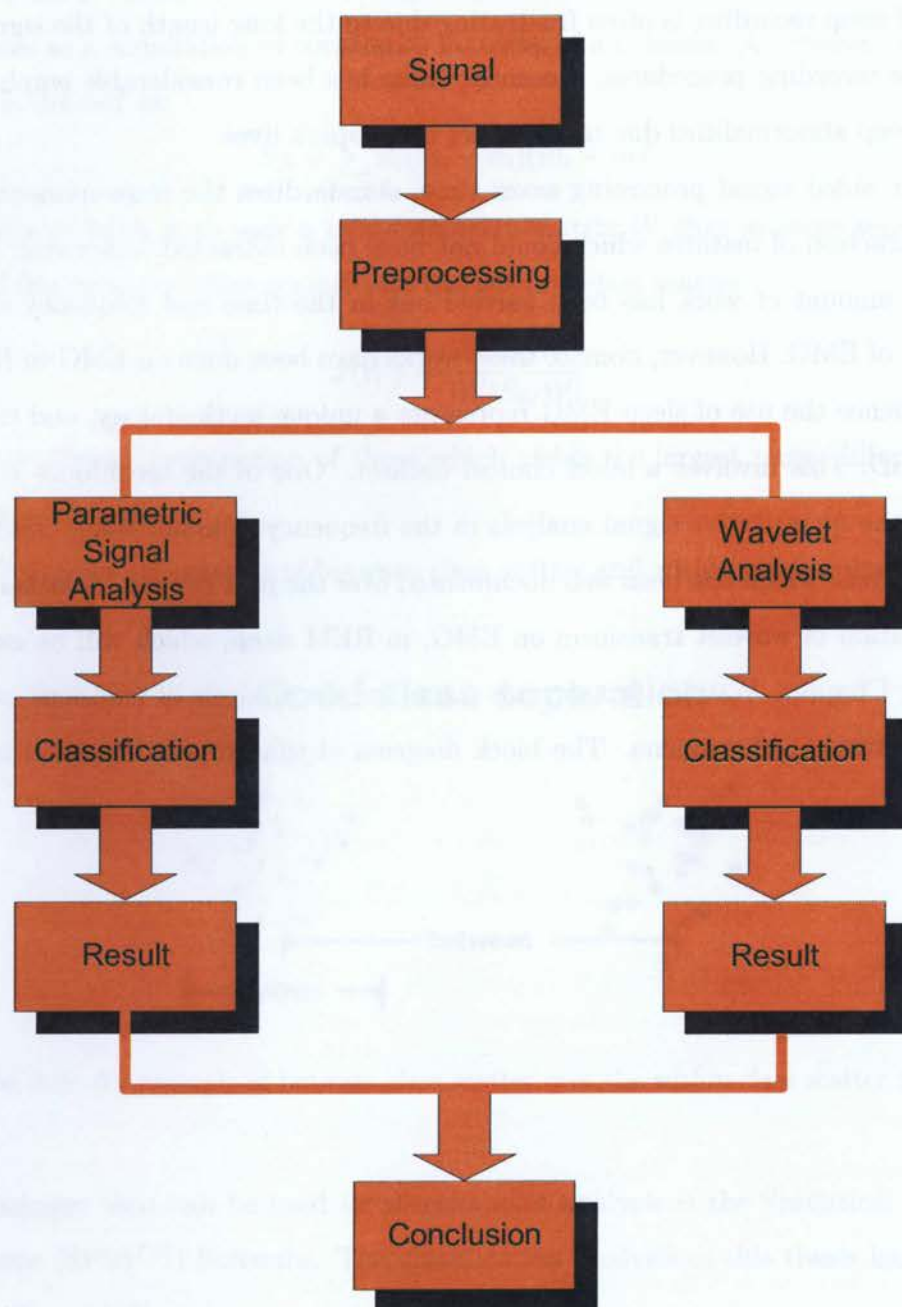


Figure 2.7: The Proposed Method

Chapter 3

Parametric Signal Analysis of EMG in Sleep

3.1 Introduction

As discussed in Chapter 1, the digitization of biomedical signals paves the way for a multitude of new applications. One of these applications is Computer-Aided Diagnosis (CAD), in which a computer program can act as “a second pair of eyes” to assist the physician with diagnostic results [53]. The result of the CAD system may be used to complement doctor’s initial findings, as well as, provide useful information to clinicians in the form of an easily applicable measure of disease activity sensitive to early neurodegeneration, and treatment response. In general, the CAD aims to improve the quality of diagnoses which would result in superior form of health care. For such an application to be realized, a feature extraction and classification scheme must be identified which provides results that discriminate between pathologies of the biomedical signals. As a way to realize this application, this Chapter will focus on feature extraction algorithms using AR modeling and Cepstrum analysis, and the next Chapter would focus on feature extraction algorithm using Wavelet Transform.

As there is no other reported works for feature extraction and classification of EMG in REM Sleep, the research discussed in this Chapter is, the first reported work in the area.

The following sections will cover the methods and techniques that will be used to classify

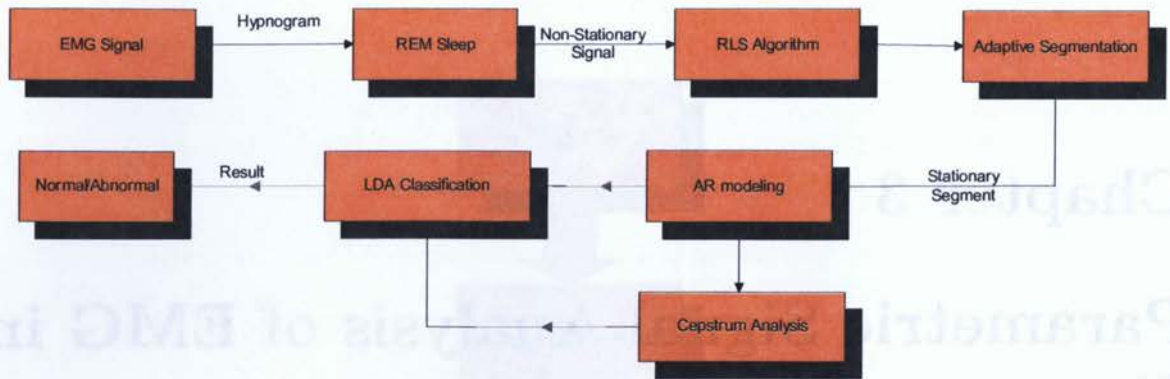


Figure 3.1: Block Diagram of the AR modeling and Cepstrum Analysis

the EMG signal in REM sleep, normal vs abnormal. First, the Adaptive segmentation technique would be explained which is being used to divide the signal into stationary segments, followed by a brief explanation of AR modeling and Cepstrum analysis. Next these coefficients would be used as features to the classifier. At last, a description of the classification scheme will also be included, followed by the experimental results. Figure 3.1 shows the algorithms and techniques for this Chapter.

3.2 Adaptive Signal Processing

As explained in the previous Chapters, biomedical signals are nonstationary, since their statistical properties change over time. This cause challenges in classifying the signal using well developed spectrum analysis methods. EMG is a nonstationary signal that its properties, such as spectrum, changes over time. However, it can be considered that EMG is locally stationary over short time intervals [33][41][68]. There are two types of techniques associated with this. One is fixed segmentation, and the other is adaptive segmentation. The fixed segmentation uses a fixed-size window for segmenting the signal. This technique is mostly used in applications such as STFT. Although fixed segmentation is simple, its performance is not sufficient since the window size may not be accurate. If the window length is too long then the sudden variation of the signal may not be captured, and if the window

length is too short poor estimation of low frequency components may occur. Therefore, this method cannot guarantee stationarity, as well as, appropriate length of each chosen window. Conversely, adaptive signal processing involves the use of optimum and statistical signal processing techniques to design signal processing systems that can modify their characteristics, during normal operation (usually in a real time). Moreover they can, achieve a clearly predefined application dependent objective, and create signal processing tools which would be able to monitor time variations of statistical properties of the signals, and divide them into locally stationary components [28]. As explained in Section 2.2 there are many adaptive signal processing techniques, used in the literature. In this work, the RLS algorithm has been applied on the EMG signal, and stationary segment boundaries have been marked. There are many advantages for RLS algorithm, compared to other algorithms [39]. Some of these benefits are listed below:

- Fast rate of convergence, compared with other algorithms such as LMS (Least Mean Square).
- Good tracking capability which insures numerical stability [40].
- Better steady state approximation of tap weights, since infinite memory RLS averages the value of each tap weight.

The details of RLS algorithm for segmentation are discussed below.

3.2.1 Recursive Least Square Algorithm

The principle of adaptive filters is that they can improve their performance, during normal operation, by learning the statistical characteristics through processing current signal observations. The goal of any adaptive filter such as RLS is to “find and track” the optimum filter corresponding to the same signal operating environment, with complete knowledge of the required statistics. The performance of such adaptive filters are evaluated using the concept of stability, speed of adaptation, quality of adaptation, and tracking capabilities. Therefore,

the distinguishing feature of the adaptive filters is that they can modify their response to improve their performance during operation, without any intervention from the user [38].

RLS algorithm has been already explained in Section 2.3, however, the summary and initialization of the algorithm is shown below.

1. Select an initial tap-weight vector at time $n = 0$, usually $\underline{w} = 0$ and $\underline{P}(0) = \delta^{-1}I$ where δ is a small constant and I is the identity matrix.

2. For $n = 1, 2, \dots, N$

Receive an input sequence $\underline{x}(n)$ for each time sample, where $\underline{x} = [x(n)x(n-1)x(n-2)\dots x(n-N)]$

Compute the error $\alpha(n) = d(n) - \underline{w}^T(n-1)\underline{x}(n)$

Compute the inverse autocorrelation function $\underline{P}(n) = \underline{P}(n-1) - \underline{K}(n)\underline{x}^T(n)\underline{P}(n-1)$

Compute the Kalman gain factor $\underline{K}(n) = \frac{\underline{P}(n-1)\underline{x}(n)}{1 + \underline{x}^T(n)\underline{P}(n-1)\underline{x}(n)}$

Update the weight vector $\underline{w}(n) = \underline{w}(n-1) + \underline{K}(n)\alpha(n)$

3. Calculate the sum error squared s using the two adjacent tap-weight vector $s(n) = \|\underline{w}(n) - \underline{w}(n-1)\|^2$.

A procedure based on this idea was developed in the present study to divide the EMG signal into locally stationary components. This procedure operates in two steps [31]: The first step is the primary boundary detection algorithm in which the beginning of the stationary boundary would be marked; and the second step is the decision process on the location of the final segment boundary in which the end part of the stationary boundary is marked. Since having a segment with few number of samples might result in under modeling, a minimum of 120 samples is being used as the minimum segment length L_{min} [31]. The order of the filter was set to be 7 in order to be low enough to detect transitional changes in the input, and also to provide fast convergence. The desired response was the original signal delayed by 5 samples. The delay of 5 samples was chosen based on [31]. The latter work was done on vibroarthrography (VAG) signals in which the autocorrelation functions of typical VAG

signals were observed to drop off rapidly after about five samples; hence the authors assumed the signal with this delay to be independent of the original signal. Similar idea was used in this work because of the similarities of the two signals. Following would be the procedure of segmenting the signal:

- Define a threshold value, three times the standard deviation of the $s(n)$ vector [31].
- Compare each element of the $s(n)$ vector with the threshold, and store all the points above the threshold as the Primary Segment Boundary (PSB), $PSB = [a_0, a_1, \dots, a_P]$.
- Compare the adjacent elements of PSB to each other, and if the difference of the two values is less than the minimum length L_{min} keep the first element and compare it to the next one.

If $a_i - a_{i-1} \geq L_{min}$, keep a_i as one of the segment boundaries and check the next element, if $a_i - a_{i-1} < L_{min}$, remove a_i and continue.

- Mark the end boundaries.

Thus this algorithm, divides the EMG signal into few locally stationary components, and each segment is then considered as a separate signal. The number of segments in both the normal and abnormal signals varied, since they all have different characteristics and behaviors.

After dividing the signal into stationary components different algorithms could be applied. In the next two sections AR modeling and cepstral analysis will be applied to each of these segments and the classification result of these will be explained at last. Figure 3.2 shows the block diagram of the RLS algorithm.

3.3 AR Modeling

Computer aided EMG processing saves time, standardizes the measurements and enables the extraction of features which could not have been extracted, otherwise. Time domain

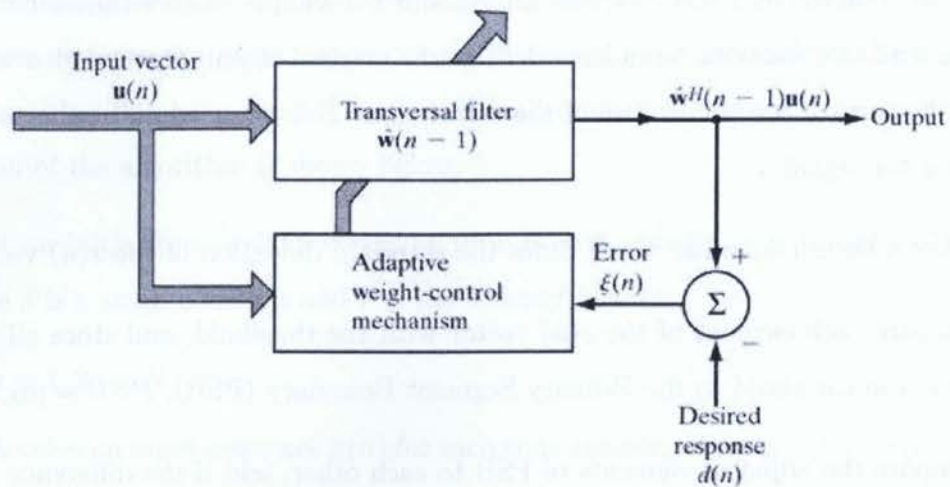


Figure 3.2: Block Diagram of RLS Algorithm

techniques that rely on extracting features for classification directly from the EMG waveform, are somewhat difficult to measure automatically, whereas their manual measurement depends on one's intuition. Alternatively, signal analysis in the frequency domain reveals the frequency characteristics of the EMG signals. However, this analysis based on the FT power spectrum estimates, suffers from reduced frequency resolution and spectral leakage effects. The limitations inherent in the classical power spectrum density (PSD) estimation methods can be overcome, (or improved), by parametric modeling techniques such as AR or autoregressive moving average (ARMA). These techniques give better PSD estimators, higher spectral resolution, and avoid spectral leakage effects compared to nonparametric or classical spectral estimation techniques [41].

Parametric modeling is a typical method in dealing with random signals, as long as the signals are stationary. The basic idea for parametric modeling is that the present value of model output is assumed to be the linear combination of several past values of model output plus the linear combination of present and past values of model input. For such model either an adaptive model can be used, in which the values of the parameters gets updated at the

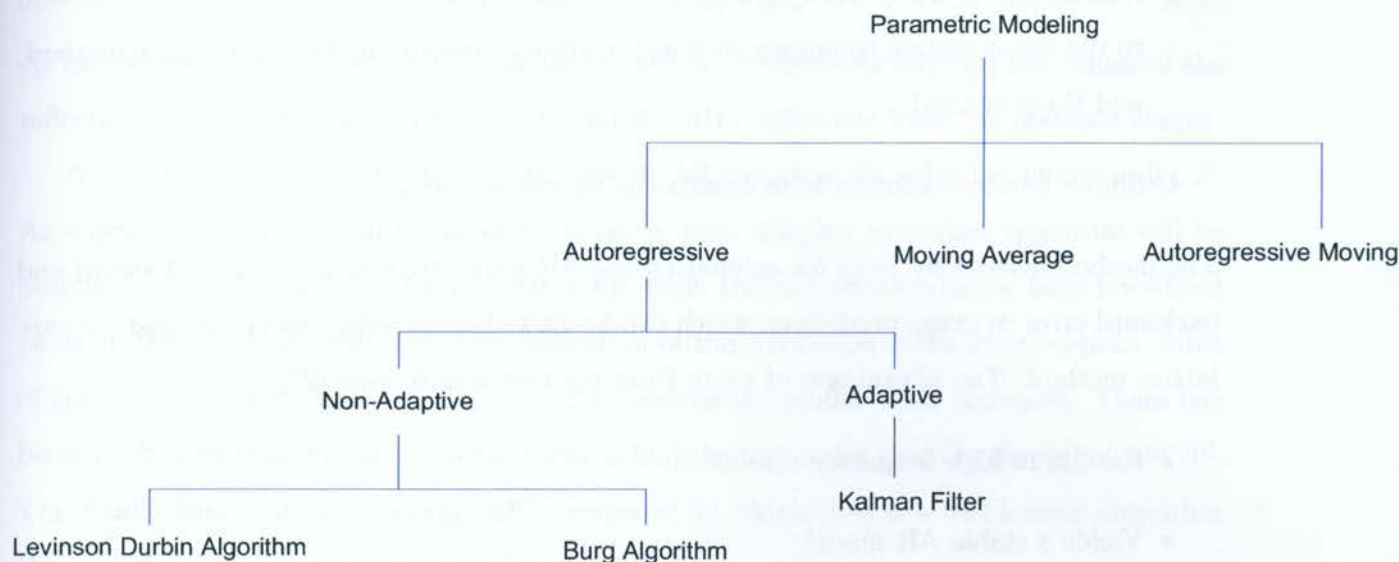


Figure 3.3: Techniques for Parametric modeling, using a rational transfer function and algorithms for autoregressive parameter estimation

arrival of new data sample, or a non-adaptive model can be used, in which the signal has to be divided into stationary segments, i.e. sufficiently short segments [68]. The effectiveness of the algorithm used to estimate the values of the parameter has a key role in parametric modeling techniques [68]. Levinson–Durbin algorithm and the Burg algorithm are the two most popular algorithms for non-adaptive modeling, and Kalman filtering is an adaptive algorithm modeling. Figure 3.3 is summarization of these algorithms [68].

The relative simplicity and reliability of the Burg algorithm has made non-adaptive AR modeling, by far, the most popular method of time series analysis. For the information of the reader, this work used the application of AR modeling for several reasons:

1. Some biomedical signals such as speech signal have an underlying autoregressive structure.
2. Generally, any signal can be modeled as an AR process, as long as, an appropriate model order is selected.

3. Availability of many algorithm to estimate the model parameters, and find a solution to the linear system equations such as the autocorrelation method, covariance method, and Burg method.
4. Shows better resolution than traditional Fourier spectrum.

The method derived by Burg for estimating the AR parameters minimized the forward and backward error in linear predictors, which can be viewed as an order–recursive least–square lattice method. The advantages of using Burg method is as follows [27].

- Results in high frequency resolution.
- Yields a stable AR model.
- Computes more efficiently.

As stated before, the Burg algorithm is based on minimizing the sum of the squared forward and backward prediction error. The cost function is defined as follows [1][27]

$$\xi = \sum_{n=m+1}^N f_m^2(n) + b_m^2(n) \quad (3.1)$$

where $f_m(n)$ and $b_m(n)$ are forward and backward prediction error for order of m respectively and N is the length of the input data.

$$f_m(n) = f_{m-1}(n) - \gamma_m b_{m-1}(n-1) \quad (3.2)$$

$$b_m(n) = b_{m-1}(n-1) - \gamma_m f_{m-1}(n) \quad (3.3)$$

The above two equations are the recursion equations for forward and backward prediction error updates, and γ is the reflection coefficient. The reflection coefficient is being used to calculate the AR coefficients. Equation 3.4 calculates the reflection coefficient and Equation 3.5 finds the AR parameters using the γ values.

$$\gamma_m = 2 \frac{\sum_{n=m}^{N-1} f_{m-1}(n) b_{m-1}(n-1)}{\sum_{n=m}^{N-1} [f_{m-1}^2(n) + b_{m-1}^2(n-1)]} \quad (3.4)$$

$$a_{m,k} = a_{m-1,k} - \gamma_m a_{m-1,m-k} \quad (3.5)$$

As can be seen from Equation 3.5, calculating the AR coefficients requires the values of the reflection coefficients of the current stage and the AR coefficients from the previous stages.

Another very important aspect of the use of AR model is the selection of the order P . As a general rule, if a small model order is being used, a highly smoothed spectrum will be obtained. And if a model order selected is too high, the risk of introducing false low-level peaks in the spectrum will exist. One indication of the AR model is the mean-square value of the residual error. This residual error decreases as the model order increases. There has been much work done by various researchers to find the best order that fits the signal [69][70]. The Akaike Information Criterion (AIC) proposed by Akiake [69] is a well known algorithm that is based on selecting the order that minimizes

$$AIC(P) = \ln \hat{\sigma}_{wP}^2 + 2P/N \quad (3.6)$$

where $\hat{\sigma}_{wP}^2$ is the estimated variance of the linear prediction error. As the order increases the $\hat{\sigma}_{wP}^2$ decreases and therefore, $\ln \hat{\sigma}_{wP}^2$ decreases, however, $2P/N$ increases, with an increase in P . Yet at some P the minimum value could be obtained [27][69].

Therefore, in general, the AIC tries to find the order that minimizes the prediction error. Hence, that order could be used as a model order to estimate the model parameters. Yet the signal is required to be stationary. In this work AIC was applied on each stationary segment for each signal. After calculating the AIC values for each segment for different model orders, and averaging all the AIC values for each signal the minimum AIC was obtained. Figure 3.4 shows AIC for these signals. Figures 3.5 and 3.6 show one segment of the normal and abnormal subject respectively, followed by the the spectrum of the signal. AR coefficients are shown, as well followed, by the model spectrum. As can be seen in both of these figures the model spectrum matches the spectrum of the signal, and this could be another proof that the model order used, was appropriate.

The model order based on this algorithm was chosen to be 25. Accordingly 25 AR coefficients were calculated for each segment; and were used as features to classify the signal into normal and abnormal. This has been shown in Section 3.5.

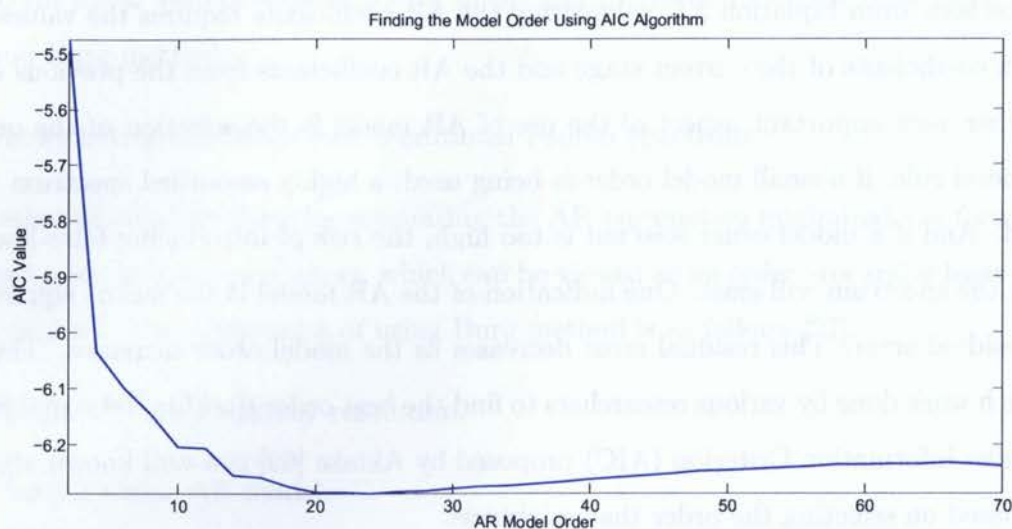


Figure 3.4: Finding the optimum model order, using the AIC criterion

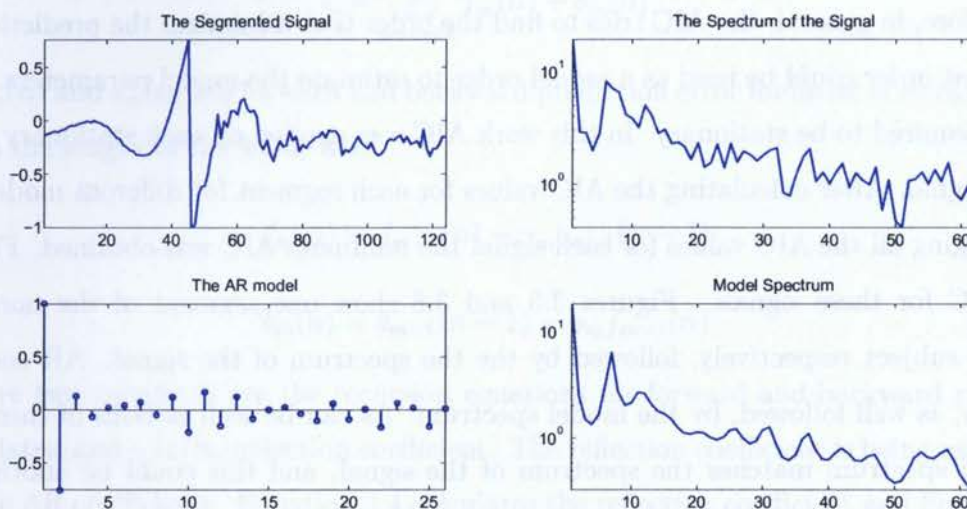


Figure 3.5: Model spectrum of the Normal Signal

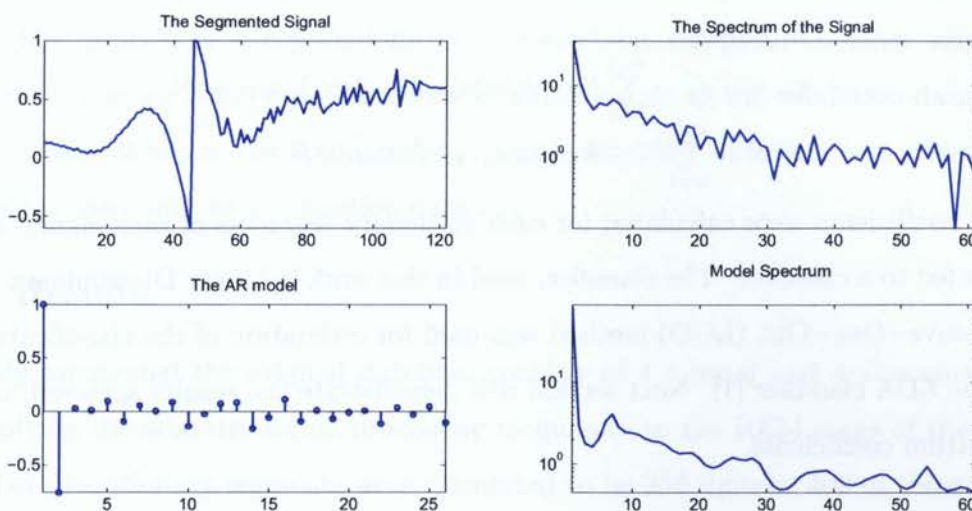


Figure 3.6: Model spectrum of the Abnormal Signal

Another type of application that has been widely used in signal analysis, specially in speech analysis, is cepstrum. Due to higher identification accuracy in speech recognition compared to AR as well as the non-linearity involved, Cepstrum analysis has been investigated in this work. The next section explains the cepstral analysis in detail [41].

3.4 Cepstrum Analysis

The motivation of using cepstrum as a diagnostic tool in this study was due to their documented higher identification accuracy in speech recognition, compared to the AR coefficients and the non-linear mapping provided [71][72]. Cepstral analysis have been applied to speech recognition for a long time where it has been proved to be a valuable tool, and has been extensively studied in the corresponding literature [24][44][46][73]. As shown in the previous Chapter and again in Equation 3.7, 25 cepstrum coefficients (c_1 to c_{25}) were directly

calculated from the AR coefficients (a_1 to a_{25}).

$$\begin{aligned} c_1 &= -a_1 \\ c_n &= -a_n - \sum_{k=1}^{n-1} (1 - k/n) a_k c_{n-k} \quad \text{for } 1 < n \leq P \\ c_n &= - \sum_{k=1}^{n-1} (1 - k/n) a_k c_{n-k} \quad \text{for } n > P \end{aligned} \quad (3.7)$$

These coefficients were calculated for each stationary segments of each signal, and then they were fed to a classifier. The classifier, used in this work is Linear Discriminant Analysis (LDA). Leave-One-Out (LOO) method was used for estimation of the classification accuracy of the LDA classifier [1]. Next section will demonstrate the results achieved, using AR and Cepstrum coefficients.

3.5 Time-Series Analysis of Sleep EMG

Automatic (machine) recognition, description, classification, and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as: biology, psychology, medicine, marketing, computer vision, artificial intelligence, and remote sensing [60][61].

The motivation behind using pattern classifier is that the coefficients achieved from AR modeling and Cepstrum analysis are abstract coefficients. Abstract coefficients are coefficients that do not have meaning by themselves, and a classifier has to be used to judge their behavior. Aside from distinct classification, the level of abnormality and normality is of great importance for neurologists and sleep specialists as this may represent a signal that can be used to follow disease progression or response to therapy.

One of the simplest classifier used, is the Linear Discriminant Analysis (LDA). LDA produces a discriminant function that maps the input into the classification space. LDA searches for those vectors in the underlying space that best discriminate among classes, rather than those that best describe the data [64]. The goal of LDA is to seek a transformation matrix W that maximizes the ratio of the between-class scatter to the within-class scatter.

One of the techniques that could be used to validate the behavior of the classifier is the LOO method. The idea behind this method is to exclude one of the available samples,

and train the classifier based on the rest of the samples, then test the data with that one sample. This procedure is repeated for all the available samples. In other words, LOO involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as a validation data.

3.5.1 Results

As already mentioned the original database consists of 4 normal and 4 abnormal signals. After applying the adaptive signal processing techniques to the REM stage of these signals the number of stationary segments were calculated to be 956 signals. 409 of these segments were marked as abnormal and 547 of them were marked as normal. To test the performance of the proposed feature extraction and classification scheme, all the stationary segments were used. The objective of the proposed system for the EMG signals were as follows:

1. To validate the techniques used in this work such as AR modeling and Cepstrum Analysis.
2. To find the level of normality and abnormality in different signals.
3. To find the disease progression in different subjects and/or their response time to therapy.
4. To prevent subjects from Parkinson Disease by studying their neuromuscular behavior in REM Sleep.

3.5.2 Classification Result

Classification using LDA was performed using the statistical analysis package SPSSTM (Statistical Package for the Social Science) [67]. The following subsection will present the classification results as a confusion matrix, which shows how many signals were correctly classified and how many signals were misclassified. The results of LDA and LOO for AR modeling are

shown in Tables 3.1 and 3.2, and the results of LDA and LOO for Cepstrum Analysis are shown in Tables 3.3 and 3.4 respectively.

From Table 3.1, LDA classification results of EMG in REM Sleep are shown, indicating that the correct classification rates were 89.8% (normal as normal) and 69.7% (abnormal as abnormal). Misclassification rates were computed to be 10.2% (normal as abnormal) and 30.3% (abnormal as normal). An overall classification rate of 81.2% was achieved for the EMG in REM sleep using AR modeling, and an average misclassification rate of 18.8% was achieved. From Table 3.2, LOO classification results of EMG in REM Sleep are shown, indicating that the correct classification rates were 88.7% (normal as normal) and 68.5% (abnormal as abnormal). Misclassification rates were computed to be 11.3% (normal as abnormal) and 31.5% (abnormal as normal). An overall classification rate of 80.0% was achieved for the EMG in REM sleep using AR modeling, and an average misclassification rate of 20.0% was achieved.

As shown in Table 3.3, the LDA classification rates of the EMG signal in REM sleep using Cepstrum were: 86.5% (normal as normal) and 68.0% (abnormal as abnormal). Misclassification rates were computed to be 13.5% (normal as abnormal) and 32.0% (abnormal as normal). An overall classification rate of 78.6% was achieved for the EMG in REM sleep using Cepstrum Analysis and an average misclassification rate of 21.4% was achieved.

However, the result of LOO classification in Table 3.4 indicates 86.3% (normal as normal) and 67.5% (abnormal as abnormal) were correctly classified. Misclassification rates were computed to be 13.7% (normal as abnormal) and 32.5% (abnormal as normal). An overall classification rate of 78.2% was achieved for the EMG in REM sleep using cepstral analysis and an average misclassification rate was 21.8%. Figures 3.7 to 3.10 show the distribution of normality and abnormality in different subjects. In these Figures N and A refers to Normal and Abnormal respectively. This is a valuable asset for doctors and neurologists since this may represent a signal that can be used to follow disease progression or response to therapy.

In order to prove the classification results, first, 3 dominant AR coefficients and 3 dom-

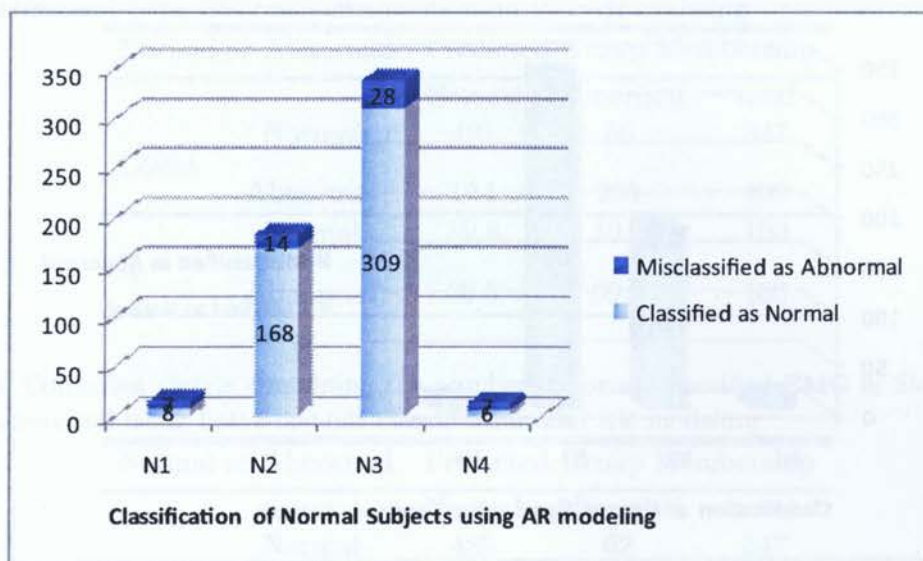


Figure 3.7: The level of Normality in Normal Subjects using AR modeling

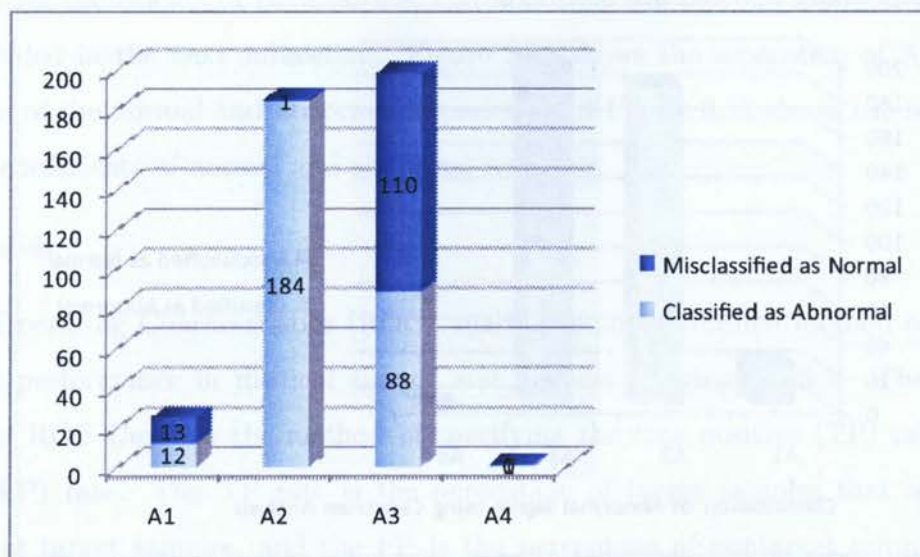


Figure 3.8: The level of Abnormality in Abnormal Subjects using AR modeling

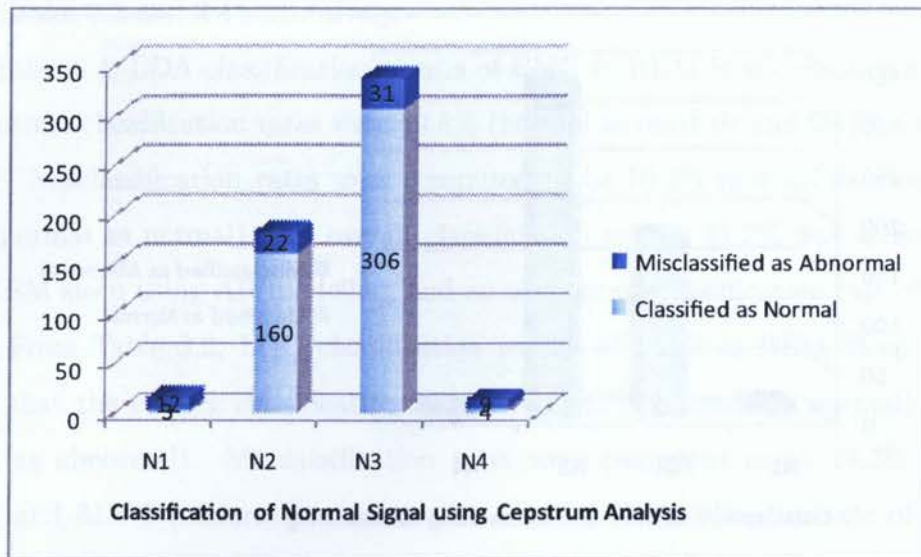


Figure 3.9: The level of Normality in Normal Subjects using Cepstrum

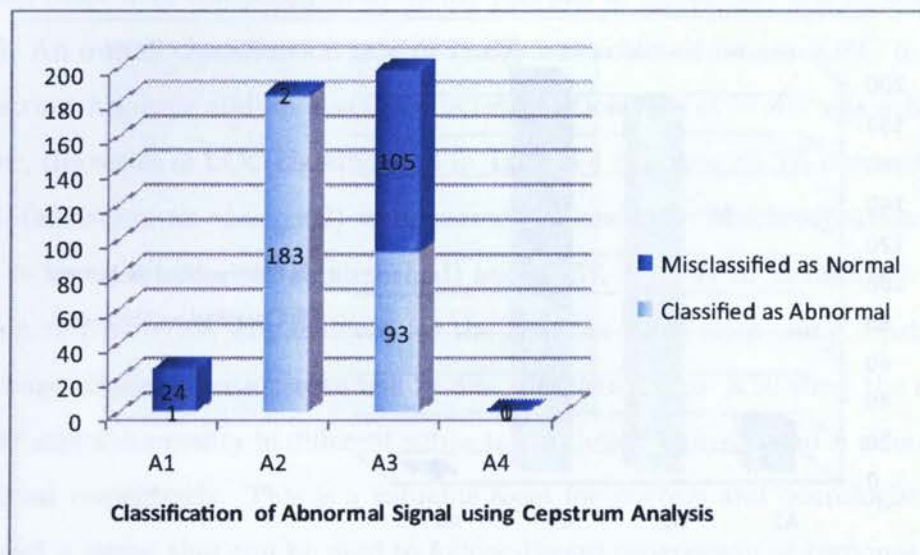


Figure 3.10: The level of Abnormality in Abnormal Subjects using Cepstrum

Table 3.1: Confusion matrix containing the number of correct classified EMG in Sleep as either normal or abnormal using Discriminant classification and AR modeling

Normal or Abnormal		Predicted Group Membership		
		Normal	Abnormal	total
Count	Normal	491	56	547
	Abnormal	124	285	409
%	Normal	89.8	10.2	100
	Abnormal	30.3	69.7	100

Table 3.2: Confusion matrix containing the number of correct classified EMG in Sleep as either normal or abnormal using Leave-one-out classification and AR modeling

Normal or Abnormal		Predicted Group Membership		
		Normal	Abnormal	total
Count	Normal	485	62	547
	Abnormal	129	280	409
%	Normal	88.7	11.3	100
	Abnormal	31.5	68.5	100

inant cepstral coefficients of both normal and abnormal are plotted, followed by the ROC curve included in the next subsection. Figure 3.11 shows the separation of AR modeling coefficients of the normal and abnormal segments, and Figure 3.12 shows the separation of cepstrum coefficients of normal and abnormal segments.

ROC Curve

Receiver Operating Characteristics (ROC) analysis is an established method of measuring diagnostic performance in medical signals and medical imaging studies. The basic idea behind the ROC Curve is the method of specifying the true positive (TP) rate and false positive (FP) rate. The TP rate is the percentage of target samples that are correctly classified as target samples, and the FP is the percentage of nontarget samples that are incorrectly classified as target samples. An ROC curve is a plot of operating points showing the possible tradeoff between a classifier's TP rate versus its FP rate. The TP is often called

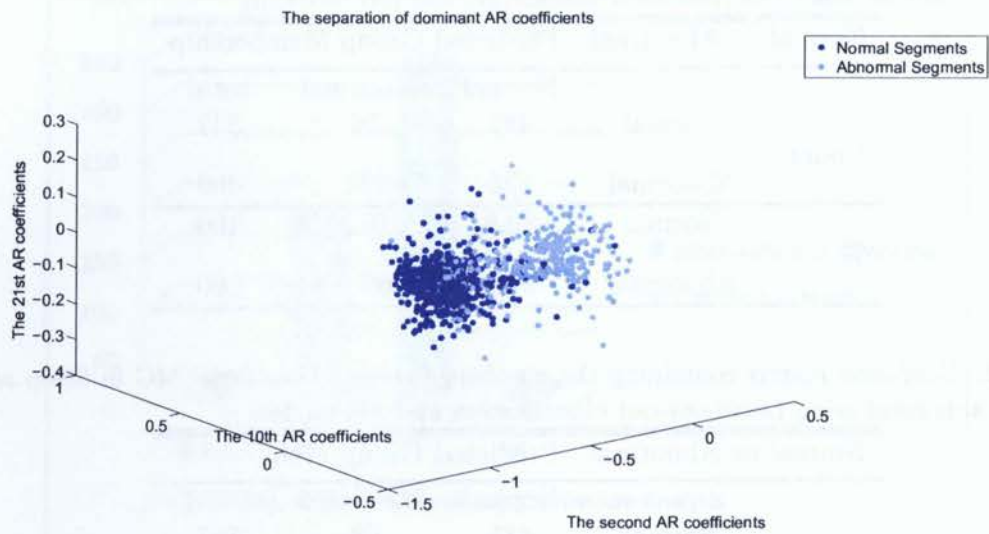


Figure 3.11: The Separation of Dominant AR coefficients

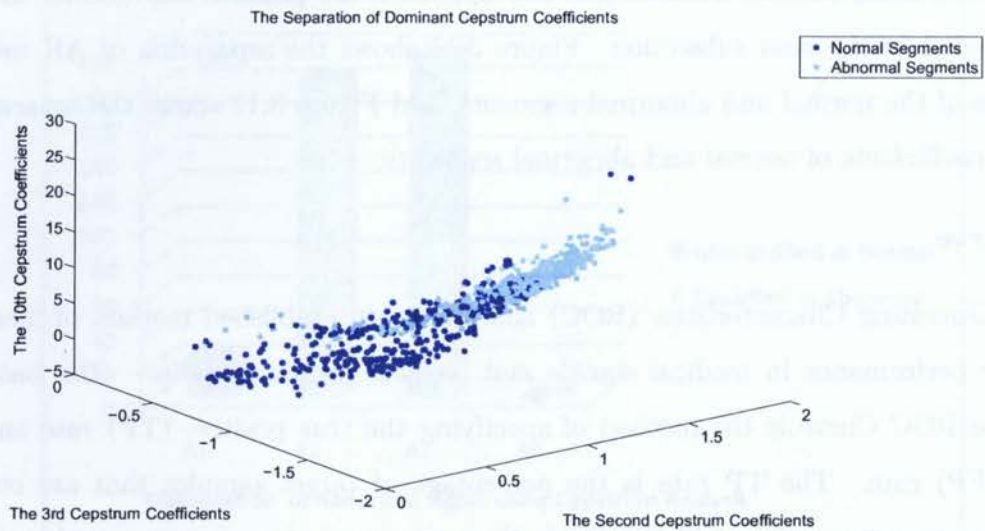


Figure 3.12: The Separation of Dominant Cepstral coefficients

Table 3.3: Confusion matrix containing the number of correct classified EMG in Sleep as either normal or abnormal using Discriminant classification and Cepstrum Analysis

		Predicted Group Membership		
		Normal	Abnormal	total
Count	Normal	473	74	547
	Abnormal	131	278	409
%	Normal	86.5	13.5	100
	Abnormal	32.0	68.0	100

Table 3.4: Confusion matrix containing the number of correct classified EMG in Sleep as either normal or abnormal using Leave-one-out classification and Cepstrum Analysis

		Predicted Group Membership		
		Normal	Abnormal	total
Count	Normal	472	75	547
	Abnormal	133	276	409
%	Normal	86.3	13.7	100
	Abnormal	32.5	67.5	100

sensitivity and $(1-FP)$ is referred to as specificity. In other words, ROC curve describes the inherent tradeoff between sensitivity and specificity of a diagnostic test by plotting the sensitivity vs. specificity points obtained from a threshold of the decision stage of the proposed algorithm. Measure of effectiveness of an algorithm is then given by the area under the ROC curve (AUC) [74].

In order to plot the ROC Curve for the proposed method, the two classes (normal and abnormal), and the predicted group membership of each of the signals were analyzed. From these two the TP and FP samples were calculated. Figure 3.13 shows the ROC curve of the AR modeling and Figure 3.14 shows the ROC curve of the Cepstrum Analysis. The AUC of the AR modeling was calculated to be 0.797 (79.7%) and the AUC of the Cepstrum analysis was calculated to be 0.772 (77.2%), which closely matches the result from the confusion matrices.

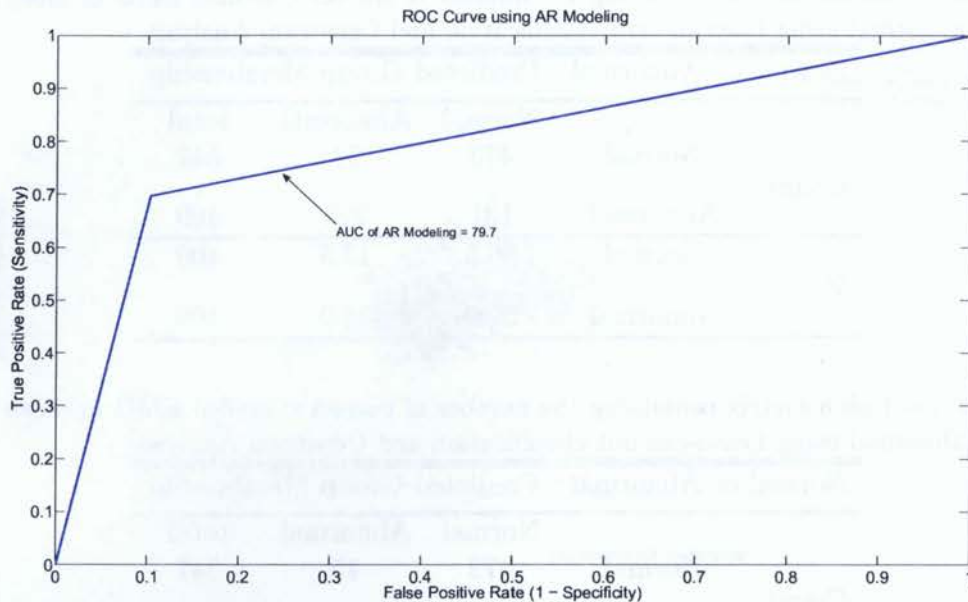


Figure 3.13: Receiver Operating Characteristic of the AR modeling

3.6 Discussion

As discussed before, a total of 8 normal and abnormal EMG in REM sleep were used for this experiment. These 8 signals were first segmented into stationary components using adaptive signal processing techniques. 956 samples were extracted from them in which 547 of them were normal segments and 409 of them were abnormal segments. On each of these stationary segments, modeling techniques were applied. To evaluate the performance of these techniques a classifier was used. These signals were correctly classified (normal as normal and abnormal as abnormal) at a rate of 79.8% for AR modeling while achieving a misclassification rate of 20.2%, and the correct classification result for Cepstrum analysis was at a rate of 77.2% and a misclassification rate of 22.8%.

As shown by the result, cepstrum does not significantly improve the accuracy, compared to AR modeling unlike in speech processing. Although cepstral coefficients show better separability in the feature space, and emphasize the spectral difference in the low-frequency

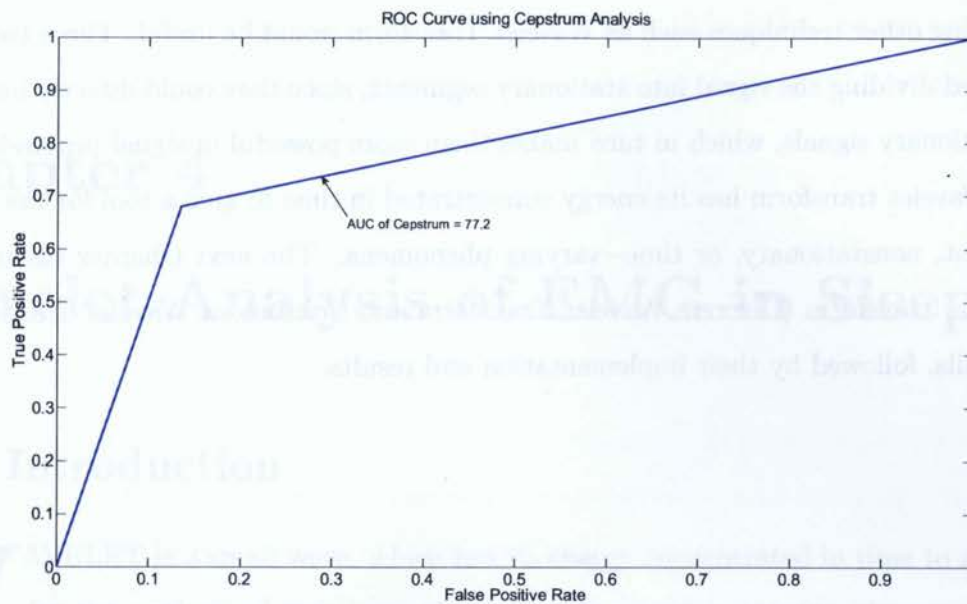


Figure 3.14: Receiver Operating Characteristic of the Cepstrum Analysis

band, however, the frequency band of the chin EMG is very lower than the speech signal. This might be the reason by which cepstrum does not increase the accuracy, by much.

Aside from distinct classification, the level of abnormality and normality is of importance for neurologists and sleep specialists, as this may represent a signal that can be used to follow disease progression or response to therapy. Since, this is the first work in the area of EMG signal during REM Sleep, the results are promising, and show great potential for applications such as therapy and prediction of fatal diseases. This is especially true since all techniques were done in a fully automated manner, without any intervention or assistance from the user. This means that such a system could, in fact, be used as a tool, which could either (1) approximate the level of normality and abnormality in different subjects and/or (2) an earlier diagnosis to different diseases such as Parkinson to prevent for further development of the disease.

Nevertheless, since both these techniques might lack the stationarity properties, and also

since the diagnostic yield of the cepstral coefficient and the AR spectral measures were similar, applying other techniques such as Wavelet Transform would be useful. These techniques do not need dividing the signal into stationary segments, since they could directly be applied to nonstationary signals, which in turn makes them more powerful in signal processing techniques. Wavelet transform has its energy concentrated in time to give a tool for the analysis of transient, nonstationary, or time-varying phenomena. The next Chapter discusses the two wavelet transform (Discrete Wavelet Transform and Continuous Wavelet Transform) in more details, followed by their implementation and results.

Chapter 4

Wavelet Analysis of EMG in Sleep

4.1 Introduction

WAVELET is a small wave, which has its energy concentrated in time to give a tool for the analysis of transient, nonstationary, or time-varying phenomena. They are functions that satisfy certain mathematical requirements, and are used in representing data or other functions. Wavelet not only has oscillating wave-like characteristics but also has the capacity to allow simultaneous time and frequency analysis. The Fourier transform, provides information about the frequency domain; however, time localized information is essentially lost in the process. The problem with this is the inability to associate features in the frequency domain with their location in time, since frequency spectrum will result in changes throughout the time domain. In contrast to the Fourier transform, the wavelet expansion maps the signal into a two dimensional array of coefficients. In other words, this two dimensional representation allows localizing the signal into time and frequency. The wavelet analysis procedure is to adopt a wavelet prototype function, called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. These operations are applied to the mother wavelet to calculate the wavelet coefficients, which represent the correlation between the wavelet and a localized section of the signal. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions),

data operations can be performed, using the corresponding wavelet coefficients. This gives the ability to look at different scales and resolutions, which means an approximation of the signal look stationary, while at the detailed level discontinuities become apparent [75].

4.2 Motivation Behind the Study

As already explained in the previous Chapters, the stationary property, in its usual definition, means the independence of statistical properties relative to an absolute time. Nevertheless, the most natural local quantities which intrude into the description of nonstationary random signals, are related to time [76]. However, when analyzing signals with nonstationary nature, it is often beneficial to be able to acquire a technique that could be used without segmenting the signal into stationary components, as opposed to the one explained in Chapter 3. Furthermore, it would be practical to find techniques that enable one to find a correlation between the time and frequency domains of a signal, in order to study both time and frequency aspects simultaneously. The operation of translation and scaling seems to be basic to many practical signals and signal generating processes, and their use is one of the reasons that wavelets are efficient expansion functions. Another advantage of WT is that, it uses long duration windows for capturing the low frequency components, and short duration windows for capturing high frequency components; based on the assumption of rapid changes in high frequency components and slow changes in low frequency ones. This means that at high frequencies the WT is sharper in time, while at low frequencies the WT is sharper in frequency.

Hence, the motivation behind using wavelet in this work, is the ability to analyze the signal without dividing it into stationary components; as well as the high efficiency of its use in many practical applications. To the best of Author's knowledge, there has been no reported literature on the application of WT on the EMG in REM Sleep. Therefore, the techniques used in this thesis is novel.

The following sections will cover the methods and techniques that will be used to classify the EMG signals in REM sleep, normal vs abnormal, using WT. First the concept of DWT

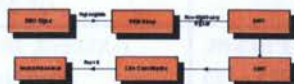


Figure 4.1: Block Diagram of the Wavelet Analysis

and CWT will be examined, followed by their applications on both sleep and nonsleep EMG. Then description of the classification scheme used will be included, and finally the experimental results. Figure 4.1 shows the algorithm and the techniques for this Chapter.

In the past, numerous research has been dedicated to EMG, using WT [56][77]–[80]. To provide a comparative analysis of the techniques used in this thesis, the research in the area of EMG analysis, using WT, is included in Table 4.1. Their main focus is extraction and classification of Motor Unit Action Potential (MUAP). The researchers have used variety of WT methods, including DWT, CWT and WPT (Wavelet Packet Transform).

4.3 Discrete Wavelet Transform

DWT transforms a discrete signal from time domain into time–scale domain. This results in a set of coefficients, which are organized to enable not only a spectrum analysis of the signal, but also to enable a spectral behavior study of the signal in time. In DWT a signal is represented by inner products that are temporal shifts and dilation of a prototype function ψ , often called mother wavelet [81]. Wavelets can be realized by iteration of filters, which

Table 4.1: Related works in Wavelet Transform

Source	Method	Application	Result
Gazzoni et al. [77]	CWT	Detection and Classification of MUAP	Automatic Detection of MUAP
Hu et al. [78]	Relative Wavelet Packet Energy (RWPE)	Extract features from surface EMG	Uses less features/ accuracy is higher
Leao et al. [56]	CWT	Evaluation of stretch reflex response from EMG	Suitable for extraction/ analysis of stretch reflex
Diab et al. [79]	DWT	Uterine EMG	High accuracy on stimulated signal low accuracy on real signal
Arveti et al. [80]	CWT	Effectiveness of active hand using EMG	High Accuracy 90%

are one of the most widely used signal processing functions. In other words, the lower resolution coefficients can be calculated from the higher resolution coefficients by a tree structured algorithm called a filter bank. In DWT, the filtering is being used to measure the amount of details in the signal, and the scale is determined by upsampling and downsampling operations. This allows an efficient calculation of the expansion coefficients. In other words, this could be achieved by decomposing a signal, breaking it into two components, where each carry information about source signal.

Filters from the filter bank, used for decomposition, come in pairs: low-pass and high-pass. The filtering proceeds by down-sampling (the obtained filtering result is “re-sampled”, so that every second coefficient is kept). Low-pass filtered signal contains information about slow changing component of the signal. They looking similar to the original signal only the number of samples is two times shorter. Conversely, the high-pass filtered signal contains information about fast changing component of the signal [81]-[86]. There are many advantages in using DWT, which are itemized below:

- Efficient to compute.
- Easy to implement

Next subsection explains the wavelet decomposition and filter banks, in more detail.

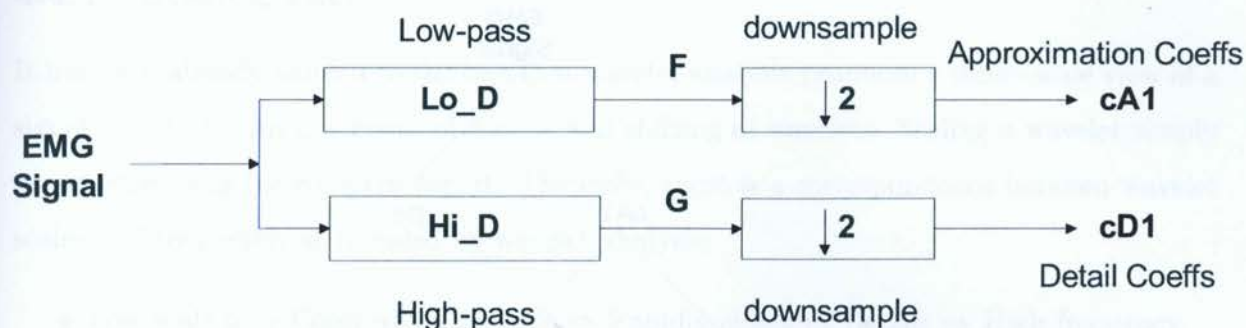


Figure 4.2: 1-Level Wavelet Decomposition

4.3.1 Wavelet Decomposition

As explained in the previous sections, DWT is being done by using a set of filters. The output of different filter stages are the wavelet and scaling function transform coefficients. The filtering is done by splitting the signal spectrum into two equal parts, a low-pass and a high-pass part. The high-pass part usually has the most details of the signal, often known as noise. Yet, the low-pass part contains less details which could be split again and again until the number of the bands are satisfactory. Therefore, the decomposition process can be iterated, with successive approximations being decomposed in turn; so that one signal is broken down into many lower resolution components, called the wavelet decomposition tree. This is shown in Figures 4.2 and 4.3.

Continuous analysis is often easier to be interpreted, since its redundancy tends to reinforce the traits, and makes all information more visible. This is specially true of very subtle information. Thus, the analysis gains in “readability” and in ease of interpretation what it loses in terms of saving space.

4.4 Continuous Wavelet Transform

As explained before, the idea behind the time–frequency joint representations is to segment the signal of interest into several parts, and then analyze the parts separately. It is clear that

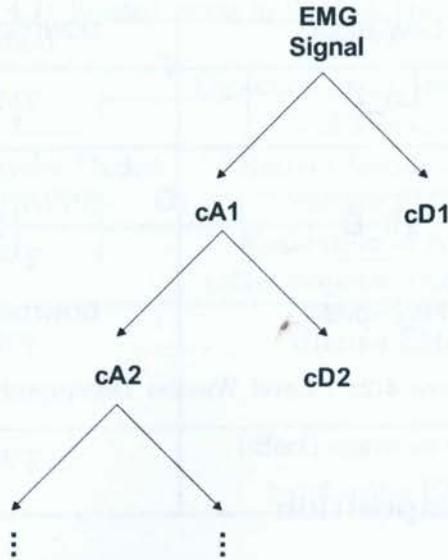


Figure 4.3: Wavelet Decomposition Tree

analyzing a signal in this way will provide more information about the when and the where of different frequency components. CWT is shifted along the signal, and for every position the spectrum is calculated. Then, this process is repeated many times with a slightly shorter (or longer) window for every new cycle. In the end, the result will be a collection of time–scale representations of the signal, all with different resolutions. The CWT is in the mathematical form below:

$$W(a, b) = \frac{1}{\sqrt{a}} \int x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (4.1)$$

In this Equation a is scale factor, b is the translation factor, and $\frac{1}{\sqrt{a}}$ is a factor for energy normalization across different scales. In other words, the continuous transform displays the temporal structure of the various different frequency components of the signal [87]. The scaling factor and the translation factor are two important phenomena in CWT applications, and they could vary according to the application in which they are being used.

4.4.1 Scaling Factor

It has been already alluded to the fact that wavelet analysis produces a time–scale view of a signal, which in turn is a result of scaling and shifting of wavelets. Scaling a wavelet simply means stretching (or compressing) it. Therefore, there is a correspondence between wavelet scales and frequency, as revealed by wavelet analysis:

- Low scale $a \Rightarrow$ Compressed wavelet \Rightarrow Rapidly changing details \Rightarrow High frequency.
- High scale $a \Rightarrow$ Stretched wavelet \Rightarrow Slowly changing / coarse features \Rightarrow Low frequency.

As a result, when the scale factor, a , is enlarged its effect on frequency is compression, as the analysis window is contracted by the amount of $1/a$. That means, the more stretched the wavelet, the longer the portion of the signal with which it is being compared, and thus the coarser the signal features, being measured by the wavelet coefficients. And the more compressed the wavelet, the smaller the portion of the signal being compared, and thus more rapidly changing detailed features, being produced and measured. Figure 4.4 is an example of different scaling in CWT.

Fitting of the wavelet, CWT, to the data results in calculation of a resemblance index between the signal $x(t)$ and the wavelet located at the position b and scale a . The higher this value, the more correlated the signal is with the mother wavelet. Unlike DWT, CWT can operate at every scale, from the original signal up to some maximum scale, by trading off the need for detailed analysis with available computational complexity. The scales that were used in this work were 15 – 100 and 150 – 300 in order to measure both the low frequency components (i.e. higher scales) and the high frequency components (i.e. lower scales). The result from these scale showed that most of the energy is in the higher scales and therefore, using higher scales would be a better choice. However, due to the length of the signal and the computational time that it takes, both higher and lower scales could not be analyzed simultaneously, and only one of them was analyzed at a time.

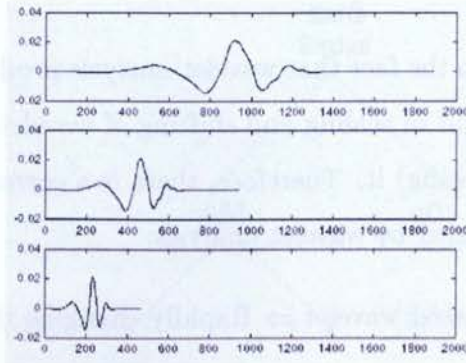


Figure 4.4: Scaling of Wavelet

4.4.2 Mother Wavelet

Another very important factor in both CWT and DWT is the selection of the mother wavelet. A suitable mother wavelet selection is the kernel of feature extraction [79]. It must be well adapted to the events, to be classified. Multiresolutional Analysis (MRA) assures that once a scaling function is specified, the associated function $\psi(t)$ can be generated. This happens prior to the analysis, rather than being derived from the data set. There are many different wavelets shapes available. They ranged in shape, from square to triangular waveforms, to Gaussian and Mexican Hat shapes. Within a given family of wavelets (i.e. Daubechies) the waveform can be ranged from simple to primarily biphasic, to much smother and multiphasic form. Each has its own properties, such as: orthogonality, regularity, symmetry, and support width. These are discriminative properties that are important to keep in mind when choosing the mother wavelet. The basic requirement for mother wavelet for DWT is orthogonality since it avoids redundancy in decomposition of the signal, and ensures the unique reconstruction for finite energy signals. Orthogonality means:

- Scaling and translated functions are orthogonal.
- Scaling and wavelet functions are orthogonal.
- Scaling and wavelet functions are orthogonal in function space.

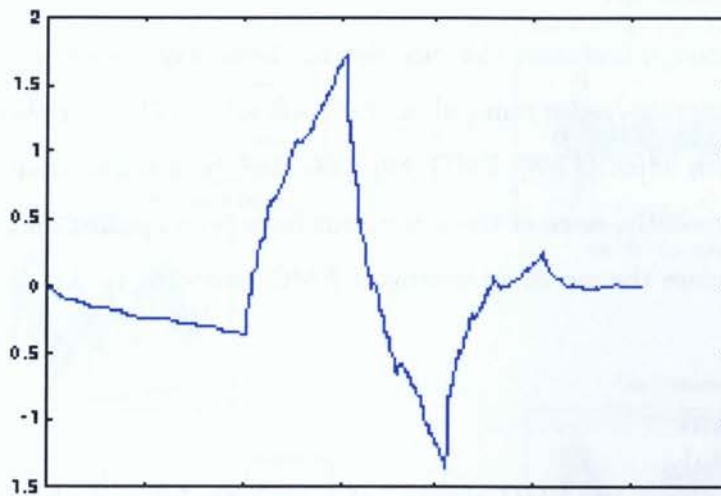


Figure 4.5: Daubechies Wavelet

Daubechies Wavelets (Db) along with Symlet (sym) and Coiflets are among the family of wavelets that have these properties [81][87]. Within each family of wavelets, these are wavelet subclasses, which can be distinguished by the number of coefficients and level of iterations. Another very important feature of each family of wavelets is number of vanishing moments which is closely related to the number of coefficients. Because of its properties and ease of use, the mother wavelet that was used in this work was the Daubechies wavelet. This was applied to both DWT and CWT.

Daubechies Wavelet

As explained before the main characteristic of the Daubechies family of wavelet is the orthogonality. Another very useful characteristic of this family is that they have a maximum number of zero moments of the wavelet, which would result in a high degree of smoothness for the scaling and wavelet functions. Figure 4.5 shows the Daubechies (Db4), used in this work. Next section will show the results achieved from applying wavelet on EMG signal during REM sleep.

4.5 Wavelet Analysis of Sleep EMG

It has been shown through literature that wavelet has been widely dedicated to many applications such as compression, noise removal, and classification. These applications vary from biomedical signals such as ECG [88], EMG [89] EEG [90], biomedical images [91] and many more [92]. Yet as of recently, none of these methods have been applied on EMG signal, during REM sleep; therefore the use of an overnight EMG recording to detect RBD represents a novel approach.

4.5.1 Algorithm

As explained in Chapter 1 the EMG consists of tonic (steady) and phasic (intermittently elevated) bursts in the REM sleep. The phasic REM activity is defined as any burst of EMG activity lasting for $0.1 - 2s$ and has an amplitude of at least $50\mu V$, while tonic REM constitutes the remainder [9]. For patients with RBD tonic tone is higher than in normal subjects. The increase in tonic EMG activity might reflect disease progression [19].

The EMG signal in REM Sleep was originally processed using DWT. As stated before, the sampling rate to record the EMG signal is 256 samples/s. Therefore 1-level decomposition would consist of an approximation of $0 \sim 64Hz$ and a detail signal of $64 \sim 128Hz$. Since most of EMG spectral characteristics are in the range of $0 \sim 64Hz$, approximation coefficients can be a better representation of the decomposed signal than the detailed coefficients.

Although the DWT provides good space–frequency localization and is scale invariant; it is a well known fact that the DWT is shift variant [54]. This means that for different translations of an input signal, a different set of DWT coefficients would be generated [53][82][83]. This shift variant property of the DWT is widely known, and several solutions have been proposed. Thus far, in this work, to overcome the problem, the DWT was not further applied to the signal, and CWT was used; this is because the redundancy of some applications of CWT can improve the shift invariance of the transform as well as analysis gains in “readability”, and in ease of interpretation what it loses in terms of saving space [93]. Therefore, the one level decomposition DWT was used to capture the tonic and phasic twitches for each signal,

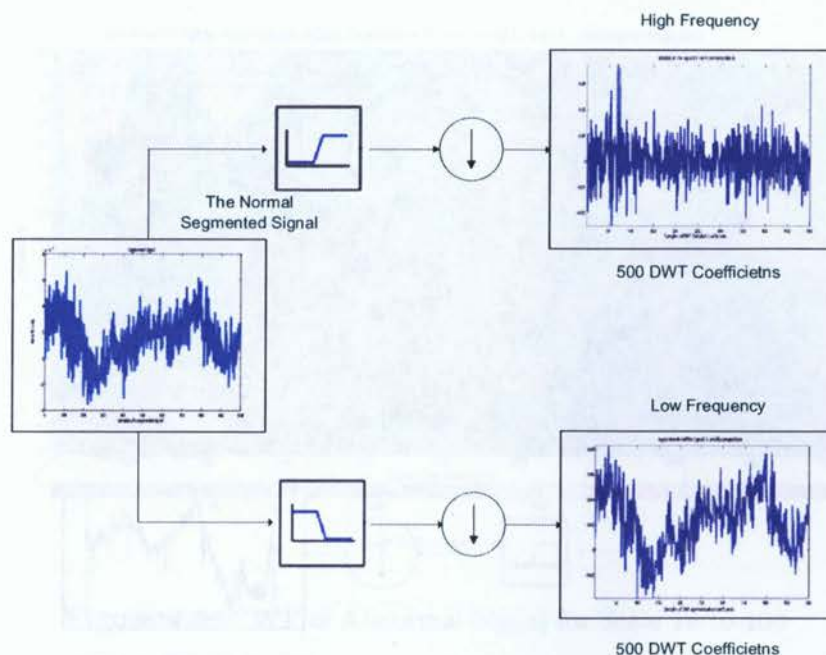


Figure 4.6: 1 Level Decomposition EMG of Normal Segment

and then the CWT was applied to the approximation coefficients of the DWT to extract useful features from the CWT coefficients. Figures 4.6 and 4.7 show 1-level decomposition of EMG for normal and abnormal segments respectively.

As stated before, the WT was mainly used to overcome the adaptive segmentation of the nonstationary signals. However, this was not possible due to the long length of the signal as well as the redundancy of the CWT coefficients. To solve this issue, the EMG signal in REM sleep was segmented into fixed length components, and the wavelet analysis was applied on each of these segments separately. The extent of this fixed segment was chosen to be 1000 samples, which corresponds to almost 4s of data. This was on an average significantly longer than the segments, used in modeling work, presented in Chapter 3.

The plot of the CWT shows the signal in time and scale. This representation clearly indicates at what time and which scale the energy is distributed throughout the signal. This energy representation is very important in classifying the signal into normal and abnormal,

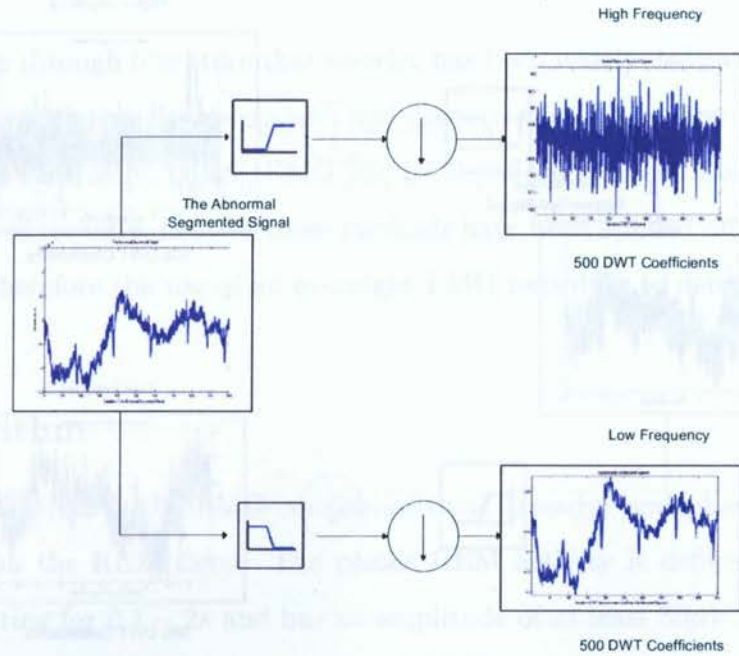


Figure 4.7: 1 Level Decomposition EMG of Abnormal Segment

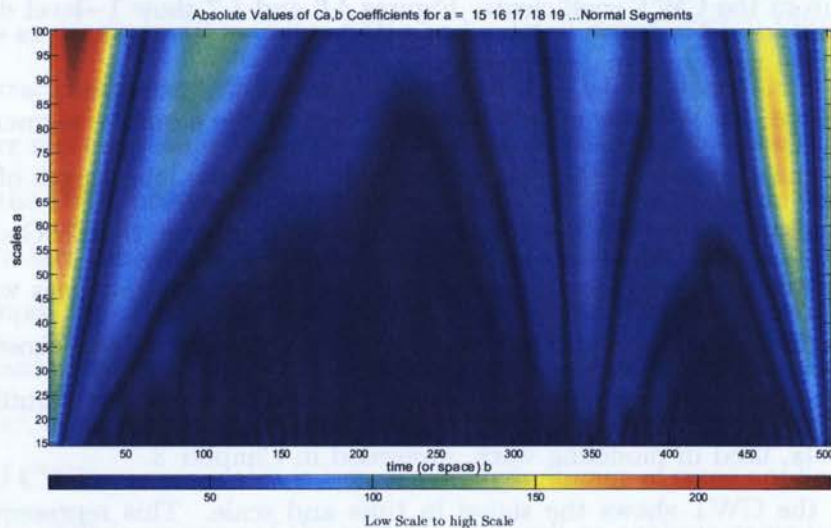


Figure 4.8: CWT of Normal Signal for Scale 15 to 100

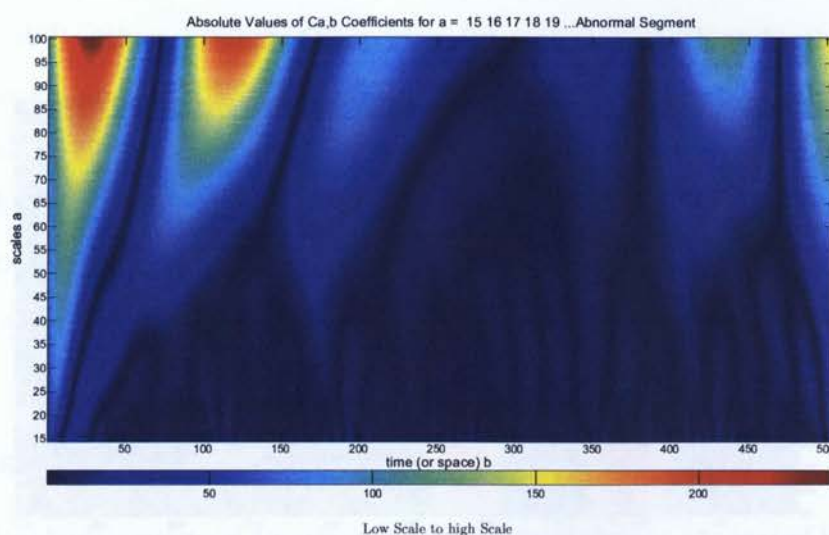


Figure 4.9: CWT of Abnormal Signal for Scale 15 to 100

since the abnormal signal is expected to have higher energy distribution throughout each segment. Therefore, the coefficients extracted from CWT, which shows the energy distribution of the signal, could be used as features input to the classifier.

At first the scale used to classify the signal was very low, this means that since the scale is low, the high frequency signals are being analyzed. However, since most of the information was in low-frequency components the higher scales were analyzed in more detail. As a result, the range of the scale used was 15 to 100. Figures 4.8 and 4.9 show the CWT representation of normal and abnormal subjects respectively. These two figures are chosen from two different segments of both normal and abnormal signals. In addition, the distribution of energy in each of these signals, for different scales, are very much different from each other. Nevertheless from both these Figures, it is shown that the maximum energy is within the higher scales. Consequently the higher scales (150 ~ 300) were analyzed. Figures 4.10 and 4.11 show the CWT representation of scales 150 to 300.

The algorithm used in this work tries to predict the scale that maximizes the CWT coefficients. The higher the coefficients, the more correlated the signal to the mother wavelet

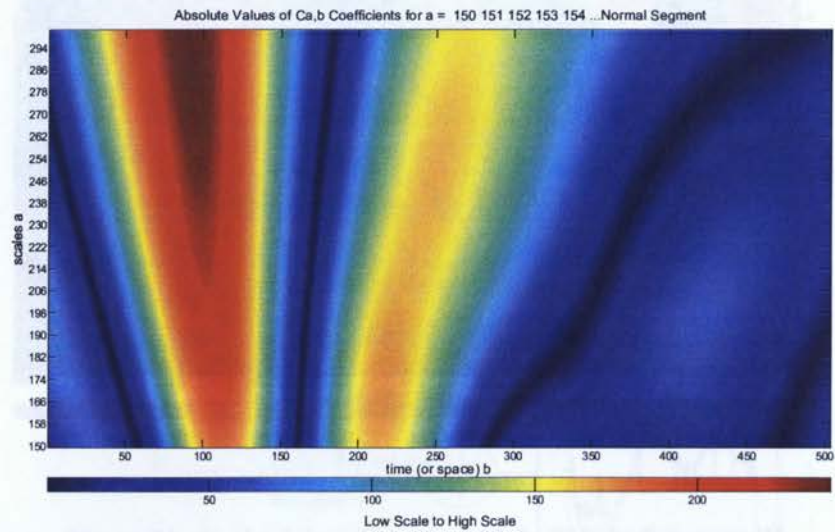


Figure 4.10: CWT of Normal Signal for Scale 150 to 300

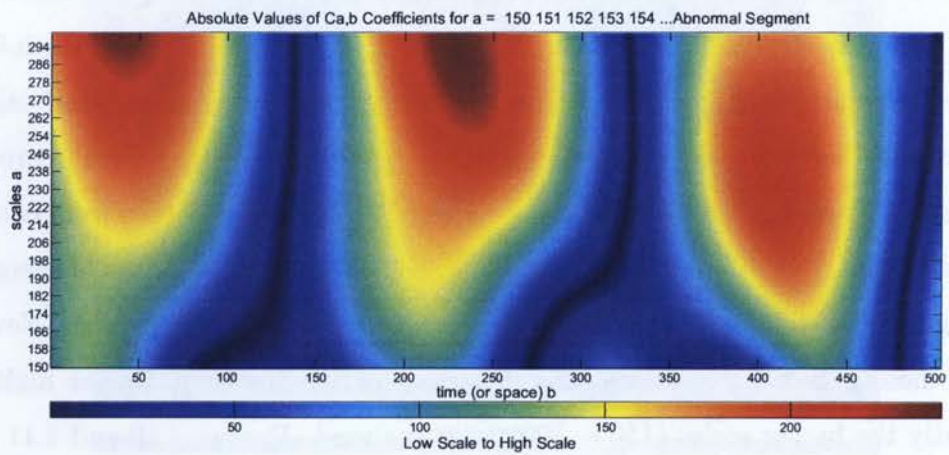


Figure 4.11: CWT of Abnormal Signal for Scale 150 to 300

and thus higher the energy. The reason for this procedure was to avoid comparing different scales of the CWT with each other. Comparing different scales with each other would result in wrong outcome and/or variable accuracy. Since the energy was mostly distributed in higher scales the scale that gave the highest energy coefficients was 100 and 300 respectively for all the 8 different signals. The result of the CWT coefficients for these scales were fed to a classifier, and the classification accuracy was calculated. The next subsection shows the classification accuracy of classifying the EMG segments into normal and abnormal, using WT.

4.5.2 Results

The original database consists of 4 normal and 4 abnormal signals. As stated before, the WT, specially, CWT could not be directly applied to the signal, because of the long computational time and the memory that it uses to compute the coefficients. Therefore, the signals were divided into smaller components in such a way that the computational time is shorter than of other applications such as adaptive signal processing and AR modeling, explained in the previous Chapter. Since each of the signals had a different length, the number of segments varied from signal to signal. As a result, when segmenting the signal into fixed (1000 samples) components, the number of fixed segments were calculated to be 2350 signals. 1641 of these segments were marked as abnormal and 709 of them were marked as normal. Most of the biomedical researchers suffer from a low number of dataset to validate their results. Nevertheless it is difficult to predict what an optimum dataset size maybe. Therefore, to test the performance of the proposed feature extraction and classification scheme, all of the fixed segments were used. The objective of the proposed system for the EMG signals were as follows:

1. To validate the techniques used in this work;
2. To compare the results of the three applications used in this work (AR modeling, Cepstrum Analysis and Wavelet Transform);

3. To find the level of the normality and abnormality in different signals;
4. To find the disease progression in different subjects and/or their response time to therapy;
5. To prevent subjects from Parkinson Disease by studying their neuromuscular behavior in REM Sleep;

4.5.3 Classification Result

Classification using LDA method was performed using SPSSTM [67]. The following subsection will present the classification results as a confusion matrix, which shows how many signals were correctly classified and how many signals were incorrectly classified. The LOO method was also used to prove that these results will work on a different dataset. The results of LDA and LOO for scale 15 ~ 100 are shown in Tables 4.2 and 4.3, and for scales 150 ~ 300 are shown in Tables 4.4 and 4.5.

Table 4.2: Confusion matrix: containing the number of correct classified EMG in Sleep, as either normal or abnormal, using Discriminant classification and WT for Scale 15 to 100.

		Predicted Group Membership		
		Normal	Abnormal	total
Count	Normal	691	18	709
	Abnormal	143	1498	1641
%	Normal	97.5	2.5	100
	Abnormal	8.7	91.3	100

From the Table 4.2, LDA classification results of EMG in REM Sleep are derived, indicating that the correct classification rates were 97.5% (normal as normal) and 91.3% (abnormal as abnormal). Misclassification rates were computed to be 2.5% (normal as abnormal) and 8.7% (abnormal as normal). An overall classification rate of 93.1% was achieved for the EMG in REM sleep using WT, and an average misclassification rate was 6.9%. In Table 4.3, LOO classification results of EMG in REM Sleep are shown, indicating that the correct

classification rates were 97.5% (normal as normal) and 91.3% (abnormal as abnormal). Misclassification rates were computed to be 2.5% (normal as abnormal) and 8.7% (abnormal as normal). An overall classification rate of 93.1% was achieved for the EMG in REM sleep using WT and also an average misclassification rate of 6.9% was achieved. The latter shows that the system will work for a different group of dataset.

Table 4.3: Confusion matrix: containing the number of correct classified EMG in Sleep, as either normal or abnormal, using Leave-one-out method classification and WT for Scale 15 to 100.

		Predicted Group Membership		
		Normal	Abnormal	total
Count	Normal	691	18	709
	Abnormal	144	1497	1641
%	Normal	97.5	2.5	100
	Abnormal	8.8	91.2	100

The results for the higher scales show a higher accuracy, which are determined to be as follows: From Table 4.4, LDA classification results of EMG in REM Sleep are driven, indicating that the correct classification rates were 97.9% (normal as normal) and 94.6% (abnormal as abnormal). Misclassification rates were computed to be 2.1% (normal as abnormal) and 5.4% (abnormal as normal). An overall classification rate was 95.6% for the EMG in REM sleep, using WT, and an average misclassification rate was 4.4%. From Table 4.5, LOO classification results of EMG in REM Sleep are driven, indicating that the correct classification rates were 97.9% (normal as normal) and 94.5% (abnormal as abnormal). Misclassification rates were computed to be 2.1% (normal as abnormal) and 5.5% (abnormal as normal). Therefore an overall classification rate of 95.5% was achieved for the EMG in REM sleep, using WT, and an average misclassification rate was 4.5%.

The results achieved using WT, compared to AR modeling and Cepstrum analysis, is more promising, and shows that analyzing both the time and frequency (scales) of the signal

Table 4.4: Confusion matrix: containing the number of correct classified EMG in Sleep, as either normal or abnormal, using LDA classification and wavelet WT for Scale 150 to 300.

Normal or Abnormal		Predicted Group Membership		
		Normal	Abnormal	total
Count	Normal	694	15	709
	Abnormal	89	1552	1641
%	Normal	97.9	2.1	100
	Abnormal	5.4	94.6	100

Table 4.5: Confusion matrix: containing the number of correct classified EMG in Sleep, as either normal or abnormal, using Leave-one-out method classification and WT for Scale 150 to 300.

Normal or Abnormal		Predicted Group Membership		
		Normal	Abnormal	total
Count	Normal	694	15	709
	Abnormal	90	1551	1641
%	Normal	97.9	2.1	100
	Abnormal	5.5	94.5	100

for longer duration is more accurate in detecting RBD in different subjects. Figures 4.12 to 4.16 show the distribution of normality and abnormality in different subjects. In these Figures N and A refer to Normal and Abnormal subjects respectively. This is a valuable asset for neurologists since this may represent a signal that can be used to follow the disease progression or the response to therapy.

In order to prove the classification result the three dominant Wavelet Coefficients of both normal and abnormal are plotted, followed by the ROC curve in the next subsection. Figure 4.16 shows the separation of WT coefficients of the normal and abnormal segments.

ROC Curve

In order to plot the ROC Curve for the proposed method, the two classes (normal and abnormal) and the predicted group membership of each of the signals were analyzed. From these two, the TP and FP samples were calculated. Figures 4.17 and 4.18 show the ROC

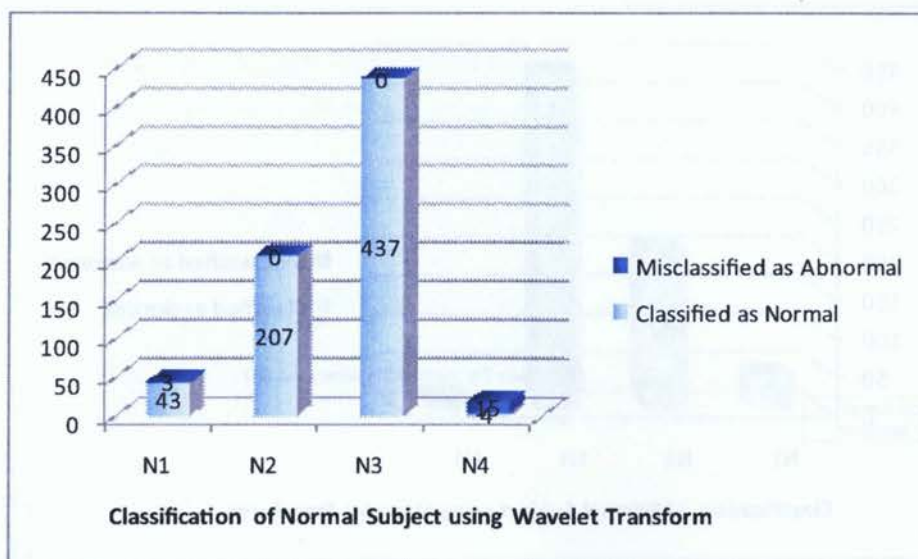


Figure 4.12: Distribution of Normality in Normal subjects

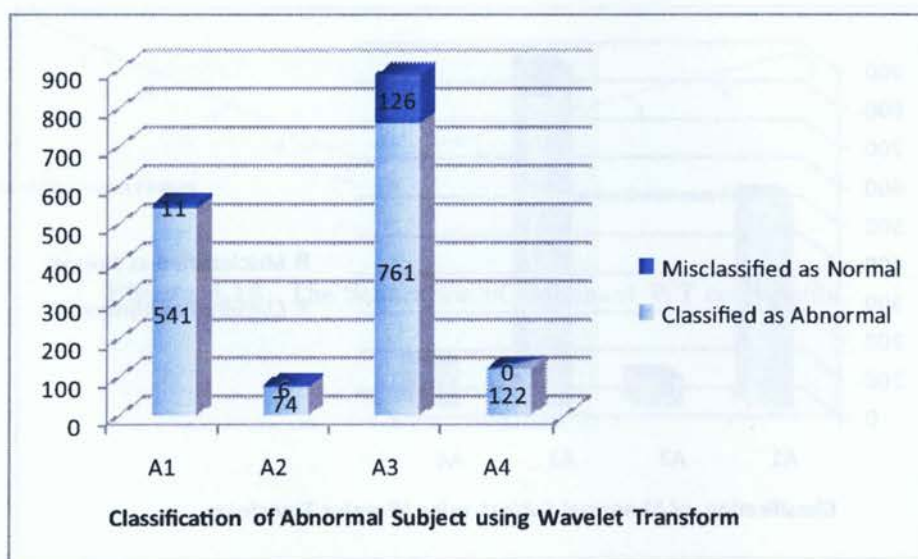


Figure 4.13: Distribution of Abnormality in Abnormal subjects.

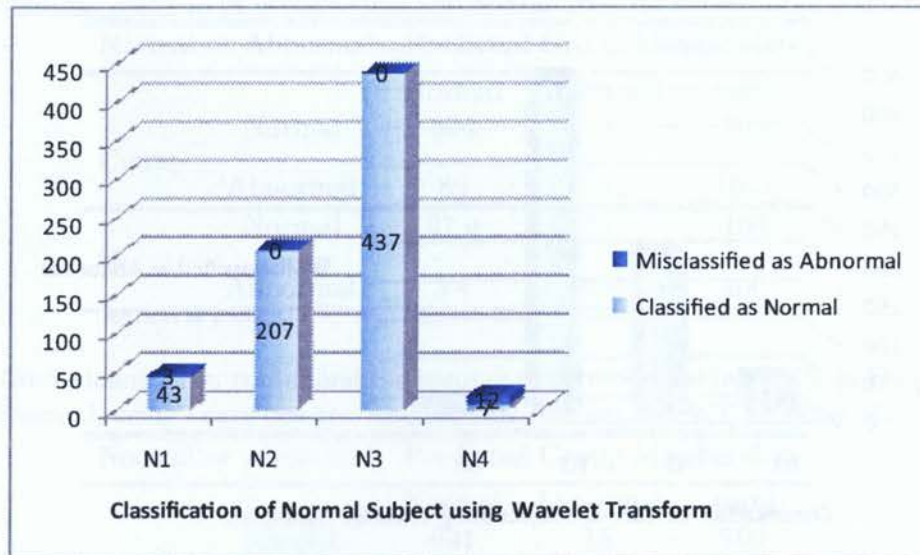


Figure 4.14: Distribution of Normality in Normal subjects for Scale 150-300.

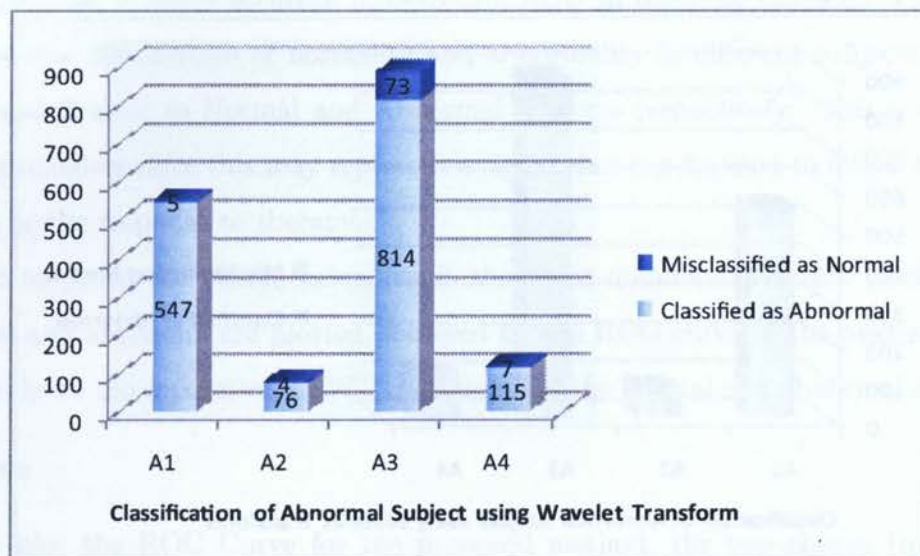


Figure 4.15: Distribution of Abnormality in Abnormal subjects for Scale 150-300.

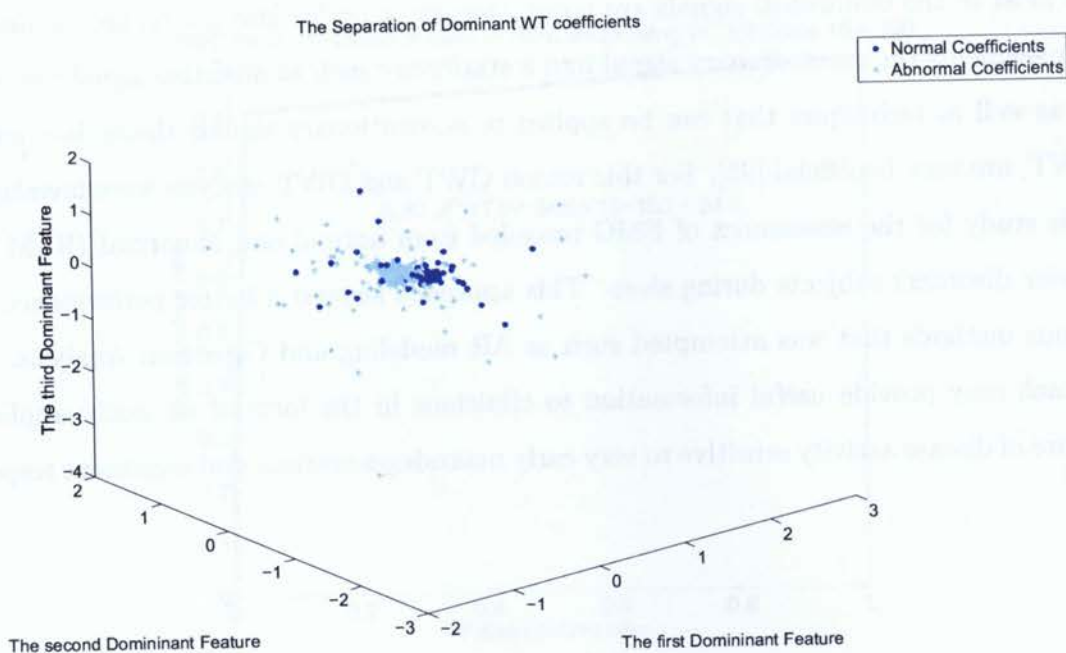


Figure 4.16: The Separation of Dominant WT coefficients

curve of the wavelet analysis for scale 15 – 100 and 150 – 300 respectively. The AUC of these curves are calculated to be 0.944 (94.4%) and 0.962 (96.2%). These values, compared to those obtained in Chapter 3, show that WT is a better technique for detecting the RBD. Figure 4.19 demonstrates this fact in more detail.

4.6 Discussion

Since most of the signal processing techniques have to be done on stationary signals and also since most of the biomedical signals are nonstationary or cyclo-stationary, techniques that either segments the nonstationary signal into a stationary such as adaptive signal processing [94], as well as techniques that can be applied to nonstationary signals themselves such as the WT, are very beneficial [95]. For this reason CWT and DWT analysis were investigated in this study for the assessment of EMG recorded from normal and abnormal (REM sleep behavior disorder) subjects during sleep. This approach showed a better performance than previous methods that was attempted such as AR modeling and Cepstrum Analysis. This approach may provide useful information to clinicians in the form of an easily applicable measure of disease activity sensitive to very early neurodegeneration, and treatment response.

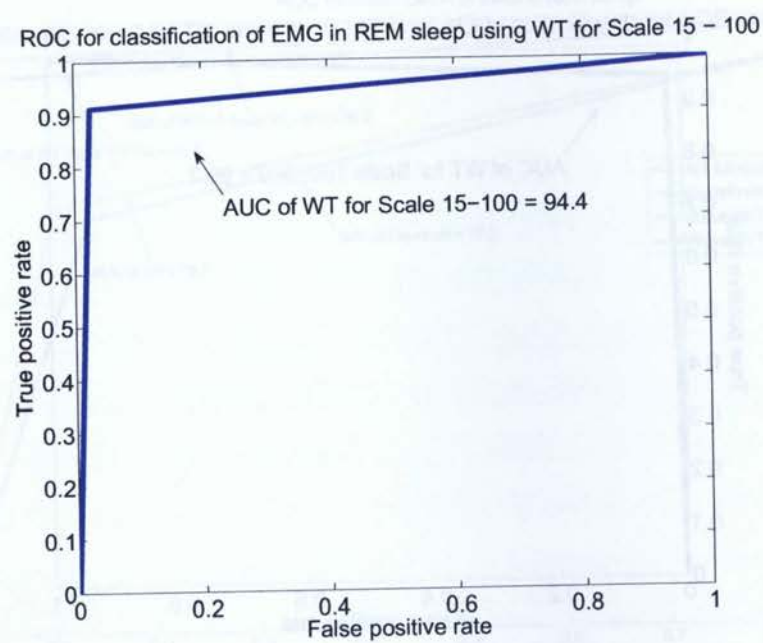


Figure 4.17: ROC for classification of EMG in REM sleep using WT for Scale 15 -100

4.6 Discussion

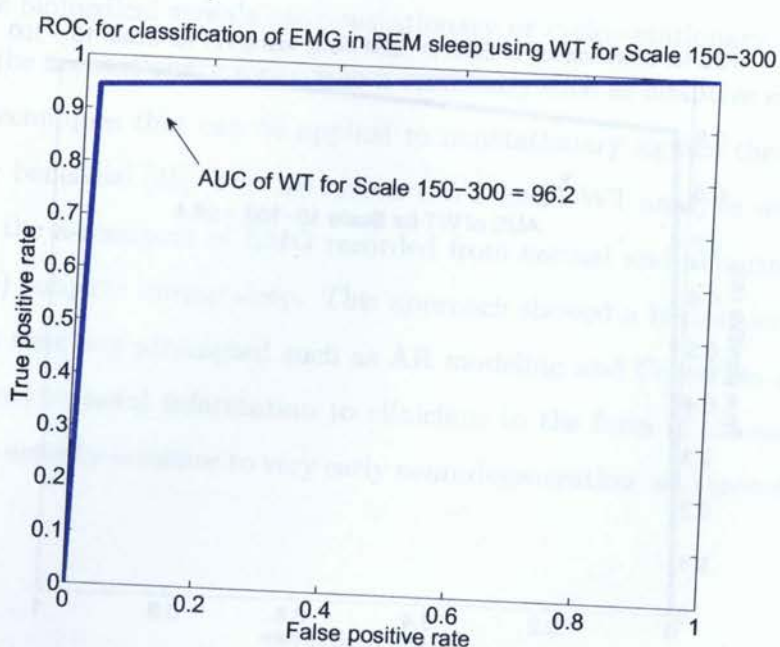


Figure 4.18: ROC for classification of EMG in REM sleep using WT for Scale 150 -300

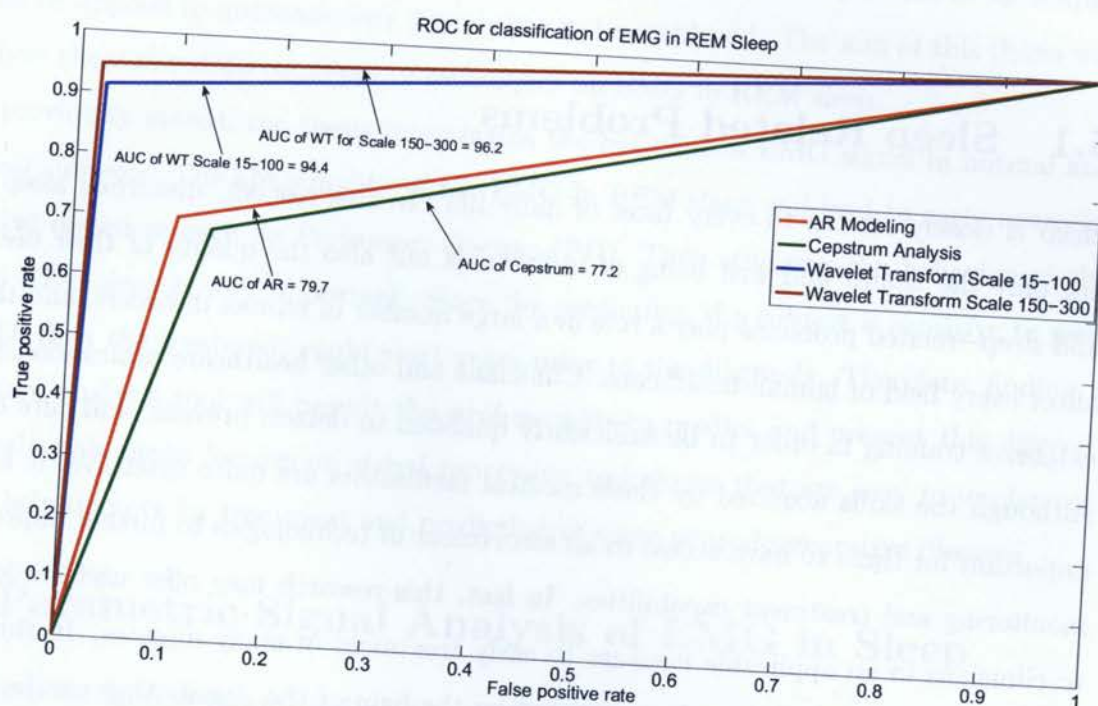


Figure 4.19: ROC for classification of EMG in REM Sleep

Chapter 5

Conclusions

5.1 Sleep Related Problems

Sleep is closely related to every facet of daily life. In this respect, disturbed sleep affects not only the health and well being of individuals but also the quality of their life. Sleep and sleep-related problems play a role in a large number of human disorders, and therefore affect every field of human healthcare. Clinicians and other healthcare professionals receive extensive training in order to be sufficiently qualified to detect, prevent, and cure diseases. Although the skills acquired by these medical facilitators are quite extensive, it is just as important for them to have access to an assortment of technologies to further improve their monitoring and treatment capabilities. In fact, this research may offer useful information to clinicians in an applicable measure to early treatment of many diseases. In this respect many incurable diseases may be prevented by the help of the signals that can be detected before it becomes very late.

The introduction of a number of new techniques during the past few decades, including polysomnographic (sleep study) surface measurements of central nervous system (CNS) activity, eye movements, and muscle activity, has allowed sleep to be described in electrophysiological terms. Therefore, studying sleep behavior and particularly the sleep signals are very important, since it affects all humans. In consequence this thesis was mainly dedicated to the analysis of sleep and a common type of disorder called Rapid Eye Movement Sleep Behavior Disorder (RBD).

It is a well known fact, in the real world that, there exists abundant kinds of signals. All those signals carry lots of information that are of human's interests. People develop diverse techniques to analyze, interpret, manipulate, and process those signals. Biomedical signals are a type that have strong relationship with human body or human organs. It has been shown that the type of most of these signals are nondeterministic and further nonstationary or quasistationary. Yet, in order to be able to apply signal processing techniques to these types of signals, either the signal has to be divided into stationary components or techniques that can be applied to nonstationary signals has to be employed. The aim of this thesis was to analyze these two different types of techniques on EMG in REM sleep.

As previously stated, the thesis investigates the behavior of EMG signal in normal and abnormal subjects. The abnormality of the EMG in REM sleep will lead to early neurodegenerative disorders such as Parkinson Disease (PD). Thus studying the behavior of the EMG in REM sleep is very important. Since, by predicting, if a subject is sensitive to generate PD, then the treatment could start years prior to the diagnosis. Therefore, finding a good signal analysis tool will permit the professionals to predict and prevent this disease. As a result, this thesis focuses on signal processing techniques that are easy to implement and can help doctors for treatment and prediction of some neurodegenerative diseases.

5.2 Parametric Signal Analysis of EMG in Sleep

As discussed in Section 3.5 a total of 8 normal and abnormal EMG in REM sleep were used for experimentation. These 8 signals were first segmented into stationary components, using adaptive signal processing techniques. Total of 956 samples were extracted from them, 547 were normal segments and 409 were abnormal segments. On each of these stationary segments modeling techniques were applied. To evaluate the performance of these techniques a classifier was used. These signals were correctly classified (normal as normal and abnormal as abnormal) at a rate of 80.6% for AR modeling while achieving a misclassification rate of 19.4%, and the correct classification result for Cepstrum analysis was at a rate of 78.4% and a misclassification rate of 21.6%.

As shown in Section 3.5.2, Cepstrum did not significantly improve the accuracy, compared to AR modeling unlike in speech processing. Although Cepstral coefficients show a better separability in the feature space and emphasize the spectral differences in the low-frequency band, the frequency band of the chin EMG is much lower than the speech signal. Therefore, cepstral analysis did not increase the accuracy by much.

Nevertheless, since both these techniques might lack the stationarity properties, and also because the diagnostic yield of the cepstral coefficients and the AR spectral measures were similar, applying other techniques such as Wavelet Transform was useful. These techniques do not need dividing the signal into stationary segment since they could directly be applied to nonstationary signals which in turn makes them more powerful in signal processing techniques.

5.3 Wavelet Analysis of EMG in Sleep

Since most of the signal processing techniques have to be done on stationary signals and also since most of the biomedical signals are nonstationary or cyclo-stationary, techniques that either segments the nonstationary signal into stationary forms such as adaptive signal processing [94], as well as techniques that can be applied to nonstationary signals themselves such as the Wavelet Transform are very beneficial [95]. For this reason Continuous and Discrete Wavelet Transform analysis were investigated in this study for the assessment of EMG recorded from normal and abnormal (REM sleep behavior disorder) subjects during sleep. This approach showed a better performance than previous methods, that was attempted, such as AR modeling and cepstral analysis. This approach may provide useful information to clinicians in the form of readily applicable tools to diseases that are preventable.

As discussed in Section 4.5.2 the 4 normal and 4 abnormal signals had to be segmented into smaller components due to the computational time and memory storage issues that the author had. As a result, the number of fixed segments were calculated to be 2350 signals. 1641 of these segments were marked as abnormal and 709 of them were marked as normal. DWT was applied to each of these segments and further CWT was applied to the

DWT's approximation coefficients. Since the distribution of energy coefficients were seen to be higher as shown in Figures 4.10 and 4.11. Thus higher scales were also applied on the segments. However, again because of the computational time and memory issues both high and low scales could not be analyzed simultaneously. To evaluate the performance of this technique Linear Discriminant Analysis (LDA) was applied. The signals were correctly classified (normal as normal and abnormal as abnormal) at a rate of 94.4% while achieving a misclassification rate of 5.6% for scale 15–100, and for scale 150–300 the correct classification rate was 96.25% and a misclassification rate of 3.75%.

Aside from distinct classification, the level of abnormality and normality was very important, as this may represent a signal that can be used to follow a disease progression or a response to therapy. The results were promising and show great potential for applications such as therapy and prediction of serious neurological diseases. This was specially true since all techniques were done in a fully automated manner, without any intervention or assistance from a user. Hence, such a system could, in fact, be used as a tool, which could either (1) approximate the level of normality and abnormality in different subjects and/or (2) an earlier therapy to different diseases such as Parkinson to prevent for further development of the disease.

5.4 Future Work

Some of the following points could be addressed in the future research endeavors,

1. Other types of adaptive signal processing could be used such as Recursive Least Square Lattice (RLSL) algorithm, which is based on calculating the conversion factor values of each input sample for adaptive segmentation, and further provide segment boundaries.
2. Extract dominant poles from the AR coefficients and characterize them in terms of frequency and energy distribution to the signal.
3. Reduce the computational time of the CWT which could result in applying the CWT to the whole signal.

4. All these algorithms could be applied to different stages of sleep to find other sleep disorders.
5. The effect of number of features extracted as well as the database size on classification performance could be examined.

List of Acronyms

ACF - Autocorrelation Function
ACh - Acetylcholine
AR - Autoregressive
ARMA - Autoregressive Moving-Average
AUC - Area Under the Curve
CAD - Computer Aided Diagnosis
CNS - Central Nervous System
CWT - Continuous Wavelet Transform
Db - Daubechies
DWT - Discrete Wavelet Transform
ECG - Electrocardiogram
EEG - Electroencephalogram
EMG - Electromyogram
EOG - Electrooclogram
FFT - Fast Fourier transform
FP - False Positive
LDA - Linear Discriminant Analysis LMS - Least-mean-square
LOO - Leave-One-Out NSF - National Sleep Foundation
NREM - Non-Rapid Eye Movement
PD - Parkinson Disease
PDF - Probability Distribution Function
PSD - Power Spectrum Density
RBD - Rapid eye movement Behavior Disorder
REM - Rapid Eye Movement
RLS - Recursive least-squares
RLSL - Recursive least-squares lattice

ROC - Receive Operating Characteristics

RWPE - Relative Wavelet Packet Energy

SEM - Spectral error measure

SPSS - Statistical Package for the Social Science

STFT - Short-time Fourier transform

SYM - Symplet Wavelet TP - True Positive

VAG - Vibroarthrogram

WPT - Wavelet Packet Transform

WT - Wavelet Transform

Publications

Journal Papers

- M. Shokrollahi, S. Krishnan, D. Jewell, B.J. Murray, "Parametric Modeling of EMG in Sleep", Under Review, *Biomedical Signal Processing and Control*, August 2009.

Conference Papers

- M. Shokrollahi, S. Krishnan, D. Jewell, B.J. Murray, "Autoregressive and Cepstral Analysis of Electromyogram in REM Sleep", In Proc., *World Congress in Medical Physics and Biomedical Engineering (WC)*, September 2009.
- M. Shokrollahi, S. Krishnan, D. Jewell, B.J. Murray, "Analysis of the Electromyogram of Rapid Eye Movement Sleep using Wavelet Techniques", In Proc., *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009)*, September 2009.
- M. Shokrollahi, S. Krishnan, D. Jewell, B.J. Murray, "Analysis of the Electromyogram of Rapid Eye Movement Sleep", *ELCE Graduate Research Symposium: Innovations*, Ryerson University, April 2009.

Bibliography

- [1] R. M. Rangayyan. "Biomedical signal analysis: a case-study approach", *New York, N.Y.: Wiley-Interscience*, 2002.
- [2] J. Clarke. "Mortality from Parkinson's disease", *Neurol Neurosurg Psychiatry*, 68: 254-255, 2000.
- [3] Data 360. "X page U.S. Death Rates by Cause", <http://www.data360.org/dsg.aspx>.
- [4] J. M. Monti, and D. Monti. "Sleep and quality of life in clinical medicine", *Humana Press, NJ*, PP. 29-51, 2008.
- [5] "National Sleep Foundation - The Latest Sleep News and Information", <http://www.sleepfoundation.org>.
- [6] C. M. Shapiro, M. J. Flanigan. "Function of Sleep", *AMC of Sleep Disorders. BMJ Publishing Group*, London 1993.
- [7] Faculty of the Harvard Medical School. "Sleep Mechanics: Sleep Disorder", <http://www.aolhealth.com/sleep-disorders/learn-about-it/sleep-mechanics>.
- [8] E. Estrada, H. Nazeran, J. Barragan, J. R. Burk, E. A. Lucas, and K. Behbehani. "EOG and EMG: Two Important Switches in Automatic Sleep Stage Classification", *Proceeding of the 28th Annual International Conference, New York City, USA*, August 2006.
- [9] N. Takeuchi, N. Uchimura, Y. Hashizume. "Parasomnia Melatonin therapy for REM sleep behavior disorder", *Psychiatry and Clinical Neuroscience*, 55, 267-269, 2001.
- [10] Rechtschaffen and A. Kales. "A manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages in Human Subject", *Washington, DC:US Government Printing Office*, 1968
- [11] M. S. Aldrich. "Principles and Practice of Sleep Medicine", *Boston: W.B. Saunders Company*, PP. 1051-1058, 2000.
- [12] M. A. L. Nicolelis. "Methods for Neural Ensemble Recordings", *Second Edition (Frontiers in Neuroscience). Null: CRC*, Chapter 8, 2007.

- [13] M. M. Ohayon. "Epidemiology of insomnia: what we know and what we still need to learn", *Sleep Medicine Reviews*, 6 (2): 97-111, 2002.
- [14] C. M. Morin, M. Leblanc, M. Daley, J. P. Gregoire, C. Merette. "Epidemiology of insomnia: prevalence, self-help treatments, consultations, and determinant of help-seeking behaviors", *Sleep Medicine*, 7: 123-130, 2006.
- [15] G. St-Jean, C. H. Bastien. "Sleep and quality of life in clinical medicine", *Humana Press, NJ*, PP. 37-45, 2008.
- [16] D. Leger, and S. R. Pandi-Perumal. "Sleep Disorders Their Impact on Public Health", *Information Healthcare, NY*, 2006.
- [17] S. Chokroverty. "Sleep disorders medicine basic science, technical considerations, and clinical aspects", *Butterworth-Heinemann, Boston*, 1999.
- [18] S. H. Carlos, M. W. Mahowald. "REM Sleep Behavior Disorder: Clinical, Developmental, and Neuroscience Perspectives 16 Years After its Formal Identification in Sleep", *Sleep*, 25, 120-38, 2002,
- [19] B. Frauscher, A. Iranzo, B. Hogl. "Quantification of electromyographic activity during REM sleep in multiple muscles in REM sleep behavior disorder", *Sleep*, 31 (5):724-31, May 2008.
- [20] G. Borreguero, A. B. Caminero, D. L. Llave. "Decreased phasic EMG activity during rapid eye movement sleep in treatment-naïve Parkinson's disease: effects of treatment with levodopa and progression of illness", *Movement Disorder* 17(5):934-41, September 2002.
- [21] C. H. Schenck, M. W. Mahowald. "Rapid Eye Movement Sleep Parasomnia", *Neurologic Clinics*, 2005
- [22] B. F. Boeve , M. H. Silber, and T. J. Ferman. "REM Sleep Behavior Disorder in Parkinson's Disease and Dementia with Lewy Bodies". *Journal of Geriatric Psychiatry and Neurology* 17th ser. 146(2004) 146-57.
- [23] D. K. Lindner. "Introduction to signals and systems", *WCB/McGraw-Hill, Boston*, 1999.
- [24] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck. "Discrete-Time Signal Processing", *Upper Saddle River: Prentice Hall*, 1999.
- [25] C. L. Phillips, J. M. Parr. "Signals, systems, and transforms", *Upper Saddle River, N.J: Prentice Hall*, 1999.
- [26] J. V. Vegte. "Fundamentals of Digital Signal Processing", *Upper Saddle River: Prentice Hall, NJ, USA*, 2001.

- [27] J. G. Proakis. "Digital signal processing principles, algorithms, and applications", *Prentice-Hall, Inc. Upper Saddle River, NJ, USA*, 1996.
- [28] D. Manolakis, G. Dimitris. "Statistical and adaptive signal processing spectral estimation, signal modeling, adaptive filtering, and array processing", *McGraw-Hill, Boston*, 2000.
- [29] S. Ross. "Introduction to Probability Models", *Academic Press, USA, California*, 2003.
- [30] S. Krishnan, R. M. Rangayyan, C. D. Bell, and C. B. Frank. "Adaptive Time-Frequency Analysis of Knee Joint Vibroarthrographic Signals for Noninvasive Screening of Articular Cartilage Pathology", *IEEE Transactions On Biomedical Engineering*, 47,6, pp. 773-783, June 2000.
- [31] Z.M.K. Moussavi, R.M. Rangayan, G.D. Bell, C.B. Frank, K.O. Ladly, and Y.T. Zhang. "Screening of Vibroarthrographic signals via Adaptive Segmentation and Linear Predation Modeling", *IEEE Transactions on Biomedical Engineering* , 43(1):15-23, 1996
- [32] U. Appel, A. V. Brandt. "Adaptive sequential segmentation of piecewise stationary time series," *Journal of Information Science*, 29:27-56, 1983.
- [33] G. Bodenstein and M. Praetorios. "Feature Extraction from the Electroencephalogram by Adaptive Segmentation", *Proceeding of the IEEE* 65, pp. 642-52, 1977.
- [34] D. Michael, J. Houchin. "Automatic EEG analysis: a segmentation procedure based on the autocorrelation function", *Electroencephalography and clinical neurophysiology*, 46, pp. 232-35, 1979.
- [35] S. Tavathia, R. M. Rangayyan, C. B. Frank, G. D. Bell, K. O. Ladly, and Y. T. Zhang. "Analysis of knee vibration signals using linear prediction," *IEEE Transactions on Biomedical Engineering* 39 pp. 959-70, 1992.
- [36] S. Krishnan, and R.M Rangayyan. "Recursive Least Square Lattice Based Adaptive Segmentation and Autoregressive Modeling of Knee Joint Vibroarthrographic signals", *1996 Canadian Conference on Electrical and Computer Engineering (CCECE'96)*, pp. 339-342, 1996.
- [37] S. Krishnan. "Adaptive Signal Processing of Knee Joint Vibroarthrographic signals", *Master's Thesis, Dept. of Electrical and Computer Engineering, The University of Calgary, Alberta*, 1996.
- [38] M.G. Larimore, C.R. Johnson, and J.R. Treichler. "Theory and Design of Adaptive Filters", *Upper Saddle River: Prentice Hall*, 2001.
- [39] T. Wigren, A. E. Nordsjo. "Compensation of the RLS Algorithm for Output Nonlinearities", *IEEE Transactions on Automatic Control* 44, 10, October 1999.

- [40] A. Benallal, A. Gilloire. "Improvements of the Tracking Capability of the Numerically Stable Fast RLS Algorithms for Adaptive Filtering," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-89)*, 10, pp. 1031-34, May 1989.
- [41] Pattichis, S. Constantinos, and G. E. Andreas. "Autoregressive and Cepstral Analysis of motor unit action potential", *Journal of Medical Engineering and Physics* 21, pp. 405-419, 1999.
- [42] Hefftner, Gisela, W. Zucchini, and G. G. Jaros. "The Electromyogram (EMG) as a Control Signal for Functional Neuromuscular Stimulation-PartI: Autoregressive Modeling as a Means of EMG Signature Discrimination", *IEEE Transactions on Biomedical Engineering*, 4, 35, pp. 230-37, 1988
- [43] L. Liu, J. He, and G. Palm. "Signal Modeling for Speaker Identification," *IEEE International Conference on Acoustics, speech, and Signal Processing*, 2, pp. 665-668, 1996.
- [44] A. V. Oppenheim, and R. W. Schafer. "From Frequency to Quefrency: A history of the Cepstrum," *IEEE Signal Processing Magazine* pp. 95-106, September 2004.
- [45] B.P. Bogert, M.J.R. Healy, and J.W. Tukey. "The quefrency alanalysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking", in *Time Series Analysis*, M. Rosenblatt, 209243, 1963.
- [46] W. Kang, J. R. Shiu, C. K. Chen, J.S. Lai, H.W. tsao, T.S. Kuo. "The application of cepstral coefficients and maximum likelihood method in EMG pattern recognition", *IEEE Transactions on Biomedical Engineering* 42, 8, pp. 777-85, 1997.
- [47] J.D. Markel, A. H. Gray, "Linear Prediction of Speech", *Springer-Verlag New York, Inc. NJ, USA*, 1976.
- [48] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification", *Journal of Acoustic Society Amer.*, 556, pp. 1304-1312, June 1974.
- [49] A.L. Graps. "An introduction to wavelets", *IEEE Computational Sciences and Engineering*, 2, 2, pp. 50-61, 1995.
- [50] C. K. Chui. "Introduction to wavelets", *Boston Academic Press*, 1992.
- [51] C. K. Chui. "Wavelets a tutorial in theory and applications", *Boston Academic Press*, 1992.
- [52] G. Bachman, L. Narici, and E. Beckenstein. "Fourier and wavelet analysis". *Springer: New York*, 2000.

- [53] A. Khademi. "Multiresolution Analysis for Classification and Compression of Medical Images", *Master's Thesis, Department of Electrical and Computer Engineering. Ryerson University, Toronto, ON, Canada*, 2006.
- [54] S. Mallat. "A wavelet tour of signal processing", *New York: Academic Press*, pp. 98-103, 1998.
- [55] M.I. Todorovska. "Estimation of instantaneous frequency of signals using the continuous wavelet transform", *University of Southern California, Department of Civil Engineering*, pp. 01-07 2001.
- [56] R. N. Leao, J. A. Burne. "Continuous wavelet transform in the evaluation of stretch reflex responses from surface EMG", *Journal of Neuroscience Methods*, 133,115-125, 2004.
- [57] W. Qiao H.H. Sun. "Continuous Wavelet Analysis as an Aid in the Representation and Interpretation of Electrogastrographic Signals", *Biomedical Engineering Conference, 1996., Proceedings of the 1996 Fifteenth Southern*, pp. 140-141, 1996.
- [58] G. Strang. "Wavelets and filter banks", *Wellesley Cambridge Press*, 1996.
- [59] J . Pauk. "Different Technique for EMG signal processing", *Vibròmechanika. Journal of Vibroengineering* 10, pp. 571-576, 2008.
- [60] K. Fukunaga. "Introduction to statistical pattern recognition", *Academic Press, Boston*, 1990.
- [61] W. Zhao, R. Chellappa and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition", *Automatic Face and Gesture Recognition*, pp. 336-341, 1998.
- [62] R. O. Duda, P. E. Hart and D. G. Strok. "Pattern classification", *USA: Wiley*, 2001.
- [63] T. S. Tabatabaei. "Speech-Based Human Emotion Recognition", *Master's Thesis, Department of Electrical and Computer Engineering. Ryerson University, Toronto, ON, Canada*, 2007.
- [64] H. C. Kim, D. J. Kim, and S. Y. Bang. "Face Recognition Using LDA Mixture Model", *Prococessing of 16th International Conference on Pattern Recognition*, 2, pp. 925-928, 2002.
- [65] Y. H. Hu, S. Palreddy and W. Tomkins. "A patient adaptable ECG beat classification using a mixture of experts approach," *IEEE Transactions on Biomedecial Engineering*, 44, pp. 891-900, September 1997.

- [66] T. Sugiura, H. Hirata, Y. Harada and T. Kazui. "Automatic discrimination of arrhythmia waveforms using fuzzy logic", *Processing of 20th Annual International Conference of the IEEE Engineering in Medical and Biology society*, 20, 1, 108-111, 1998.
- [67] SPSS Inc. "SPSS Advanced Statistics User's Guide", *Chicago, IL*, 1990 .
- [68] J. Pardey, S. Roberts, and L. Tarassenko. "A review of parametric modelling techniques for EEG analysis", *Journal of Medical Engineering and Physics*, 18, 211, January 1996.
- [69] H. Akaike. "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, AC-19, pp. 716-723, December 1974.
- [70] W. Gresch, D.R. Sharpe. "Estimation of power spectra with finite-order autoregressive models", *IEEE Transactions on Automatic Control*, AC 18, pp. 367-369, August 1973.
- [71] B.S. Atal. "Automatic Recognition of speakers from their voices", *IEEE Prococessing* 64: pp. 460-75, 1976.
- [72] A.H. Gary and J.D. Markel. "Distance measure of speech processing", *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24, 5, pp. 380-391, June 1976.
- [73] K. Tokuda, T. Kobayashi, S. Imai. "Adaptive Cepstral Analysis of Speech", *IEEE Transactions on Speech and Audio Processing*, 3, 6, November 1995.
- [74] K. Woods, K. W Bowyer. "Generating ROC Curves for Artificial Neural Networks", *IEEE Transactions on Medical Imaging* 16, 3, June 1997.
- [75] M. Vetterli, C. Herley. "Wavelets and Filter Banks: Theory and Design", *IEEE Transactions On Signal Processing* 40, 1992.
- [76] P. Flandrin. "Time-frequency/time scale analysis". *Academic Press, San Diego*, 1999.
- [77] M. Gazzoni, D. Farina, D. Merletti. "A new method for the extraction and classification of single motor unit action potentials from surface EMG signals", *Journal of Neuroscience Methods*, 136, 2004.
- [78] X. Hu, Z. Wang, X. Ren. "Classification of surface EMG signal using relative wavelet packet energy", *Computer Methods and Programs in Biomedicine*, 79, pp. 189-195, 2005.
- [79] M. O. Diab, C. Marque, M. Khalil. "An unsupervised classification method of uterine Electromyography signals using wavelet decomposition", *IEEE Processing of the 26th Annual International conference of the EMBS*, September 2004.
- [80] M. Arveti, G. Gini, M. Folgheraiter. "Classification of EMG signals through wavelet analysis and neural networks for controlling an active hand prosthesis", *IEEE 10th International Conference On Rehabilitation Robotics 2007*, 531-536, June 2007.

- [81] L. Brechet, F. M. Lucas, C. Doncarli, and D. Farina. "Compression of Biomedical Signals with Mother Wavelet Optimization and Best Basis Wavelet Packet Selection", *IEEE Transactions on Biomedical Engineering* 54: pp. 2186-192.
- [82] A. Khademi, S. Krishnan. "Shift Invariant Discrete Wavelet Transform Analysis for retinal image classification", *Medical and Biological Engineering and Computing*, 45: pp. 1211-1222, 2007.
- [83] S. Ioannidou, V. Karathanassi. "Investigation of the Dual-Tree Complex and Shift Invariant Discrete Wavelet Transforms on Quickbird Image Fusion", *IEEE Geoscience and Remote Sensing Letters*, 4,1, Janouary 2007.
- [84] M. Hilton. "Wavelet and wavelet packet compression of electrocardiograms", *IEEE Transactions on Biomedical Engineering*, 44,(5) pp. 394402, May 1997.
- [85] A. Maitrot, M. F. Lucas, C. Doncarli, and D. Farina. "Signal-dependent wavelet for electromyogram classification", *Medical and Biological Engineering and Computing*, 43, pp. 48792, 2005.
- [86] S. Mallat. "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11,7 pp. 674693, July 1989.
- [87] M. Flanders. "Choosing a wavelet for single-trial EMG", *Journal of Neuroscience Methods*, 116, pp. 135-177, 2002.
- [88] I. Romero, N. R. Grubb, G. R. Clegg, C. E. Roberston, P. S. Addison, J. N. Waston. "T-Wave Alternans Found in Preventricular Tachyarrhythmias in CCU Patients Using a Wavelet Transform-Based Methodology", *IEEE Transactions on Biomeical Engineering*, 55, 11, November 2008.
- [89] D. Farina, M. F. Lucas, C. Doncarli. "Opimized Wavelets for Blind Separation of Non-statioanry Surface Myoelectric Signal". *IEEE Transaction on Biomedical Engineering* 55, 1, January 2008.
- [90] R. K. Sinha. "Artificial Neural Network and Wavleet Based Automated Detection of Sleep Spindeles, REM Sleep and Wake States", *Journal of Medical Systems*, 32: pp. 291-299, 2008.
- [91] M. Unser. "Texture Classification and Segmentation using Wavelet Frames", *IEEE Transaction On Image Processing*, 4, 11, November 1995.
- [92] J. Romberg, H. Choi, R. Baraniuk, N. Kingsburry. "Multiscale classification using complex wavelets and hidden markov tree models", *IEEE Proceeding on Image Processing*, 2, pp. 371-374, September 2000.

- [93] C. S. Burrus, R. A. Gopinath, G. Guo. "Introduction to wavelets and wavelet transforms a primer", *Upper Saddle River, N.J: Prentice Hall*, 1998.
- [94] M. Shokrollahi, S. Krishnan, D. Jewell, B.J. Murray. "Autoregressive and Cepstral Analysis of Electromyogram in REM Sleep", In Proc., *World Congress in Medical Physics and Biomedical Engineering (WC)*, September 2009.
- [95] M. Shokrollahi, S. Krishnan, D. Jewell, B.J. Murray. "Analysis of the Electromyogram of Rapid Eye Movement Sleep using Wavelet Techniques", In Proc., *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009)*, September 2009.