SCREENING OF SLEEP MOVEMENT DISORDERS VIA EMG SIGNAL ANALYSIS

by

Mehrnaz Shokrollahi,   B.Eng, M.A.Sc.,

Ryerson University, Toronto, Canada, 2007,

Ryerson University, Toronto, Canada, 2009,

A dissertation

presented to Ryerson University

in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Electrical and Computer Engineering.

Toronto, Ontario, Canada, 2015

Author's Declaration

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

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

I further authorize Ryerson University to reproduce this dissertation 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.

I understand that my dissertation may be made electronically available to the public.

Abstract

Screening of Sleep Movement Disorders Via EMG Signal Analysis

© Mehrnaz Shokrollahi 2015

Doctor of Philosophy

in the Program of

Electrical and Computer Engineering,

Ryerson University.

It is estimated that 50 to 70 million Americans suffer from a chronic sleep disorder, which hinders their daily life, affects their health, and incurs a significant economic burden to society. Untreated Periodic Leg Movement (PLM) or Rapid Eye Movement Behaviour Disorder (RBD) could lead to a three to four-fold increased risk of stroke and Parkinson's disease respectively. These risks bring about the need for less costly and more available diagnostic tools that will have great potential for detection and prevention. The goal of this study is to investigate the potentially clinically relevant but under-explored relationship of the sleep-related movement disorders of PLMs and RBD with cerebrovascular diseases.

Our objective is to introduce a unique and efficient way of performing non-stationary signal analysis using sparse representation techniques. To fulfill this objective, at first, we develop a novel algorithm for Electromyogram (EMG) signals in sleep based on sparse representation,

and we use a generalized method based on Leave-One-Out (LOO) to perform classification for small size datasets.

In the second objective, due to the long-length of these EMG signals, the need for feature extraction algorithms that can localize to events of interest increases. To fulfill this objective, we propose to use the Non-Negative Matrix Factorization (NMF) algorithm by means of sparsity and dictionary learning. This allows us to represent a variety of EMG phenomena efficiently using a very compact set of spectrum bases.

Yet EMG signals pose severe challenges in terms of the analysis and extraction of discriminant features. To achieve a balance between robustness and classification performance, we aim to exploit deep learning and study the discriminant features of the EMG signals by means of dictionary learning, kernels, and sparse representation for classification. The classification performances that were achieved for detection of RBD and PLM by means of implicating these properties were 90% and 97% respectively.

The theoretical properties of the proposed approaches pertaining to pattern recognition and detection are examined in this dissertation. The multi-layer feature extraction provide strong and successful characterization and classification for the EMG non-stationary signals and the proposed sparse representation techniques facilitate the adaptation to EMG signal quantification in automating the identification process.

Acknowledgements

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 advice and motivation. His guidance and moral support have helped me to succeed beyond my wildest dreams.

I would also like to thank my parents, Shohreh and Hadi, for their continual support and encouragement, which has given me the ability to pursue my goals with confidence and full determination. I am forever grateful to them for this. I would also like to express my special thanks to my husband, Kianoosh. His motivation, support, and encouragement gave me the strength to peruse my goals and dreams. I am forever delighted for having him in my life. I would also like to express my special thanks to my sister, Elnaz, who has also been my long-term friend as well as a colleague throughout my student life at Ryerson University for both Undergraduate and Master's studies. Her motivation, support and encouragement gave me the strength to pursue my goals and dreams. In addition, I am blessed for having my brother, Peyman, whose love, encouragements and enlightening and informative discussions have been invaluable to me. I would also like to thank my brother-in-law Kiarash for his help and support especially in preparing the LaTeX version of my dissertation.

Dedication

I lovingly dedicate this dissertation to the love of my life Kianoosh, my parents

Shohreh & Hadi and my super awesome brother and sister, Peyman & Elnaz

as a token of humble and sincere appreciation for their invaluable love,

enormous support and continuous encouragement. . .

Table of Contents

List of Figures

List of Tables

List of Acronyms

ARWMC Age Related White Matter Change

ANN Artificial Neural Networks

AR Autoregressive

ARAS Ascending Reticular Activating System

ATP Adenosine Triphosphate

CLT Central Limit Theorem

CNS Central Nervous System

CPAP Continuous Positive Airway Pressure

CT Computer Tomography

DWI Diffusion Weighted Image

ECG Electrocardiogram

EEG Electroencephalogram

EMG Electromyogram

EOG Electrooculogram

FLAIR Fluid Attenuated Inversion Recovery

GABA γ-Aminobutyric Acid

HF Heart Failure

ICA Independent Component Analysis

KSRC Kernel Sparse Representation for Classification

LDA Linear Discriminant Analysis

LMO Leave-M-Out

LOO Leave-One-Out

LP Linear Programming

MOD Method of Optimal Directions

MRI Magnetic Resonance Imaging

NMF Non-Negative Matrix Factorization

NREM Non Rapid Eye Movement

NSF National Sleep Foundation

OSA Obstructive Sleep Apnea

PCA Principal Component Analysis

PLM Periodic Leg Movement

PSG Polysomnography

RBD Rapid Eye Movement Behaviour Disorder

REM Rapid Eye Movement

RLS Restless Leg Syndrome

SCN Suprachiasmatic Nucleus

SRC Sparse Representation for Classification

STFT Short Time Fourier Transform

SVD Singular Value Decomposition

SVM Support Vector Machines

TF Time-Frequency

TIA Transient Ischemic Attack

VLPO Ventrolateral Preoptic

WMH White Matter Hyperintensities

Chapter 1

Introduction

"The love of the Miraculous − Dreams as Suggestion of good or evil Spirits − Comparatively little attention given to the Scientific Study of Sleep − Changes in the Respiration during Sleep − Variation in volume of different parts of the body − The Pulse slackens during Sleep − Dilation of the Surface Vessels − Contraction of the Vessels of the Brain − Increase of Perspiration during Sleep − Digestion and the Activity of the Internal Organs − State of the Blood during Sleep − State of the Peripheral Nervous System and of the Spinal Cord − State of the Muscular System and the Motor Nerves − Why is it possible to awake at any movement − Determination of the depth of Sleep − State of the Brian during Sleep − State of Will and Attention − Condition of the Pupils and of the Visual Axes − Why we close our Eyes in Sleep − Localising Theories of Sleep − The Vasomotor Theory of Sleep − Chemical Theories of Sleep − Histological Theories of Sleep − Defects of the Chief Theories − Yawning − The Psycho −

1

Physiological Theory of Sleep − Sleep in Children and in Savages − Absence of Sleep in animals deprived of Brain − Caspar Hauser − Sleep in old age − Sleep regarded as a useless or injurious habit − Experiments on prolonged absence of Sleep in Animals − In Man − Sleep is the resting − time of Consciousness" [1].

1.1   Sleep

S LEEP is a normal physiological state that has been demonstrated in every animal species, from insects to mammals. It is one of the most momentous human behaviours, roughly occupying one-third of our lives. In addition, it is characterized by apparently unconscious organisms during which they are not responsive to stimuli for a short period of time. On the contrary, brain activity may still be periodically high. Owing to relatively rapid reversibility, sleep is distinguished from states such as coma or hibernation. The amount of sleep obtained during the night varies among individuals, though this depends on age, medication intakes and psychiatric conditions as well. Most adults need about 7-9 hours per 24-hour period in order to function optimally, although there are short sleepers and long sleepers who function ideally when they sleep for 6 or 12 hours respectively. Sleep is an unconscious state which can be in part modified by sensory stimulations. In adults, sleep is accomplished when a number of changes in the Central Nervous System (CNS) bring about a set of behavioural, physiological, and psychological changes. During the past hundred years, sleep abnormalities have attracted significant interest in the scientific community, as they play an important role in human life. Although, much progress has been made in detecting these abnormalities, sleep analyses remain a challenging task due to the

fact that are lengthy and complex. Also, because of the increase in the average age expectancy, there is more need for scientists to analyze the sleep signals, as sleep abnormalities are highly proportional to the increase of age. However, to be able to diagnose, detect, and prevent sleep disorders, the physiology of the sleep has to fully be understood. The next section aims for an understanding of this physiology.

1.2   Sleep Physiology

The brain awakes! Neurons in the upper pons produce acetylcholine. This activates signals that are responsible for channeling part of the thalamus to the cerebral cortex, which is the site of consciousness. The neurotransmitters that are responsible for triggering the cerebral cortex to receive signals from the thalamus are noradrenaline, serotonin, histamine and dopamine. These neurotransmitters are naturally produced in the pons, hypothalamus and other nearby regions [2].

The brain tires! The neuron activity in the Ventrolateral Preoptic (VLPO) nucleus get stimulated by the accumulation of adenosine as the energy carrier Adenosine Triphosphate (ATP) breaks down. The circadian control, or the brain's master clock is controlled by Suprachiasmatic Nucleus (SCN). The SCN comprises of neurons that fire in a 24-hour cycle to stimulate the VLPO. This clock is regulated by signals from the retina during the day and by melatonin from the pineal gland at night [2].

The brain sleeps! γ-Aminobutyric Acid (GABA) and galanin are released when VLPO neurons are activated. Subsequently, these bind to receptors in the hypothalamus and pons

to inhibit the ascending arousal system. Figure 1.1 demonstrates the anatomy of sleep in the brain.
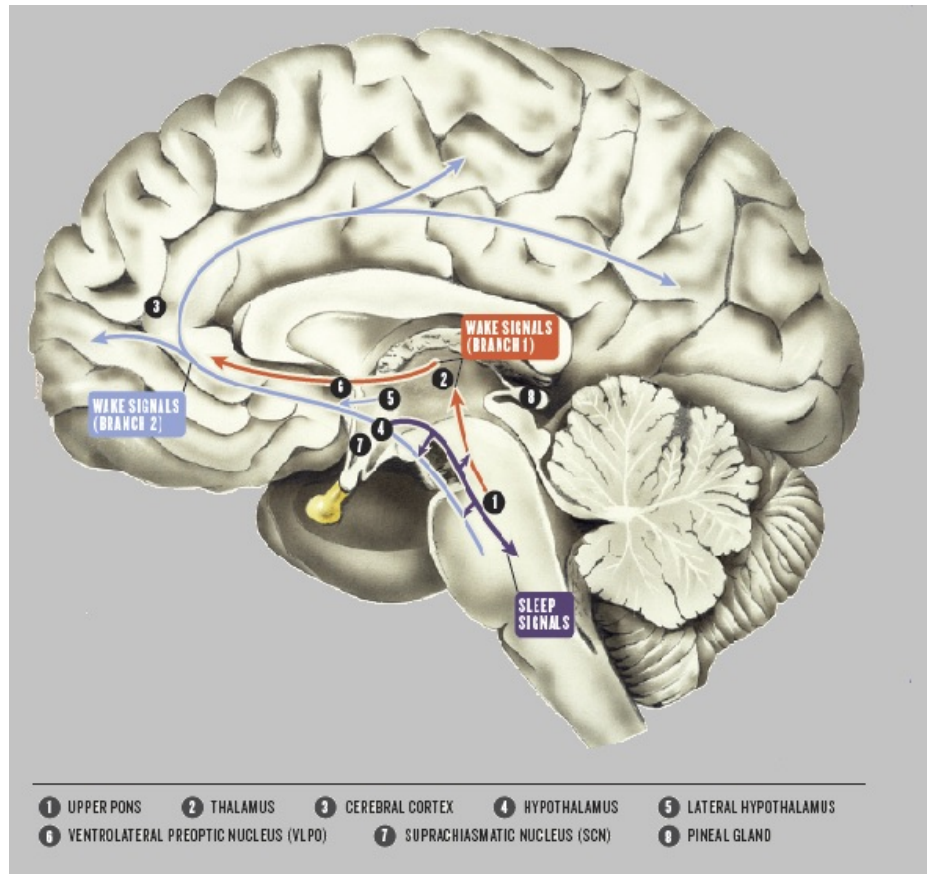
Figure 1.1: The anatomy of Sleep [2]

1.3 Sleep States

Wakefulness, Non Rapid Eye Movement (NREM) sleep, and Rapid Eye Movement (REM) sleep are the three distinct states of existence in sleep. Each state has it's unique characteristics, behaviours, and physiologic patterns. In addition, each state has specific neuro-

physiologic mechanisms associated with its generation and control. The structures of the brainstem use various neurotransmitters to manipulate higher brain structures in the midbrain and cortex [3]. The Ascending Reticular Activating System (ARAS), a set of connected nuclei in the brain that is responsible for regulating arousal and sleep-wake transitions, provides cholinergic, noradrenergic, and glutaminergic stimulation to the thalamus, hypothalamus, and basal forebrain respectively. These result in cholinergic and glutaminergic activity of the cortex. The excited cortex on the other hand manifests wakefulness which exhibits desynchronized brain activity. Sleep onset depends on various factors that influence the nucleus tractus solitarius which in turn result in inactivation of inhibitory GABAergic thalamocortical projections to the cortex. The alteration between NREM sleep and REM sleep is controlled by three different neurons, including noradrenergic neurons in the loci coeruleus, serotoninergic neurons in the raphe called REM-off cells and cholinergic neurons in the nucleus reticularis pontis oralis called REM-on cells. Other brain structures are involved in generation and control of REM sleep-related phenomena, such as eye movement and muscle atonia. The properties of wakefulness, NREM and REM are characterized by the amount of sympathetic and parasympathetic tone. As such, during wakefulness, there is a boost in sympathetic tone and a significant decrease in parasympathetic tone that maintains most organ systems in an active state. To create a state of reduced activity during NREM sleep, the sympathetic tone decreases and parasympathetic tone increases. REM sleep is characterized by increased parasympathetic activity and variable sympathetic activity associated with increased activation of certain brain functions [4, 5]. The progression of sleep stages occurs in cycles of 60 to 120 minutes throughout the sleep period. Various circadian environmental and ontological factors affect the pattern of sleep stage occurrence.

A disruption to any of these phenomena will cause sleep disturbance, which in turn will affect the quality of life of individuals. While strong evidence exists to support an increase in knowledge of normal sleep and sleep disorders, many unanswered questions remain. According to the 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 [6]. In general, sleep and sleep−related problems play a role in a large number of human disorders, and affect every field of medicine. Clinicians and other health care professionals receive extensive training in order to be sufficiently qualified to detect and prevent diseases [7]. Although the skills acquired by these medical facilitators are quite well developed, 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 neuro-degenerations and treatment responses [8]. The introduction of a number of new techniques, during the past few decades, including PSG, surface measurements of CNS activity, eye movements, and muscle activity, have allowed sleep to be described in electrophysiological terms [9]. The sleep−wakefulness cycle can be characterized by the polysomnographic recording of three basic parameters: Electroencephalogram (EEG), Electrooculogram (EOG), and EMG, where the three sleep states can be generated [10]. The following will show the characteristics of each parameter and sleep state in detail [11, 12].

1.4 Electrophysiological Parameters of Sleep

1.4.1 EOG

The EOG signal is frequently used for recordings of the eye movements in sleep research [13]. The horizontal and vertical eye movements aid doctors and sleep specialists in analyzing the effect of medical drugs prescribed to a patient with different diseases, such as depression. In addition, EOG has been widely used in PSG to detect different sleep stages. Figure 1.2 displays the EOG records in sleep for horizontal (left eye and right eye) and vertical movements.

1.4.2 EEG

The brain consists of two hemispheres, the cerebellum, and brain stem. The two hemispheres are separated by the longitudinal fissure across which there is a large connective band of fibres called the corpus callosum [14]. The brain stem is a complex structure which includes midbrain, pons medulla and reticular formation. The EEG is a recording of the electrical activity of the brain, from the scalp. These recorded waveforms reflect the cortical electrical activity of the brain [15]. Despite the advent of high-resolution anatomical imaging techniques such as Magnetic Resonance Imaging (MRI) and Computer Tomography (CT), EEG continues to be an important measurement for monitoring brain wave activities. As well, EEG is most often used for the diagnosis and treatment of mental and brain diseases and abnormalities. With appropriate interpretation methods they are emerging as a key
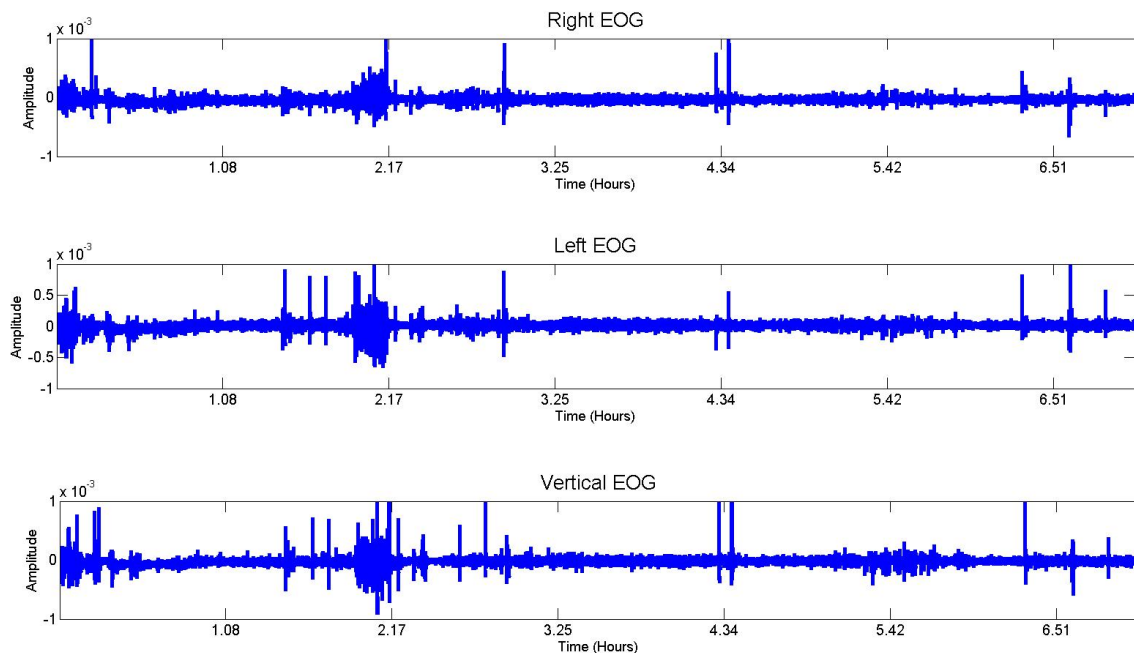
Figure 1.2: EOG signal in Sleep

methodology to satisfy the increasing global demand for more affordable and effective clinical and health care services [16]. Many brain disorders are diagnosed by visual manifestation of brain rhythms in the EEG signals. In healthy adults, the amplitudes and frequencies of such signals change from one human state to another, such as from wakefulness to sleep. Their characteristics of them also change with age. The different frequency types of the EEG waves have been used extensively in PSG to detect different sleep stages. The next subsection will briefly explain the characteristics of each brain wave.

1.4.2.1   Brain Rhythms

The frequency bands from low to high frequencies are delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ). Delta waves lie within the range of $0.5 - 4$ Hz, which are primarily associated with deep sleep but may also be present in the waking states. In this range, it is very easy to confuse artifact signals caused by chin or neck muscles. Nevertheless by applying simple signal analysis methods to the EEG, it is very easy to see when the response is caused by excessive movements [16].

Theta waves lie within the range of $4 - 7.5$ Hz and they mostly appear towards drowsiness or beginning of sleep stages [15, 16]. They are often associated with access to unconscious material, creative inspiration and deep meditation.

Alpha waves lie within the range of $8 - 13$ Hz and commonly appears as a series of sinusoidal shapes. They are usually found over the occipital region of the brain. They indicate a relaxed awareness without any attention or concentration. They are also best detected when eyes are closed, and are usually seen in visual cortex. An Alpha wave is reduced or eliminated by opening the eyes, hearing unfamiliar sounds, anxiety, mental concentration, or attention [16].

A beta wave is the electrical activity of the brain varying within the range of $14 - 25$ Hz. It is associated with active, busy, anxious thinking, active concentration, focus on the outside world, or solving concentrated problems.

The frequencies above 30 Hz correspond to the gamma range. Their amplitudes are lower than the other ranges and their occurrence is rare. However, detection of these rhythms can be used for confirmation of certain brain diseases.

As mentioned before, the amplitude and frequencies of these signals change as people tend to sleep. Theta waves appear at the beginning stages of sleep; delta waves appear at deep sleep stages. The depression or absence of the normal rhythm in a certain state of the subject could indicate abnormality. The presence of delta or theta waves in a wakeful adult would be considered to be abnormal. Focal brain injury and tumors lead to abnormal slow waves in the corresponding regions. Unilateral depression of a rhythm could indicate disturbances in cortical pathways.

The states of wakefulness and sleep are characterized as stages that are defined by stereotypical EEG, EMG, and EOG patterns. The wakefulness stage has an EEG pattern predominated by the alpha rhythm. With onset of stage 1 sleep, the alpha rhythm attenuates, and an EEG pattern of relatively low voltage and mixed frequency is seen. Progression to stage 2 sleep is defined by the appearance of sleep spindles or K-complexes. Further progression into the deepest sleep stages 3 and 4 is defined by the occurrence of high-amplitude, low-frequency EEG activity.

1.4.3   EMG

EMG 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 PSG, 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 $0.1 - 2s$ having an amplitude of at least $50\,\mu V$, while tonic constitutes the remainder [8, 17]. In mammals, sleep is divided into two broad states: REM sleep and NREM sleep. Each type

has a neurological and physiological characteristics associated with it. Next, we will explain these states in more detail.

1.5 Sleep Stages

Wakefulness The EEG signal has sinusoidal wave type characteristics, (alpha activity of $8 - 12$ Hz), intermixed with lower amplitude irregular beta waves ($13 - 35$ Hz). The EOG activity may be 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. Figure 1.3 demonstrates the EMG characteristics in this stage.
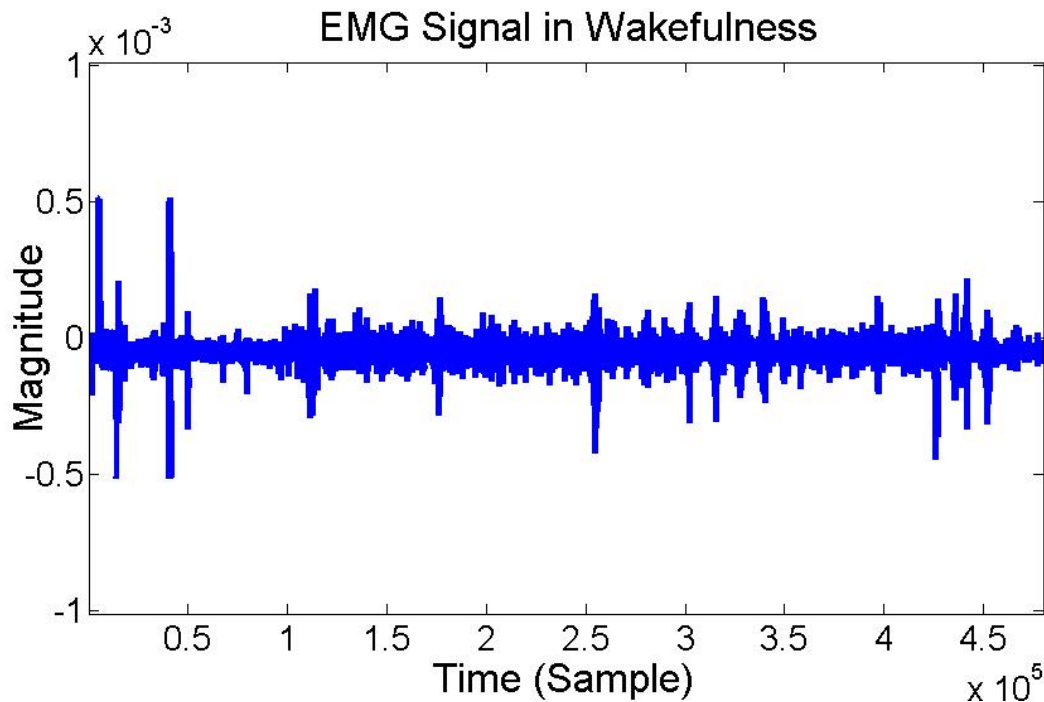
Figure 1.3: EMG Signal in Wakefulness. This segment corresponds to about 48 minutes of PSG recording.

Non-REM Sleep This is a motionless stage, eyes are closed. Sensory inputs such as noise and light no longer induce behavioural responses. In a polysomnogram, this state of sleep consists of 4 different stages:

- Stage 1: 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 theta range of $4-7$ Hz. Unless the subject is disturbed, this stage only lasts for a few minutes. Figure 1.4 shows the EMG in Non-REM sleep of Stage 1.
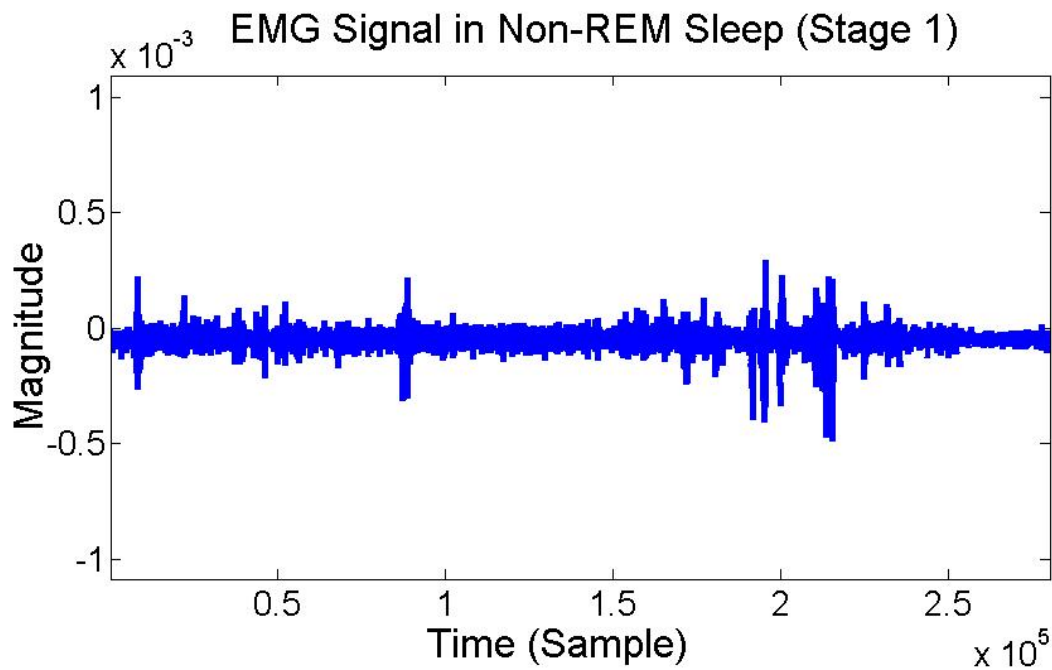
Figure 1.4: EMG Signal in Non-REM Sleep. This segment corresponds to about 48 minutes of PSG recording.

- Stage 2: 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

12

of at least $0.5\,s$. K−complexes, on the other hand, are defined to be relatively high−amplitude potentials, having a negative sharp wave followed by positive components, with a total duration of more than 0.5 . The slow waves also happen in this stage on irregular time intervals.

- Stages 3 and 4: Usually slow wave sleep is characterized by having a delta wave of $1-2$ Hz or slower, and an amplitude of $75\mu V$ or greater. If at least 20% to not more than 50% of the scoring epoch consists of delta wave activity, sleep is scored as Stage 3. Otherwise, it is classified as Stage 4.

REM Sleep : In this state dreams occur, and it is paradoxical. The brain is highly active, which is often in conjunction with bursts of REM but the muscles are completely relaxed. Yet, the flat EMG tracing is periodically interrupted by muscle twitches, and it has comparatively higher unresponsiveness compared to NREM sleep. Eyes move periodically under closed eyelids. The polysomnogram is characterized by low voltage EEG activity, which is similar to that of stage 1; theta activity of $4-7\,Hz$ exists. The REM state of sleep is characterized by two different twitches. One is the discontinuous of muscle twitches and REM, which are called the phasic events. The other is continuous processes such as desynchronizing EEG and muscle hypotonia, which are called tonic events. Figure 1.5 shows the EMG signal in REM sleep.

The need for sufficient sound sleep has worldwide importance. However, due to fast-paced lives, many people go without adequate sleep. At least 20% of the today's population suffer from a sleep disorder, in addition to $70 billion annually in lost productivity, medical bills and industrial accidents [6]. This creates the necessity of understanding and preventing of
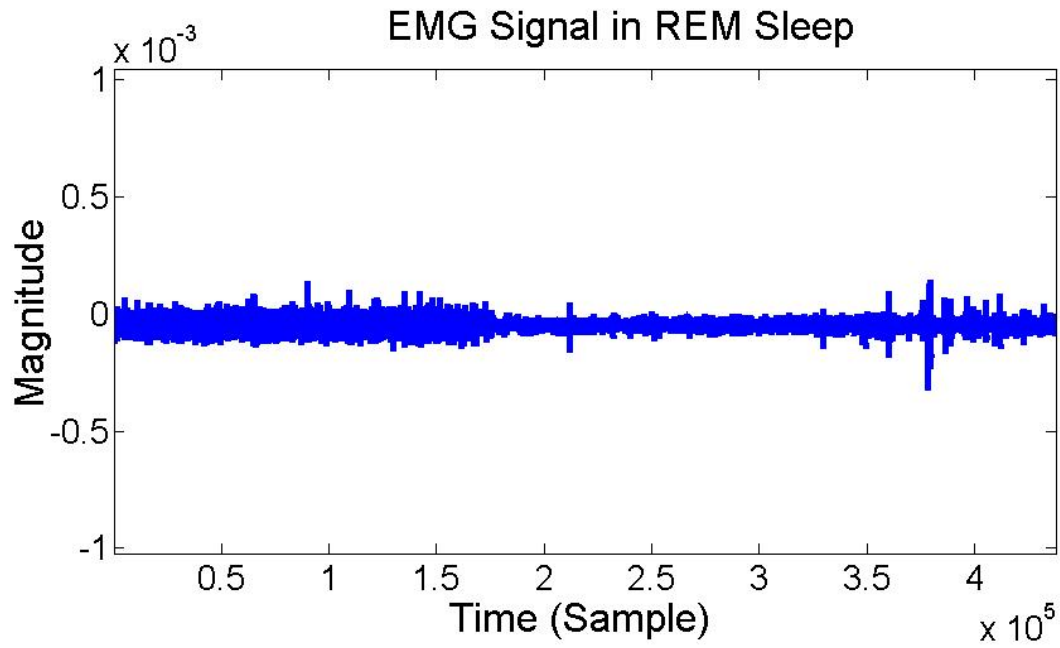
Figure 1.5: EMG Signal in REM Sleep. This segment corresponds to about 48 minutes of PSG recording.

such diseases. The next section is devoted to a general overview of different sleep disorders.

1.6   Sleep Disorders

Many sleep disorders not only diminish daytime performance, increase sleepiness, and affect mood of an individual, but can also lead to serious consequences such as high blood pressure, cardiovascular diseases, stroke, depression and even death. Therefore, recent attention has been directed to the connection of sleep disorders with other diseases. Some of the common sleep disorders are listed below:

1.6.1   Sleep Insomnia

Out of the 30% of individuals who experience insomnia symptoms, 10% of them suffer from insomnia syndrome [18, 19]. Insomnia as a symptom is a diagnostic criterion of other mental disorders such as depression, and as a syndrome it may be secondary or comorbid to another disease [7]. 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; effectively generating other problems including falling asleep while driving a vehicle.

1.6.2   Obstructive Sleep Apnea

Approximately 8% of the adult population around the world have sleep apnea. However, less than one in ten is diagnosed. Untreated sleep apnea could lead to a three- to four-fold increased risk of hypertension, stroke, or Heart Failure (HF); a potentially increased

risk of cancer mortality; and up to a four-fold increased risk of motor vehicle accidents. These risks bring about the need for less costly and more available diagnostic tools that will have great potential for detection and prevention. The current technology recommended to patients with Obstructive Sleep Apnea (OSA) is Continuous Positive Airway Pressure (CPAP). Although CPAP alleviates OSA, many patients are intolerant of it, and no therapy for OSA has been shown to improve clinical outcomes. In addition, only one in ten patients diagnosed with sleep apnea have been treated with the CPAP solutions [20]. This fact shows that robust technologies and precise tools for prevention and diagnosis of sleep apnea are still required. EMG signals have been utilized to analyze the muscle activity in OSA patients to determine the fluctuation in timing of upper airway and chest wall [21].

1.6.3   Narcolepsy

This disorder usually happens in young adults, and is illuminated by the need for naps during the day. The day time polysomnogram shows signs of drowsiness on EEG signal, associated with REM periods. On the other hand, the nighttime PSG shows an increased number of awakening periods, as well as an increased amount of stage 1. In addition to sleepiness, individuals suffering from narcolepsy are at higher risk of having accidents at home, work, or even on the road [22]. EMG signals have also been exploited in narcolepsy patients. In particular, Dauvilliers *et al.* utilized EMG signals to assess the presence of polysomnographic characteristics of REM sleep. They found higher EMG activity is present in patients with narcolepsy than the normal subjects [23].

1.6.4   Sleep Movement Disorders

1.6.4.1   Overview of Movements in Sleep

Motor activity in sleep is very common. In fact, any arousing event may cause movements; this is the primary reason for most of the abrupt changes in PSG. During sleep about 50 to 100 isolated movements of the arms and the legs typically occur. On average, axial changes of body position happen five to seven times during the night in blocks of 10 minutes occurring either before or after the REM sleep. However, during REM sleep, muscle twitching is common and may be recorded as a brief, spontaneous contraction, seen as a flicker of movement under the skin that can be recorded by means of the chin EMG recordings. In general, any other movements in sleep are exacerbated by many causes, in particular as a physical pain or discomfort, psychological disturbances, environmental factors, and a wide variety of sleep movement disorders. These disorders are explained in more detail in the following subsections.

1.6.4.2   Periodic Leg Movement Disorder

PLM are detected in approximately 80% of patients with Restless Leg Syndrome (RLS) [24], but are also seen in individuals with obstructive sleep apnea (OSA) [25], RBD [26], congestive heart failure [27], Parkinson's disease, Multiple System Atrophy [28], users of antidepressant medications [29], and even in seemingly healthy subjects, particularly the elderly [30]. To date, the largest epidemiological study reported simultaneous presence of PLM and sleep complaints in 3.9% of 18,980 subjects [31]. PLM is characterized by periodic episodes of

repetitive stereotypical triple flexion responses involving the great toe, ankle and hip, and are associated with a minimum amplitude increase of $8\mu V$ above the resting EMG voltage in the anterior tibialis surface electrode [32]. Using current scoring criteria [32], the duration of PLM is at least $0.5 - 10s$, and they recur in sequences of four or more events at $5 - 90s$ intervals. Leg movements associated with arousals from respiratory events during sleep are excluded [32]. PLM, share common physiological mechanisms with sleep-disordered breathing and in particular cardiovascular diseases [33]. The clinical significance of PLMs is unclear; however, emerging research suggests important associations with nocturnal sympathetic activation [34], inflammatory markers [35, 36], congestive heart failure [27, 37, 38], incident cardiovascular disease [39], and mortality in specific patient populations [40, 41]. Figure 1.6 shows the PLM in sleep.

1.6.4.3   REM Sleep Behaviour Disorder

RBD is characterized by elaborating movements correlated with dream during REM sleep [7, 42]. 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 [43, 44]. 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 active 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
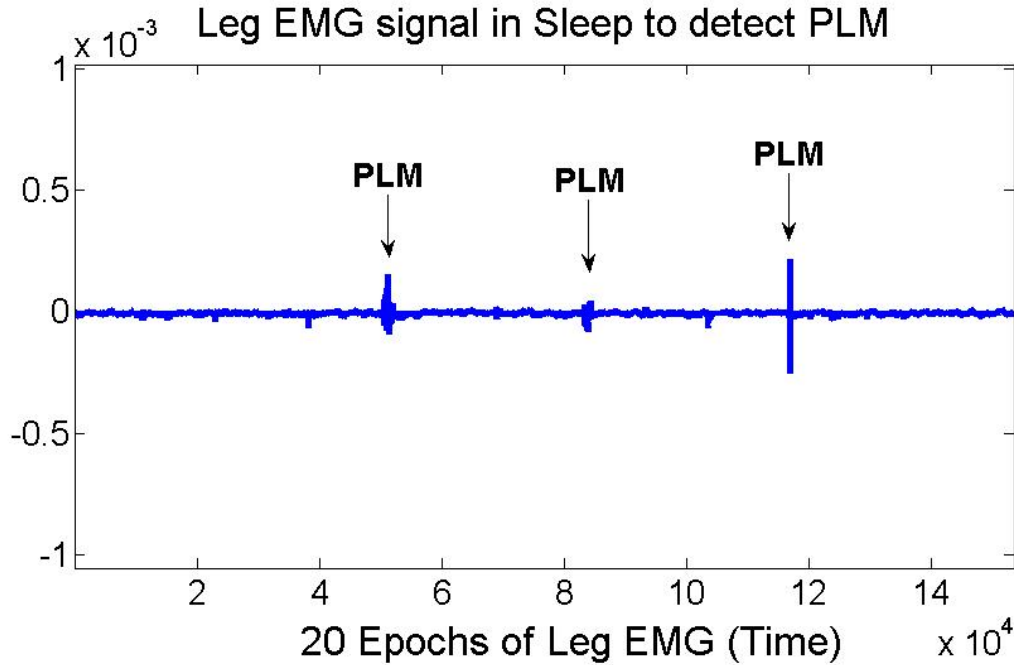
Figure 1.6: EMG in sleep with PLM markers specified This segment corresponds to about 9 minutes of PSG recording.

consequences such as paradox loss, which means loss of safe sleep. Serious injuries have the potential to occur as a result of dream acting behaviour [45]. The overnight PSG of subjects also confirms the presence of muscle tone in REM sleep. As well, most of the patients with RBD have significant abnormal REM sleep muscle activity [23]. In addition, RBD can be an early warning for the emergence of Parkinson's and other neuro-degenerative conditions antedating the illness by many years [46]. Figure 1.7 demonstrates this phenomenon.

19
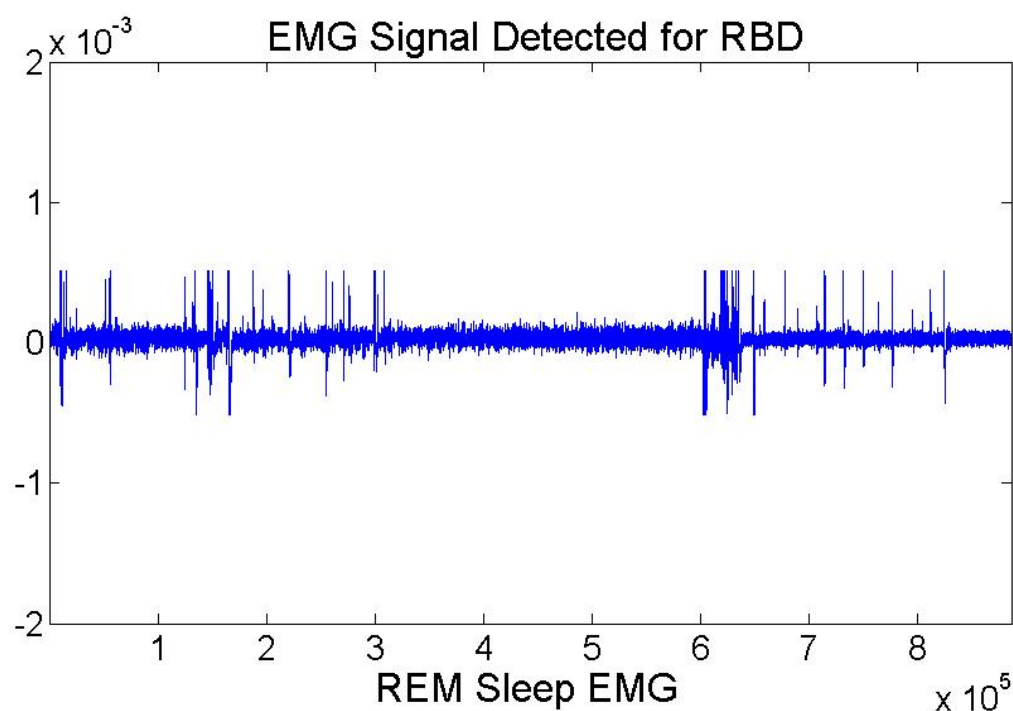
Figure 1.7: EMG Signal Detected for RBD This segment corresponds to about 55 minutes of PSG recording.

1.7   Sleep Disorder Diagnosis

Sleep and sleep-related problems play a role in a large number of human disorders, and affect every field of medicine. It is estimated that 50 to 70 million Americans suffer from a chronic sleep disorder, which hinders their daily life, affects their health and confers a significant economic burden to society. The negative public health consequences of sleep disorders are enormous and could have long-term effects, including increased risk of hypertension, diabetes, obesity, heart attack, stroke, and in some cases death. Early detection and treatment of sleep disorders can not only help clinicians apply appropriate therapeutic

procedures to control the abnormal behaviour, but can also help patients have a better quality of life. Untreated PLM or RBD could lead to a three to four-fold increased risk of hypertension, stroke, Parkinson's disease, or HF; a potentially increased risk of cancer mortality; and up to a four-fold increased risk of motor vehicle accidents. These risks bring about the need for less costly and more available diagnostic tools that will have great potential for detection and prevention. Although the skills acquired by 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.

Despite sharing common underlying pathophysiological mechanisms with sleep-disordered breathing, sleep-related movement disorders such as the closely related nocturnal PLMs and RBD have been relatively unexplored in the context of cerebrovascular disease. These are common treatable conditions that can arise after or before stroke and Parkinson's disease respectively, and preliminary evidence suggests an association with vascular risk factors and vascular events. The goal of this study is to investigate the potentially clinically-relevant but under-explored relationship of the sleep-related movement disorders of PLMs and RBD with cerebrovascular diseases. For this reason, given numerous potentially relevant health-related linkages, further investigation of sleep related movement disorder is warranted.

In addition to observational studies, PSG, imaging based techniques, and, EMG can be used for diagnosis of sleep movement related problems. Figure 1.8 lists the methods presently available for non-invasive diagnosis of sleep movement disorders.
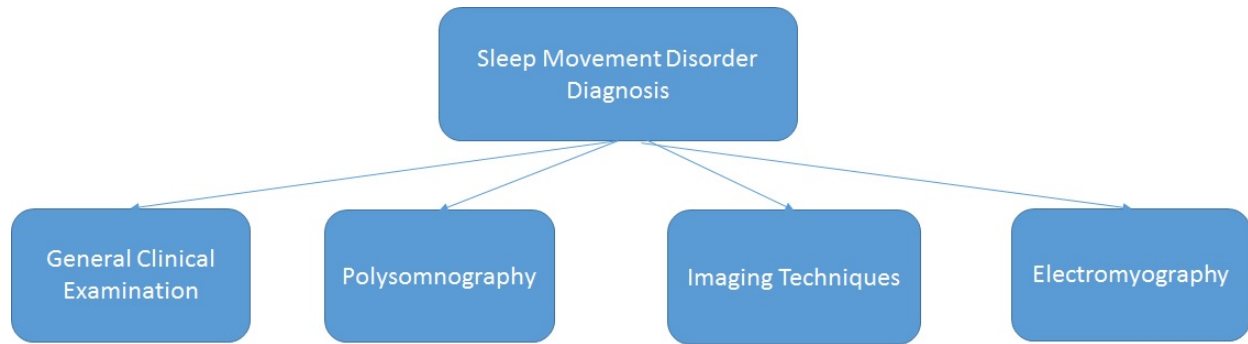
Figure 1.8: Non-invasive Diagnosis Methods for Sleep Movement Disorders

1.7.1   Polysomnography

PSG is commonly performed as a sleep recording procedure using standard recording and scoring methods [32]. During each study, monitoring of the following take place: surface EMG (anterior tibialis on the right and left legs [mostly used to detect PLM] and mentalis/submentalis), EEG (electrodes C3, C4, O1, O2), A1, A2 (reference leads at the mastoids), EOG, respiratory measures (abdominal and thoracic effort [measured with respiratory inductive plethysmography belts], nasal/oral pressure [measured with a nasal/oral pressure transducer], nasal/oral flow [measured with a thermistor]), oxygen saturation, and a 2-lead electrocardiogram. Sleep is manually staged according to criteria from the American Academy of Sleep Medicine [32] or using the traditional scoring of Rechetschaffen *et al.* [10]. All studies are interpreted by a sleep expert and scored by a registered polysomnographic technologist. Figure 1.9 shows the standard electrode placement sites used in PSG, in addition to some signals to its right. The numbers indicate locations that are part of a more complex and standardized system which is practiced around the world. In addition, the polysomnogram readings illustrate the onset of Stage for detection of eye movements

and tonic REM [10, 47]. In many clinical settings, in order to detect abnormal EEG beha-
viour in sleep, as many as 20 electrodes are used on the scalp. Polysomnographic modalities
can monitor sleep cycles to identify disrupted sleep patterns, adjust the treatments, in-
crease therapeutic options and enhance the quality of life of recording the EEG, EMG, and
Electrocardiogram (ECG).
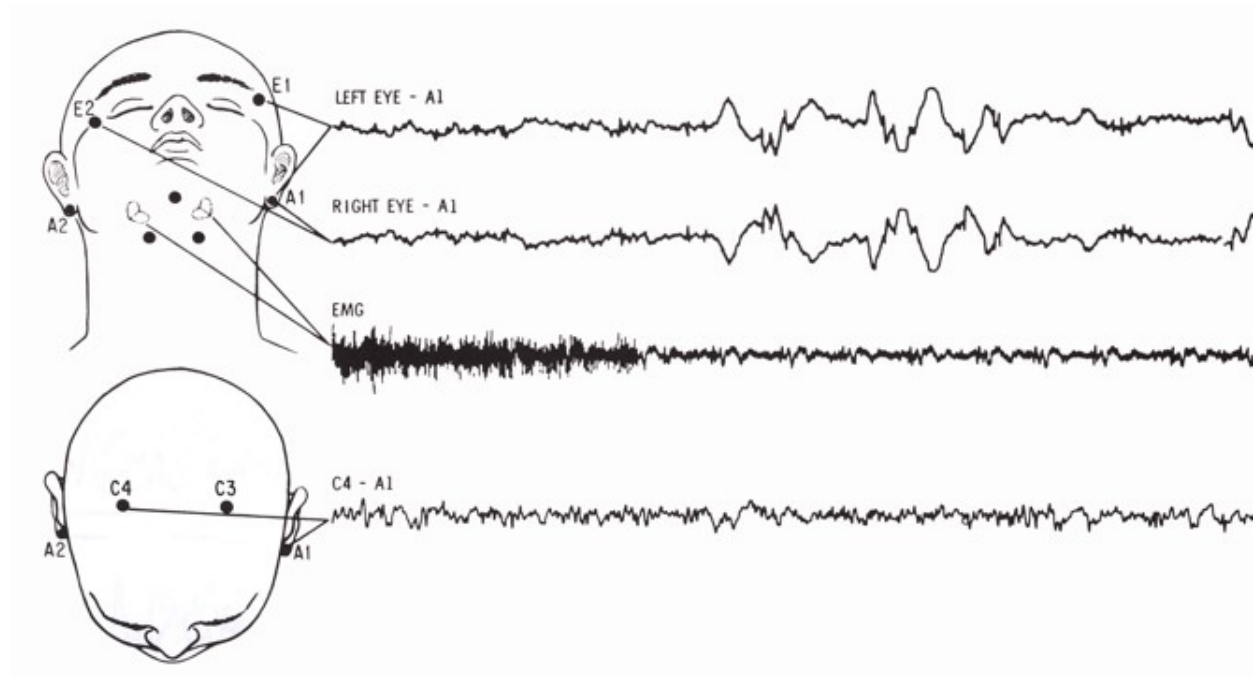
Figure 1.9: The standardized electrode placement sites used in polysomnography [48]

1.7.2 Neuroimaging

Imaging such as MRI, CT, and X-ray are non-invasive techniques that can be used to
investigate whether sleep disorders are associated with any changes in the brain structure.
Based on the current state of the research [31], it has been advocated that brain imaging is a

suitable approach to measure and evaluate the structural and functional correlations of sleep abnormalities as well as to serve in understanding the consequences of various therapeutic approaches. Therefore, by means of modern neuro-imaging techniques, a valuable tool is provided that allows us to gain insight into pathophysiological mechanisms of sleep disorders. However, MRI is expensive, inconvenient, and cannot be implemented over long periods of time. These techniques bring the necessity of using cheaper and more reliable and convenient techniques for both patients and medical experts. The next section explains a non-invasive technique that can be used for sleep movement disorder monitoring and diagnostics.

1.7.3 Electromyography

The problems with sleep signals and the limitations of the image-based techniques have motivated us to concentrate on the development of other non-invasive techniques. An interesting possibility for a safe, objective, and non-invasive clinical tool for early detection, localization, and classification of sleep movement disorders lies in the analysis of surface electromyography, also known as **EMG**. During sleep, an electrical current is produced by means of muscle activity bursts. Although there is little to no activity in the muscles in PSG, 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 having an amplitude of at least $50\mu V$, while tonic constitutes the remainder [17]. In one study it was shown that the degree of phasic EMG activity during sleep may reflect the completion of brain-stem development [49]. The type and frequency of muscle activity during sleep affect the information regarding the abnormality that causes

sleep problems.

So far in this chapter, we have discussed sleep physiology, sleep stages and different electrophysiological parameters of sleep. Furthermore, we have explained different sleep abnormalities and challenges in detection of such diseases. Also, we have highlighted that these sleep abnormalities, in particular sleep movement disorders, not only diminish the day time performance but also result in other neurodegenerative diseases such as Parkinson's and stroke.

There is sufficient indication that EMG signal analysis may be a potential tool for non-invasive screening of sleep movement disorders. However, these signals are one of the most challenging signal types, due to:

- Uncertainties in the signal and their generation process

- Varying spectral and temporal structures over time

- Non-stationary properties

- Multiple sources of vibration

- Multiple components

- Long duration.

Hence, to represent and understand these signals better, we must select a signal representation that:

- Reveals the varying structure

- Provides the best localization in a given signal

- Reflects a clear understanding about the signal production by correct signal illustrations.

These considerations inspire the research of this dissertation as follows:

The aim of the present dissertation is to develop an objective, non-invasive method for screening sleep movement disorders by applying novel and state-of-the-art signal processing and pattern recognition techniques to EMG signals.

1.8   Contribution of the Dissertation

The approaches taken in this dissertation toward development of a novel EMG signal analysis are presented in a block diagram shown in Figure 1.10 with the chapters with original contributions highlighted in a different colours. This work presents a generalized sparse representation methodology that exploits the benefits of sparsity, dictionary design algorithms, and pattern recognitions in sleep related problems. We will investigate four different implications as follows: i) long-term sleep signals; ii) non-stationarity iii)localizing to events of interest; iv) identification and classification of sleep signals. In all four implications, our main objective is to develop techniques that successfully quantify the patterns of interest in long-term EMG signals in sleep. The following section will highlight the original techniques and approaches investigated in this dissertation.

Novel and Unique long-term sleep movement EMG Database

Analysis of Sleep Disorders via EMG Signals
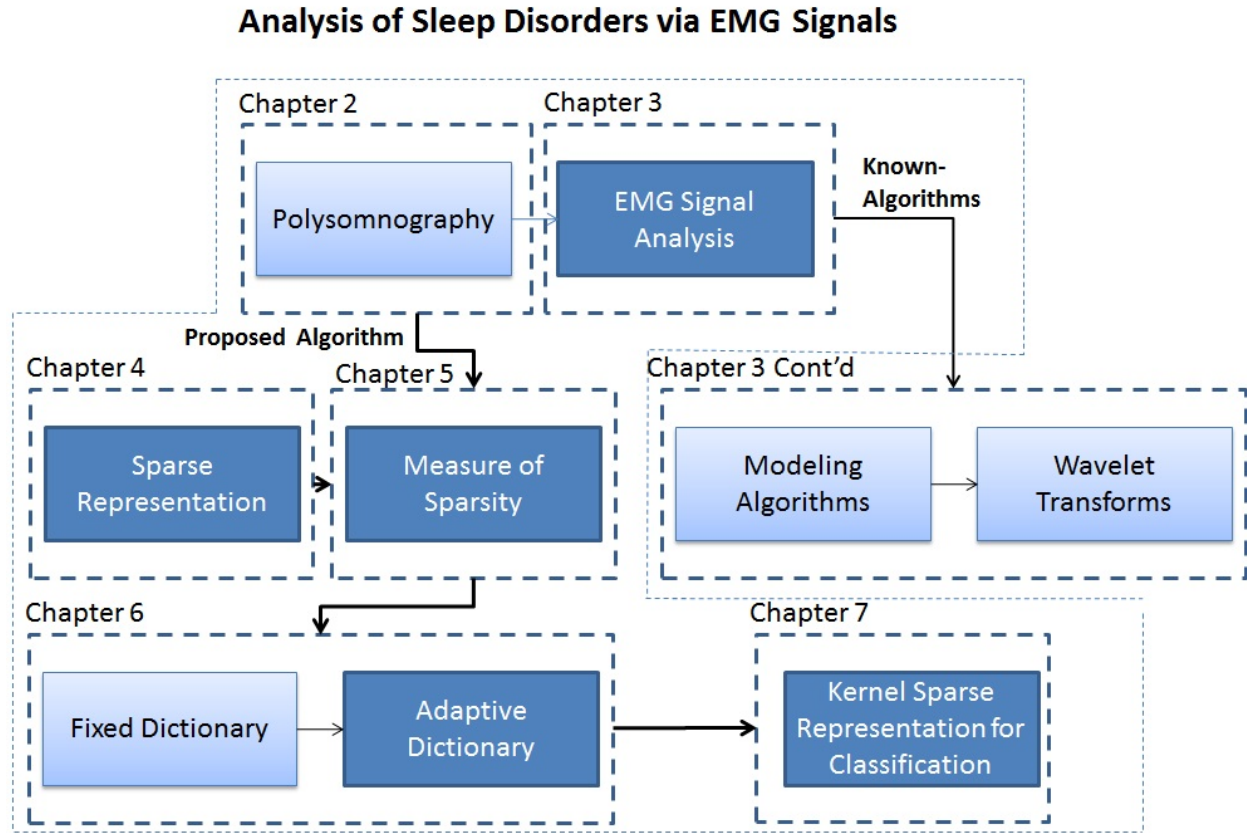
Figure 1.10: Flowchart of the proposed methodologies with highlighting the original contributions

- An overnight leg EMG recording to detect PLM

- An overnight chin EMG recording to detect RBD.

Sparse Representation for classification

- A robust generalization technique is provided for sparse representation of undersized signal types. Here, the signal decomposition is obtained using the ℓ^1-minimization technique, following which a generalization based on the LOO is performed.

- Selection of a small subset of functions using over-complete dictionaries for sparse signal expansion scheme.

- Using the measure of sparsity as a feature to perform classification rather than relying only on a single statistic for both validation and identification

- Evaluation of the representation and check to observe how accurate this representation is.

Dictionary Design Algorithms

- Using non-negative matrix factorization as well as the adaptive non-negative matrix factorization to design an optimum dictionary.

- Designing algorithms that can localize to the events of interest, especially for long-term monitoring and long-term signals.

- Designing an iterative part-based algorithm that only uses non-negative dictionary elements alternating between sparse coding of the examples based on current dictionary atoms to adapt to the data.

Kernel Sparse Representation for Classification

- Using Kernel functions to enhance the discriminatory power of the sparse representation for classification

- A novel automated method for the detection of PLMs in sleep is presented. This work has the potential to enhance detection of this clinically relevant nocturnal phenomenon in a cost-effective and efficient manner, and to facilitate future research in this evolving area of study.

1.9   Organization of the Dissertation

This dissertation is organized into 8 chapters. The flowchart in Figure 1.11 displays the evolution of this dissertation. The objectives at each stage are shown at the bottom of this chart. In Chapter 2, details of the databases, clinical data acquisition, and EMG signal acquisition are described. The different databases used in this dissertation and their characteristic details are tabulated. Chapter 3 presents a detailed review of the EMG signal analysis along with the history of signal processing, in particular on EMG signals. Characteristics of EMG signals will be outlined with emphasis on their complexities. An overview of the approaches taken is outlined and presented as well. Chapter 4, on the other hand, will present a classification approach based on sparse representation. Our proposed novel generalization algorithms based on the linear programming problem, using the LOO and Leave-M-Out (LMO) approaches, are explained. In this chapter, we will focus on quantification and detection of sleep abnormality. The performance measures obtained using long-term sleep data will be validated. This representation will be used to quantify EMG signal in both RBD and PLM disorder subjects. Chapter 5 exploits different sparsity measures and our proposed measure of sparsity.

In Chapter 6, we will present the development of an alternative dictionary design algorithm based on existing mathematical models and dictionary design algorithms. We will show that there is a link between sparsity in the dictionary and sparsity in the decomposition. Sparsity in the dictionary allows for faster and more efficient dictionary learning algorithms, whereas the sparsity in the decomposition provides a sparser representation and a decomposition

that successfully matches the signal features. To enhance performance and to achieve the best separability, in Chapter 7, the sparse approximation methods will be modified such that the objective function is enhanced with a discrimination term representing the separability properties of the signal. Chapter 8 presents a summary of the complete work with analysis of the achieved results at various stages. The novelty and the multifold benefits of the proposed work are highlighted. A discussion on the potential of the proposed methodology as long-term signal monitoring and non-stationary signal analysis tool will be presented. The future directions on enhancing this methodology are presented as well.

1.10   Chapter Summary
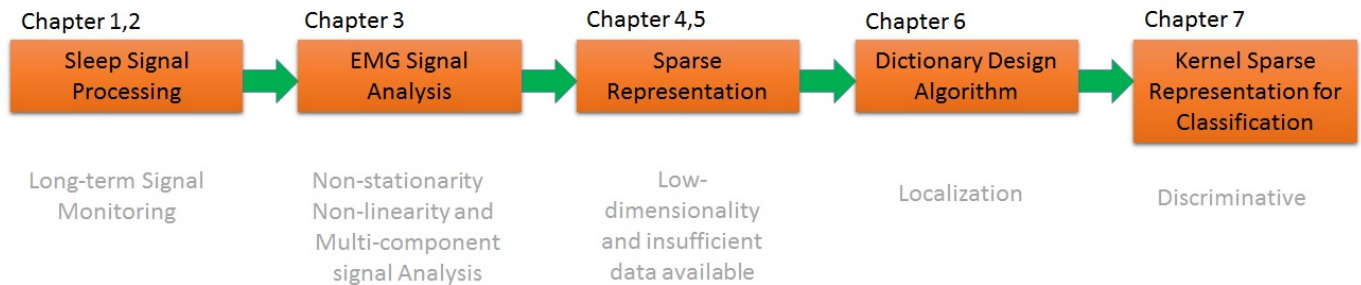
Figure 1.11: Flowchart of the proposed contributions

Sleep, sleep physiology, different sleep stages, and the electrophysiological parameters of sleep were described in this chapter. Different sleep abnormalities were explained with the focus of the sleep movement behaviour disorder. As well, some of the non-invasive methods used in sleep movement disorders were briefly reviewed. Lastly, an overview of the organization of dissertation was provided.

Chapter 2

POLYSOMNOGRAPHY SIGNAL RECORDING AND DATA ACQUISITION

Screening of Sleep Movement Disorders Via EMG Signal Analysis
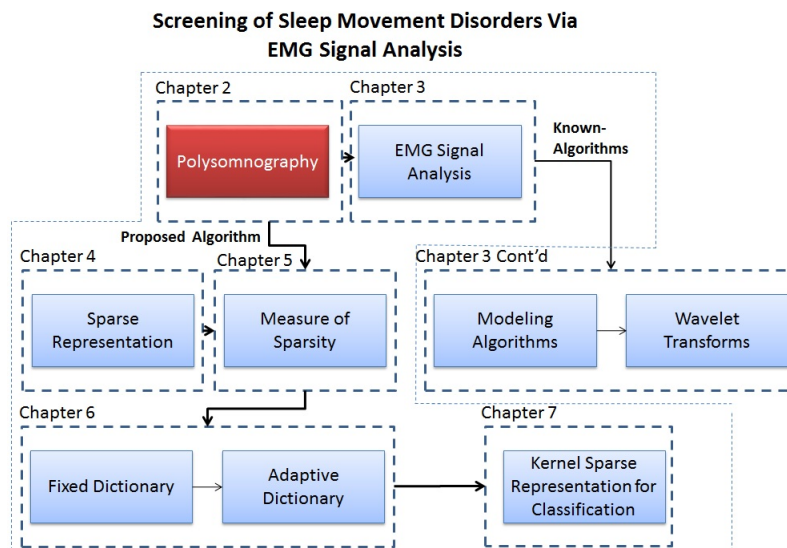
2.1   Introduction

S LEEP and dreams have played a critical role in the philosophy of human life and
have fascinated people for years. For a long time, sleep was thought to be a passive
state and for many years was believed to be associated with death. In ancient
Greek mythology, Nyx, the Goddess of Night, had twin sons called Hypnos (Sleep) and
Thanatos (Death) [50]. They were represented as young babies suckling on a breast of
mother night. The source of dreams, Morpheus, was the son of Hypnos and hence the
nephew of Death. This idea of sleep as a passive state runs through most of history, and
every night the mystery of sleep unfolds before us. However, it was not until recently that
the scientific study of sleep has been possible. The work published by a French scientist,
Henri Piéron, in 1913, is referred to as the forerunner of the modern approach to sleep
research [50, 51]. The development of polysomnography was the turning point that allowed
a better understanding of sleep and sleep disorders. This turning point mainly coincided with
the discovery of the electrical activity of the brain. At the end of the eighteenth century,
researchers using frogs as subjects found that nerve cells of animals produced electricity.
The father of electrophysiology, Emil DuBois-Reymond (1818-1896), demonstrated that the
nerves and muscle fibers had polarized states, and showed that the peripheral passage of
a nerve impulse was accomplished by an electrical discharge [50]. Around the same time,
researchers showed that nerve cells used their electrical capabilities for signaling information
to one another. Johannes (Hans) Berger (1873-1941), a psychiatrist from Jena, Germany,
was the first to record cortical electrical activity from the scalp in humans in 1924. He
noticed a significant difference between waking and sleeping alpha rhythms. During sleep,

the alpha rhythms disappeared, and electrical activity was of very low amplitude or sparse. This discovery established the most critical turning point in sleep research. As well, it allowed scientists to perform sleep screening without disturbing the sleeper and granted a window into the brain's activity with the chance of measuring it quantitatively. Shortly after, by means of a series of experiments, most major elements of sleep wave patterns, including the characteristic features that now comprise non-REM sleep, were described [52]. Researchers recorded overnight and daytime sleep and characterized it into five stages (A, B, C, D and E) listing the stages in order of appearance as well as by their resistance to change by external disturbances [53, 54]. By means of these experiments, sophisticated methods were developed to study sleep. EEG recording evolved, using amplifiers and high and low pass filters. Throughout their experiments, researchers found that certain areas of the brain produced better EEG waveforms. Consequently, certain channels of the EEG gained significant attention in the recording process of sleep. In addition to EEG, sleep researchers discovered that they were able to record other physiological parameters such as heart rate and respiration. One of the major advances in the field of sleep research was the discovery of REM sleep around 1953 when Kleitman asked one of his graduate students to observe people's eye movements as they slept [50]. However, watching for eye movements through the eyelids at night was a tedious task, so they came up with an easier method of using electrodes close to the eyelids, hence the emergence of the EOG. Together with EEG, EOG provided extended ability to evaluate the physiology of sleep. With this method, the term "rapid eye movement" (REM) was defined; this was different from the slow eye movement that happened at the onset of sleep in addition to heart rate increase and respiratory changes. Additionally, when awakened during rapid eye movements, subjects

were often able to recall dreams. At the time, however, the recordings were made only for short periods of time, to save time and paper, and to allow the observer to take naps between sampling episodes. Nevertheless, researchers did not have a clear understanding of when to turn on and turn off their recording equipments, so they made the all-night recordings with the (EEG and EOG) during sleep. This allowed them to describe and quantify the overall patterns of sleep throughout the night. Michel Valentin Marcel Jouvet did a series of experiments on cats regarding muscle atonia during REM sleep. He demonstrated that intact pontine tegmentum is the main reason for the onset of REM sleep and that REM atonia is due to an inhibition of motor centers in the medulla oblongata. He observed that cats that have lesions around the locus coeruleus have relaxed and less confined muscle movement during REM sleep. In addition, he showed diverse behaviours correlated with dreaming of attack, defense and exploration. He referred to these behaviours as an absence of muscle potentials during the REM periods in cats and this led to identification of an independent state of alertness or the "paradoxical sleep" [50, 55]. With his work, Jouvet demonstrated that recording EMG is of great importance in identifying REM sleep. With this addition, the basics of PSG, namely EEG, EOG and EMG of postural muscles were defined.

In 1967, a committee of investigators led by Allan Rechtschaffen and Anthony Kales, who both had extensive experience in scoring sleep, established a scoring system along with their own terminology [47]. They developed the first consensus-based guidelines that could universally be used by sleep specialists for staging and scoring sleep in normal human subjects, commonly referred to as R&K, or Rechtschaffen and Kales. They recommended that each sleep staging decision should be based on a 20 or 30-second window of the physiological sig-

nals, called an epoch. In their guidelines, they recommended the use of at least 4 channels including one channel for EEG, one channel for chin EMG and two channels of EOG signals, although later this number could be increased depending on the nature of the suspected disorder [10,50]. Eventually, due to the high number of respiratory and leg movement disorders during sleep, respiratory and leg movement sensors became a routine part of the all-night diagnostic test or the polysomnography. By extending of polysomnography with these additional sources of information, a new era in sleep medicine with the aim of studying sleep abnormality was evolved. This was and has been of great interest to the clinicians, as many of the sleep disorders not only diminish daytime performance, increase sleepiness, and affect the 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 published works devoted to the study and diagnosis of sleep disorders, and researchers have made genuine progress in recognizing sleep disorders in the general population and in clinical settings. Although numerous techniques have evolved to detect, prevent and treat many sleep disorders, sleep abnormality detection is still an active area of research. The focus of this research is on the detection of sleep disorders, in particular, sleep movement disorder that eventually leads to neurodegenerative diseases such as stroke and Parkinson's.

With the introduction of new technology and the advent of laboratory computers capable of signal processing techniques, it is possible to acquire, manipulate and store multiple physiologic data during sleep. Data acquisition, data display, data manipulation, data reduction and data filing are basic and distinct processes that can define the sleep signal analysis. In Section 2.2 information about clinical data acquisition, the sensors and EMG data acquisition are presented. In addition, the different databases used in this study will

be described. The chapter summary is presented in section 2.3.

2.2 Data Acquisition

The field of EMG research has enjoyed rapid increase in popularity in past years. The progressive understanding of the human body in sleep, a keenness for discovering the benefits of analyzing the physiological behaviours during sleep, the advancement of sensor technology, and the rapid growth in computational abilities of computers all contribute to the expansion of EMG-related research, particularly in sleep disordered subjects. The EMG signal used in this study were acquired during sleep by means of two different sets of electrodes, the chin EMG and the leg EMG.

2.2.1 Chin EMG

The recording of EMG from the muscle area on and beneath the chin is strongly recommended not only for scoring the REM Stage, but also for detecting sleep disorders such as RBD. Electrodes for these studies are attached to the participants in the sleep laboratory by trained technicians just prior to the respective participant's bedtime. Usual bed and rising times are ascertained for each participant via interview and minimally one week of confirmatory sleep log monitoring, and these "usual times" are used on the polysomnography (PSG) recording nights. All PSG records acquired are scored on computer reader stations by experienced sleep technologists using standard scoring criteria. The system used for recording chin EMG signals during sleep is manufactured by "Compumedics." The "E-series"

model 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 $10-100$ Hz. The impedance of each electrode is less than $10K\Omega$ with a minimum digital resolution of 12 bits per sample. The positioning of the chin EMG electrodes is demonstrated in Figure 2.1. The sampling rate is 256 Hz. Similar electrodes are used to record EEG and EOG amongst other physiological parameters. Data collection from humans was facilitated through a protocol approved by the local research ethics board.

2.2.2   Leg EMG

All polysomnograms were recorded on digital equipment (Compumedics Neuroscan, Australia) using standard recording and scoring methods as previously described [32]. During each study, monitoring of the following took place: surface EMG (anterior tibialis on the right and left legs [used to detect PLMs] and mentalis/submentalis), EOG (electrodes C3, C4, O1, O2), A1, A2 (reference leads at the mastoids), EOG, respiratory measures (abdominal and thoracic effort [measured with respiratory inductive plethysmography belts], nasal/oral pressure [measured with a nasal/oral pressure transducer], nasal/oral flow [measured with a thermistor]), oxygen saturation, and a 2-lead electrocardiogram. All studies were videotaped and audio-visual recordings were time-synchronized to the remainder of the data [56]. All studies were interpreted by a diplomat of the American Board of Sleep Medicine and scored by a registered polysomnographic technologist.
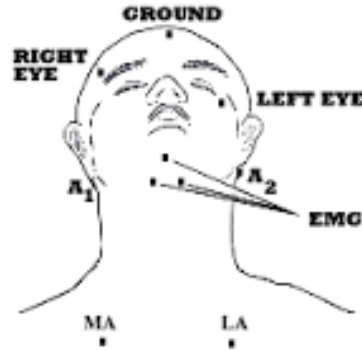
Figure 2.1: The positioning of the three electrodes of EMG signal to record the electrical activity of the chin [57]

2.2.2.1   Neuroimaging:

MRI and CT were acquired at Sunnybrook Health Sciences Centre; if CT and MRI were performed on the same patient, only MRI data was used in the analyses. 1.5 T MRI (General Electric Medical Systems) included T2 Fluid Attenuated Inversion Recovery (FLAIR) and Diffusion Weighted Image (DWI) sequences. In order to compute stroke volumes, infarcted tissues were visually identified and planimetrically traced by a trained operator based on intensity changes compared to contralateral tissue. All stroke localizations were confirmed by a formal neuroradiologist report.

2.2.2.2   Quantification of White Matter Hyperinteisities:

Global White Matter Hyperintensities (WMH) appearing hyperintense on FLAIR and hypodense on CT were assessed with the Age Related White Matter Change (ARWMC) Scale [58].

2.2.3   Database (A)

In this work, a subject was defined as normal if there was no history of violent behaviour during the night sleep; otherwise he or she was considered abnormal (diagnosed by medical sleep professionals). Thirty-six male/female volunteers with normal behaviour and subjects with elevated muscle tone were scheduled to undergo the sleep test independent of this work. All these subjects filled out a questionnaire, in which they included their informed consent, prior to their sleep log monitoring. The abnormal subjects appeared to act out their dreams, in which the exhibited behaviours mirrored the content of the dreams. Their typical behaviours included talking, laughing, yelling, gesturing, punching and kicking. Each subject slept at the hospital's clinic for one night and signals such as EEG, EMG and EOG were recorded from these subjects.

2.2.3.1   Study Population:

The present analysis included patients who were referred to hospital for sleep disorders such as insomnia, fragmented sleep disorders, sleep apnea and elevated muscle tone. All patients provided consent for retrospective analysis of their data. The patient characteristics are shown in Table 2.1.

Table 2.1: Characteristics of Study Patients Group A (n = 36)

Patient Characteristics	
Age in years (mean±SD)	48.07± 19.42
Male gender - %	33.4
Body mass Index	28.4±2.3
Height	5'3"±4.5"

Most of the subjects were taking medications such as: Hydrochlorothiazide, Lorazepam, Venlafaxine XR, Apo-Levocarb , Micardis, Effexor XR, Aggrenox, Norvasc, Hydrazide, Lipitor, or Aricept medication, or others.

2.2.4 Database (B)

The following clinical data was obtained: (a) Past medical history of myocardial infarction, coronary artery disease, hypertension, hyperlipidemia, diabetes, and prior and/or current smoking; (b) medications; (c) blood work, if obtained for clinical purposes. Sixty five patients were included in the present analysis (mean age 63.5 years, 52% male). Twenty-two patients (34%) presented with stroke, 19 (29%) with Transient Ischemic Attack (TIA), and 24 (37%) with other diagnoses.

2.2.4.1 Study Population:

The present analysis included patients who were recruited from a cerebrovascular disease research program at a tertiary stroke centre and also underwent both magnetic resonance imaging of the brain and polysomnography. Those who could not read or write, as well as non-English-speaking or aphasic patients, were excluded. All patients provided consent for retrospective analysis of their data. The patient characteristics are shown in Table 2.2.

Table 2.2: Characteristics of Study Patients (n = 65)

Patient Characteristics	
Age in years (mean±SD)	63.5 ± 13.3
Male gender - %	52
Days from Recruitment to Imaging (mean±SD)	6.7±90.1
Days from Recruitment to Sleep Study (mean±SD)	58.8±86.6
New Stroke Volume (mm3) (mean±SD)	2294.4±7868.3
Old Stroke Volume (mm3) (mean±SD)	1796.8±7825.7
Prior stroke (%) (as noted on imaging)	22
Prior Myocardial Infarction (%)	12
Coronary Artery Disease (%)	17
Body mass Index	29.4±5.1
Hypertension (%)	57
Hyperlipidemia (%)	60
Diabetes (%)	18
Current Smoker (%)	9
Renal failure (%) (creatinine $> 106 \mu mol/L$)	13
Low ferritin (%) ($< 45 \mu g/L$) ($n = 20$)	18
SSRI medication use (%)	15
Sleep Efficiency (mean±SD)	71.2±16.3
Apnea-Hypopnea Index (mean±SD)	12.3±20.2
Lowest Oxygen Saturation (mean±SD)	86.7±5.1
Arousal Index (mean±SD)	20.8±18.7
PLM index (mean±SD)	17.3±25.2

2.2.5   Experimental Procedure

2.2.5.1   Ethics:

Institutional ethics was obtained prior to the initiation of the study. Written, informed consent was provided by all study participants. This study population was derived from two trials that were registered on clinicaltrials.gov:

1. Post-stroke Triage "DOC": Simple Screening for Depression, Obstructive Sleep Apnea and Cognitive Impairment NCT02007265

2. Sleep Disorders Managed and Assessed Rapidly in Transient Ischemic Attack (TIA) and In Early Stroke (SMARTIES) NCT01528462

2.3   Chapter Summary
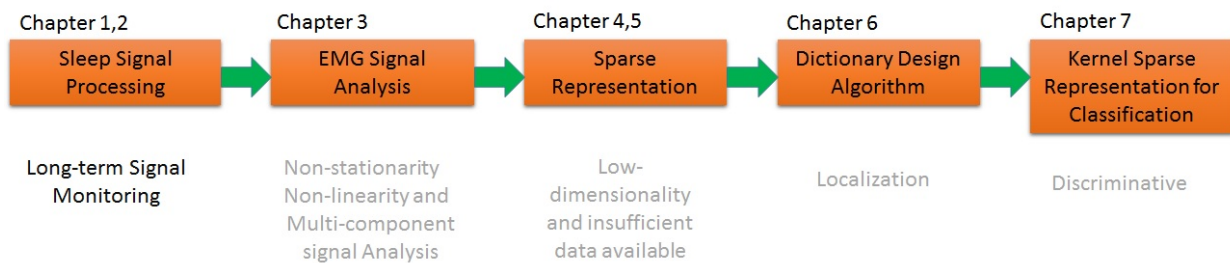
Figure 2.2: Flowchart of the proposed contributions

In this chapter, details of the databases, clinical data acquisition, and EMG signal acquisition were described. The different databases used in the dissertation and their statistical details

were tabulated. Figure 2.2 displays the contribution flowchart, and the highlighted block in this figure shows the progress of the work in this chapter. Chapter 3 focuses on EMG signal processing and will highlights the complexity of such signals.

Chapter 3

EMG Signal Analysis

Screening of Sleep Movement Disorders Via EMG Signal Analysis

Chapter 2 — Polysomnography

Chapter 3 — EMG Signal Analysis

Known-Algorithms

Proposed Algorithm

Chapter 4 — Sparse Representation

Chapter 5 — Measure of Sparsity

Chapter 3 Cont'd — Modeling Algorithms → Wavelet Transforms

Chapter 6 — Fixed Dictionary → Adaptive Dictionary

Chapter 7 — Kernel Sparse Representation for Classification

3.1   Motivation

T HE advancements in sensor technology have made it possible to gather huge amounts of data, extending the applicability of signal analysis to a wide variety of fields, such as communication, security, biomedicine, biology, physics, finance, and geology. This creates a demand for advanced signal analysis techniques to effectively process the gathered data. If a reasonable understanding is obtained, it then becomes possible to observe the corresponding signals and assess the state of the system for detection, measurement, and classification applications. Although there has been substantial progress in both data collection and processor technologies, biomedical signal analysis techniques are still performed by humans. This can make the data processing; unreliable, subjective and sometimes inefficient [59]. For example, sleep signals are recorded, for about 8 hours over the night. A technician must score the various stages of sleep based on the properties of different signals recorded, and develop a hypnogram. This procedure is often frustrating and time consuming due to the length of the signal. Automation of sleep (i.e., sleep staging) could be more reliable, cost effective as well as more efficient.

3.2   Signals

Signals are functions of one or more independent variables, and typically contain information about the behaviour or nature of some phenomenon. In other words, a signal can be considered a unique communication that occurs between the physical environment and a human being. This communication can be either voluntary or stimulated. The former indic-

ates the cases where the signal is measured directly from a physical quantity. An example is the measurement of temperature signals, or the measurement of amplitudes of an amplifier. Biomedical signals, on the other hand, are types of signals that are strongly related to the human body or human organisms and are used to stimulate communication with the physical world; for instance, recording of EEG signals of a patient, or ultrasound waves which are beamed into the body causing echoes that produce images of the inner body. It is very well known that abundant kinds of signals exist in the real physical world. All these signals carry a lot of information that is of interest, and diverse schemes have been developed to analyze, interpret, manipulate, and process them.

3.3   Signal Analysis

Signal processing algorithms are needed to analyze and understand the signals. These techniques can exploit the intrinsic properties of the signal, which allows a specific function to be applied on the signal. An analysis tool which can be tuned to characteristics of the signal is selected in such a way to extract the features of interest [60] [61]. Thus, the primary goal of signal analysis is to extract useful information to understand the signal generation process, or to extract features for signal classification purposes. Most of the methods in this area are treated under the disciplines of spectral estimation and signal modelling [62].

3.3.1   Signal Behaviour

The theoretical analysis of a signal requires the availability of a mathematical description for the signal itself. Such a description is usually referred to as a signal model; this term defines the characteristics of the most appropriate approach for analysis of such signals [63] [64]. There are many diverse types of signals, ranging from simple structures to very complex ones. The most important classifications of signal models are either deterministic or random.

3.3.1.1   Deterministic Signals

Deterministic signals are any signals which can be described by an explicit mathematical relationship. The sinusoidal signal is the simplest example of a deterministic signal, since the value of the signal at each time sample can easily be calculated. In other words, there will be no unpredictability in the signal value at any given time sample. This characteristic of the signals allows for advanced prediction of signal quantities. In the case of a continuous time signal, the mathematical relationship is a given function of time [65]. Fig 3.1 illustrates a sinusoidal signal with a sampling frequency of 8 kHz and an amplitude of one.

3.3.1.2   Nondeterministic or Random Signal

In contrast to deterministic signals, nondeterministic signals are defined as 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 an unpredictable manner. These signals are often called random; for instance, the signals

47
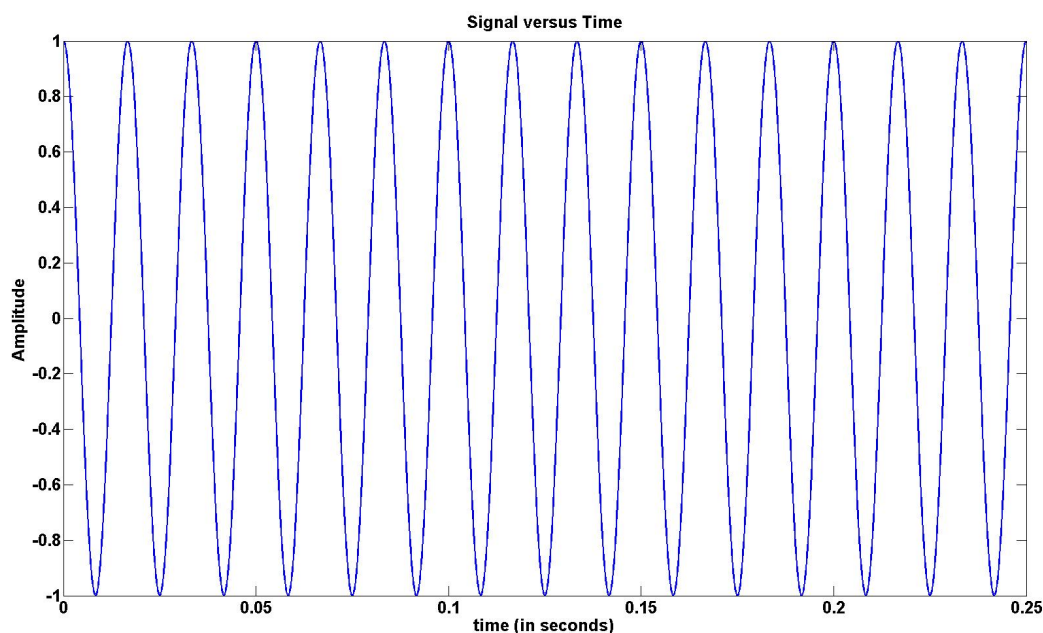
Figure 3.1: A deterministic sinusoidal signal with a frequency of 8kHz and an amplitude of one.

constructed from recording an integer of a rolling die.

Yet, although random signals evolve in an unpredictable behaviour over time, their average characteristics can often be assumed to be deterministic and can be specified by explicit mathematical formulas. At this point one could articulate, having complete knowledge of the physics of the signal, an explicit mathematical relationship at least within the limits of the uncertainty principle. This concept is a key to the modelling of a random signal such as a stochastic process. Random signals are further divided into two groups:

Stationary : A stationary signal is a signal whose 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. For

48

example, each face of the die is equally likely to occur (the probability of occurrence of each side at any time is $\frac{1}{6}$). Therefore, these types of non-deterministic signals that can be expressed using their probabilistic or statistical values are called stationary signals.

Non-stationary : A non-stationary signal, on the other hand, has a time-varying probability distribution. This causes other properties that depend on a probability distribution function (PDF) to be time-varying as well; for instance, the global temperature signals over some period of time. It is impossible to find a mathematical function, or to define an accurate fixed probabilistic value for each of the temperature values.

A majority of real-world signals generated by nature (i.e., temperature or biological signals) are non-deterministic and non-stationary. Due to the uncertainties in these signals and their generation process, the spectral and temporal structures of real-world signals are varying over time. This is the reason that non-stationary signals are the most challenging signal type. Hence, in order to appreciate and recognize the signals, better selection of a signal representation that reveals the varying structure in a given signal is of great importance. This allows a clear understanding about the signal production and behaviour. Fig 3.2 demonstrates the non-stationary signal of the annually-averaged land temperature record. Fig 3.3 on the other hand shows a biomedical (EMG) signal plotted over $4s$ with its mean computed in a moving window of $40ms$.
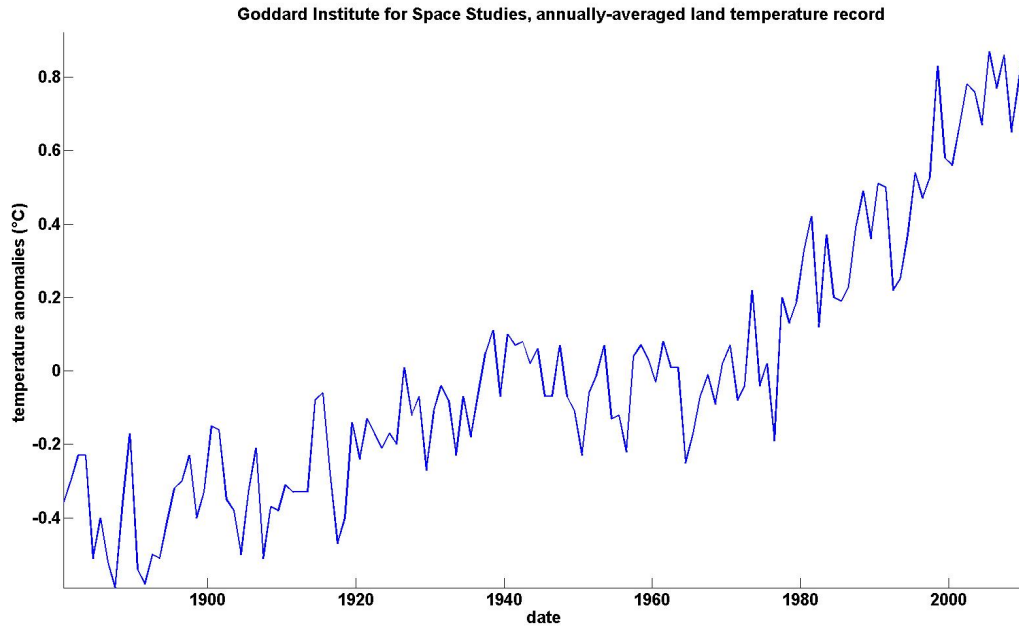
Figure 3.2: A non-stationary signal demonstrating the annually-averaged land temperature record for the years 1980 - 2010.
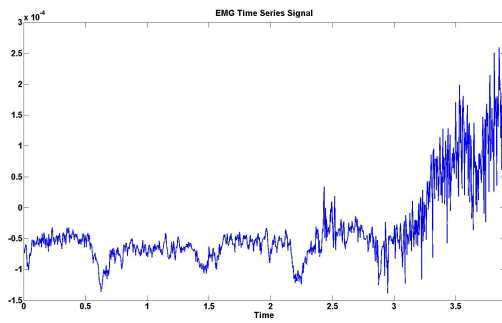
(a) EMG signal
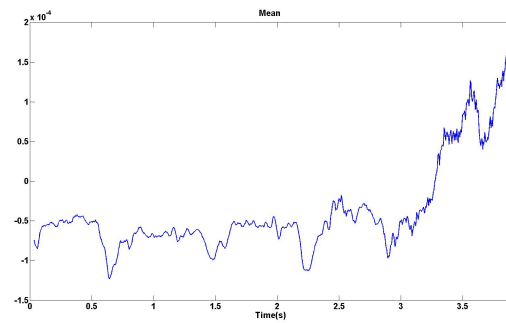
(b) Mean

Figure 3.3: (a) A 4s EMG signal is shown as an example of a real world non-stationary biomedical signal. (b) Mean of the signal is computed over moving windows of 40 ms long.

3.4   Biomedical Signal Analysis

Physiological processes are complex phenomena with inputs and outputs that can take the form either of physical material or of information derived from mechanical, electrical or biochemical actions. Most of these processes are accompanied by, or manifest themselves as signals that reflect their nature. Diseases or malfunctions in a biological system cause alterations in the system's corresponding biological processes and related signals.

Physicians, neuroscientists and health care technologists are highly skilled professionals; however, owing to the huge amount of arithmetic operations involved with analysis of signals, computer-aided diagnosis is now being employed for all available visual patterns. This also assists in efficient long term monitoring of biomedical signals, such as ECG, EEG and EMG. The properties of these signals motivate clinical and basic science research in non-invasive diagnostic and treatment modalities, and in the development of new non-invasive tools aimed towards personalized treatment of patients. For instance, EEG, EOG and EMG can be used to detect many sleep disorders, and in fact, these signals are recorded in sleep laboratories from subjects suffering from a sleep disorder, and are further analyzed by sleep experts. The properties of these signals help the sleep technicians score the various stages of sleep so they can 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. Fig 3.4 is a hypnogram illustrating sleep stages and cycles in an adult subject.
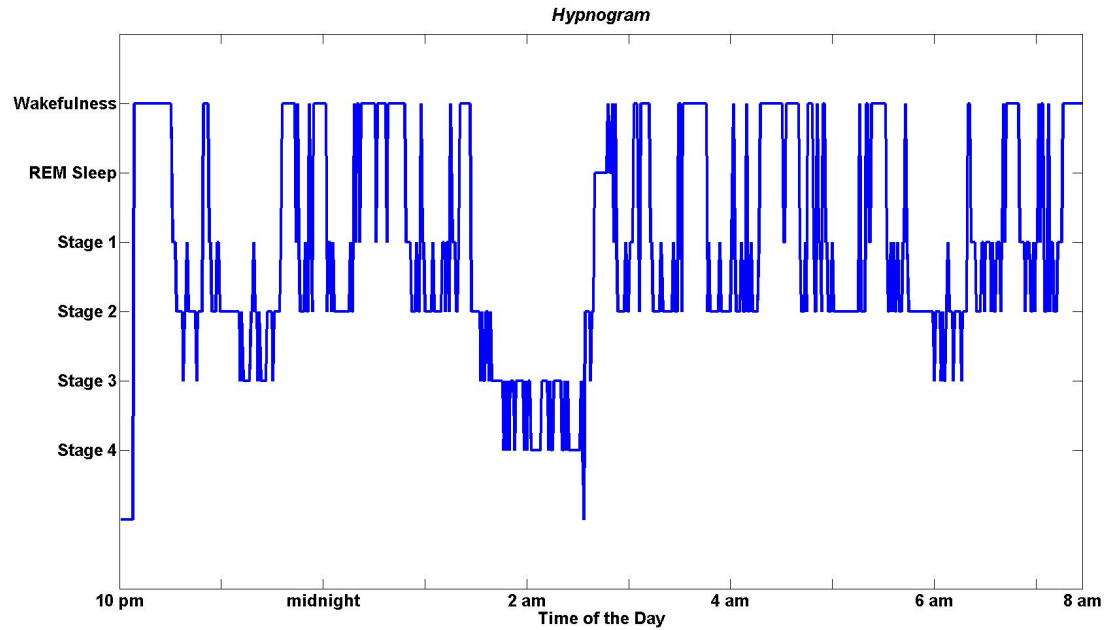
Figure 3.4: A hypnogram signal illustrating sleep stages and cycles in an adult subject.

Analysis of a physiological signal is usually multifold, mainly because the associated characteristic patterns in the signal are too complex to allow their equivalent biological relevance to be derived. In addition, the sensed biomedical signal is usually contaminated with artifacts that can very well be the most prominent signal component. In some cases, the noise present in the signal is of substantial amplitude, and it also tends to be random in nature. The next section will demonstrate the signal processing techniques developed so far.

52

3.5   A History of Signal Processing

3.5.1   Linear Transforms

In the early 1960s researches gave significant attention to the Fourier transform to diagonalize linear time invariant operators. This soon became the main tool for analyzing and designing such operators. In general, the Fourier basis approximates a signal by projecting it onto its frequency contents. Nevertheless, the Fourier transform produces oversmooth results in many of the practical cases. The Fourier transform assumes periodic extension even for finite signals, which could result in discontinuity at the boundary. In 1965, when Cooley and Tukey introduced the Fast Fourier Transform (FFT), the tool became even more popular [66]. The Discrete Cosine Transform (DCT) on the other hand, produces continuous boundaries; since it assumes an anti-symmetric extension of the signal, it is a more efficient transform [67]. Soon after, researchers used Fourier bases as a linear signal approximation tool.

$$\mathbf{x} \approx \sum_{n}(\psi_n^T \mathbf{x})\Phi_n \tag{3.1}$$

where Φ_n is the Fourier basis, $\mathbf{x}$ is the signal which is linearly approximated by projecting it onto basis elements, and ψ_n is the bi-orthogonal basis. This in general can achieve compression, which can later be replaced by sparsity.

Representing the signal with fewer coefficients was a major driving force for the development of more efficient representations. In the late 1960s Akiake used Autoregressive (AR) models to estimate the spectral characteristics of time series [68]. This soon became the common

approach for researchers to identify noisy feedback systems, a problem that could not be solved by direct application of frequency approaches [68]. This approach was even used for feature extraction techniques in the late 1980s and early 1990s. Tavathia *et al.* considered the AR coefficient's first dominant pole and the spectral power ratio as the discriminatory features to classify the vibroarthography (VAG) signals [69]. Later Moussavi *et al.* used a 40th order forward-backward linear prediction (FBLP) model on the same VAG signal with the top 10 dominant poles and some clinical information to construct the feature vector for classification. In 1997 Rangayyan *et al.* used a 40th AR model to find the parametric representation of the signal and used the first 6 coefficients for classification of the aforementioned signal. However, in all these works, AR modelling could not be applied to the biomedical signal itself, as the biomedical signals were naturally non-stationary. This led researchers to develop algorithms that could track the stationarity of signals and adaptively segment them into stationary components. Tavathia *et al.* used an adaptive segmentation method based on the spectral error measure (SEM) procedure [69]. They used a constant width reference window starting at the first data sample. Their algorithm however, did not investigate sudden changes in the signal, since a fixed window length was used for analysis. Krishnan *et al.* investigated a new adaptive segmentation based on the Recursive Least Square Lattice (RLSL) algorithm [70]. Their algorithm was more successful when compared to the previous methods due to faster convergence. Moussavi *et al.* on the other hand analyzed VAG signals using Recursive Least Square (RLS) algorithms [71]. Shokrollahi *et al.* applied the RLS algorithm to EMG signals. Their algorithm marked the stationary segment boundaries and was carried out in two steps: the first step was the primary boundary detection algorithm in which the beginning of the stationary boundary was marked, and the

second step was the decision process on the location of the final segment boundary in which the end of the stationary boundary was marked. Since having a segment with only a few numbers of samples may result in under modelling, a minimum of 120 samples was used as the minimum segment length L_{min} [72].

The following is the procedure for segmenting the signal [72].

- Define a threshold value, which is three times the standard deviation of the $s(n)$ vector.

- Compare each element of the $s(n)$ vector with the threshold and store all points above the threshold as the Primary Segment Boundary (PSB), $PSB = [a_0, a_1, ..., a_P]$.

- Compare the adjacent elements of PSB to one another, 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.

The task of segmentation was to divide the non-stationary signal into a few locally stationary components over a limited time span. At this point, conventional temporal techniques could be applied for analysis of the signal segments. Many applications, specifically speech signal processing, uses fixed segmentation techniques whereas adaptive segmentations are well-suited for the analysis of non-stationary biomedical signals. Applying fixed segmentation to biomedical signals may result not only in redundancy in the segments of interest, but also in poor identification of sudden variations.

With the advancement of technological devices, collection of large and complicated datasets has become an easy task. This process demands advanced signal processing techniques to effectively analyze the recorded data. At first, simple time-series or frequency algorithms were not well equipped to adequately analyze more complex natural data, and new, improved techniques were sought [67]. Researchers were interested in flexible representations based on the local signal characteristics. One of the first structures to be used was the Short Time Fourier Transform (STFT), which emerged as the natural extension of the Fourier transform. In this algorithm, the Fourier transform is applied locally to the segments of the signal revealing a time-frequency description. The work by Griffin *et al.* is the extension of this analysis [73]. The drawback with STFT, however was the fixed segmentation of the window length. Although the aforementioned algorithms provide discriminant information for the analysis of the signals, they suffer from a common shortcoming; in general, segmentation-based algorithms depend on clinical information that may not be available all the time.

Since biomedical signals have time-varying statistical properties, joint time-frequency (TF) analysis techniques can be expected to overcome the drawbacks associated with segmentation-based approaches [74]. During the beginning the 1980s, researchers were excited about the development of a very powerful tool known as wavelet analysis [75] [76]. Morlet [75] proposed a series of dilated and translated waveforms that pass over the elementary functions, taking the form of:

$$W = \left\{ \phi_{n,m}(x) = \alpha^{n/2} f(\alpha^n x - \beta m) \right\}_{n,m \epsilon \mathbb{Z}} \tag{3.2}$$

Wavelet transform revolutionized the signal processing community by offering a complete mathematical framework relating to multi-resolution expansion [77] [78] [79]. In general,

the multi-scale wavelet basis is constructed from a pair of localized functions, referred to as the scaling function and mother wavelet. Scaling functions are assumed to be low frequency components that span the approximation of the signals. The mother wavelet is more of a high frequency signal that varies in scale and translation and spans the signal details. In general, wavelet transforms are functions that satisfy certain mathematical requirements and can be used to decompose the signal using basis functions. The operation of translation and scaling seems to be basic for many practical signals and signal generating processes [80], [81], [82], and their use is one of the reasons that wavelets are efficient expansion functions. The continuous wavelet transform (CWT) for a function *f(t)* is defined as:

$$C(f)_{a,b} = a^{-1/2} \int_{-\infty}^{\infty} f(t)\psi(\frac{t-b}{a})dt \tag{3.3}$$

where scale parameter $a > 0$, and translation parameter b are real values. The CWT maps a one-dimensional signal to a two-dimensional time-scale joint representation. It is calculated by continuously shifting a continuously scalable function over a signal and calculating the correlation between the signal and wavelet function at given time and scale [83]. The resulting wavelet coefficients are highly redundant. This redundancy significantly improves the shift invariance property of WT, which makes CWT superior to the discrete wavelet transform (DWT) in signal classification problems due to more stable extracted features (wavelet coefficients). The DWT is computed only on a dyadic grid of points: 2^j, $j = 1, 2, \ldots$, and is defined as

$$D(f)_{j,k} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t). \tag{3.4}$$

The wavelet family $\psi_{j,k}(t)$ is generated by shrinking by a factor 2^{-j} and translating by $2^j k$ from the mother wavelet ψ, that is, $\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)$. Compared to the Fourier transform, the advantage of WT (either CWT or DWT), is that it uses long duration windows for capturing low frequency components and short duration windows for capturing high frequency components. This is based on the assumption of rapid changes in high frequency components and slow changes in low frequency ones, meaning that at high frequencies the WT is sharper in time, while at low frequencies the WT is sharper in frequency.

It has been realized that it is too difficult to achieve the local stationary component for biomedical signals, as the nonstationarity is often shown in a relatively short period of time. Also, when the length of the signal is too small, the statistical measure estimates of the signals often suffer from a small number of sample points. In addition, wavelet transform leads to a sparse representation of the signal. By selecting only a few large values of wavelet coefficients of the signal as an input of the feature vector, one can significantly reduce the feature dimension in the classification step and ultimately improve the computation efficiency and classification accuracy. Within the wavelet transform framework, one can develop (at least in principle) optimum signal performance of practical statistical signal processing techniques [65]. These processing methods start a new era of signal processing and feature extraction techniques using wavelet-transforms. Chang *et al.* uses the tree-structured wavelet transform for texture analysis; they showed that compared to Gabor filters and conventional pyramid-structured wavelet transforms, the tree-structured wavelet transform provides a more natural and effective tool for texture analysis [84]. In another similar article, Arivazhagan *et al.* used DWT and a similar set of wavelet statistical and co-occurrence matrix features to classify 3 sets of texture image datasets. Their results

showed that using wavelet transform as the feature extraction technique could increase performance up to 97.8% [85]. Lumbrou *et al.* on the other hand used different wavelet analysis techniques on musical signal for classification. In their paper they showed that the use of wavelet transform as features produces higher classification performance compared to entropy, skewness and correlation [86]. Shokrollahi *et al.* analyzed EMG signals in REM sleep to detect abnormal behaviour that can lead to Parkinson's disease. The authors used both CWT and DWT for their analysis and proved that their approach results in better performance compared to previous methods, such as AR modelling and cepstrum analysis [87].

Since 1990, sparsity has increased in popularity, especially due to the limits of approximation in orthogonal bases [67]. Sparsity has been used to describe the increasingly complex behaviour of signals by adapting the transform atoms to the signal content. One of the main optimal representations, called wavelet packet transform, was constructed in 1992 by Coifman, Meyer and Wickershauser [88], and it adaptively adjusts to signal properties and behaviours. There was a drawback to this transform: for higher dimensional signals, it remained a separable and non-oriented transform. Wavelet was soon not an ideal tool for analysis and representation because of its shiftability properties. In general, wavelets are sensitive to translation and rotation in higher dimensions. To overcome this, researchers proposed to use over-complete dictionaries since the number of atoms in an orthogonal transform has to be simply sufficient to overcome the translation and rotation invariance problem. This is the reason that researchers became more and more interested in developing over-complete transforms. Researchers were interested in approximating the signals with even fewer coefficients, called sparse representation. In 1993, Mallat and Zhang proposed

a novel sparse signal expansion scheme that was based on selection of a small subset of functions using over-complete dictionaries [89]. Using this methodology, the signal could be represented in more than one approximation, and depending on the task, the best approximation could be selected. Since then, sparse representation has been an active area of research for a variety of applications. This scheme provides a compressed representation for signals using a combination of atoms chosen from an over-complete dictionary. However, there exist few over-complete dictionaries that suit all data types. This requires the representation to span multiple dictionaries, resulting in residual errors during signal approximation. Recent research in sparse approximation involves extension of these techniques towards obtaining optimal signal representation, based on the data itself, which leads to optimal signal discriminations.

3.6   EMG Signal Analysis

The diagnostic potential of EMG signals for the non-invasive detection of movement disorders began in the 1960s. However due to limitations in recording devices, it was not until the middle of the 1980s that EMG was used as a clinical method for diagnosis of neuromuscular problems. In the early 1980 Cram *et al.* introduced a clinical method for scanning a variety of muscles using EMG sensors [90]. EMG then became an important tool in diagnosing sleep movement behaviour disorders. In 1993, Lopes *et al.* analyzed total airway resistance and respiratory muscle activity during sleep [91]. In another study Phd *et al.* analyzed the relation of periodic movements in sleep (nocturnal myoclonus) and sleep disorders [92]. In 1992 Schenck *et al.* processed EMG signals to analyze motor dyscontrol in

narcolepsy, REM sleep without atonia, and REM sleep behaviour disorder [93]. After the investigation conducted by Tan *et al.* on RBD preceding Parkinson's disease with therapeutic response to medication, EMG analysis became even more popular. The use of EMG analysis on sleep signals was continued. By 2005, Lee *et al.* used EMG to analyze Parasomnias and other sleep-related movement disorders [94]. In 2009, Shokrollahi *et al.* published work on analysis of the EMG of rapid eye movement sleep using wavelet techniques [87]. In another study, they examined the same signals using autoregressive and cepstrum analysis to detect the abnormal behaviour of subjects suffering from RBD [72]. All these works demonstrate that the EMG signal is viable for developing a non-invasive diagnostic tool.

The focus of this thesis is to analyze different EMG signals in sleep, measuring two main sleep movement disorders that lead to neurological disorders: RBD and PLM. The next section will explain the challenges involved in analyzing these signals.

3.7   Challenges of EMG Signal Analysis

EMG signal characteristics pose severe challenges in terms of analysis and extraction of discriminant features. The points below summarize a few main characteristics of the EMG signals.

- EMG signals are non-stationary, because variations of the signals are different from subject to subject. Also, owing to the uncertainties in these signals and their generation process, their spectral and temporal structures vary over time; that is, their first and second order statistics change with time. Figures 3.5 and 3.6 shows the two EMG

signals for RBD and those with elevated PLM. As can be seen in these two figures, the signals look very random and there is no specific pattern that can be detected.
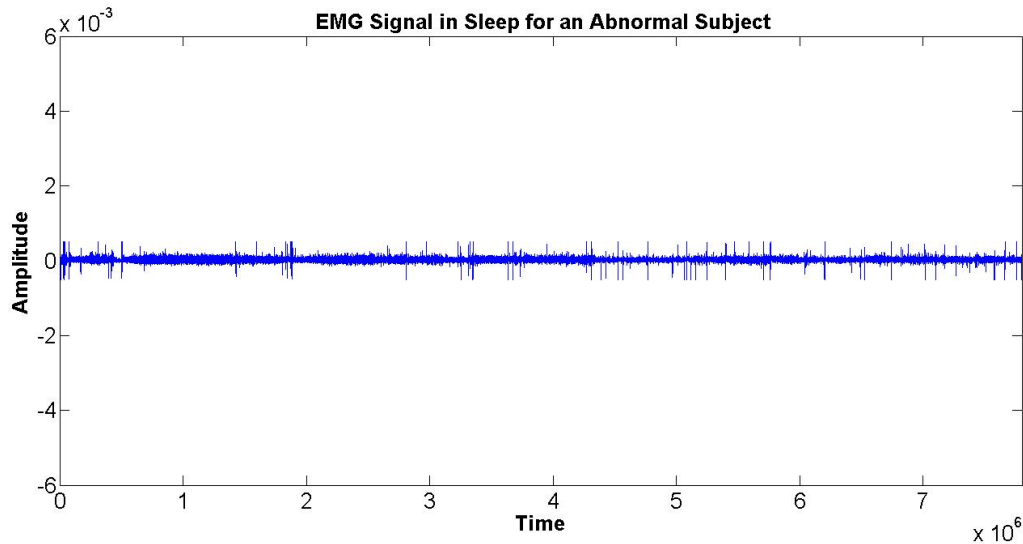
Figure 3.5: An example of EMG signal recorded from the chin for a subject with elevated muscle tone.

- EMG signals are non-linear. The non-linearity is based either on strong regularization or on neurophysiological facts; this reflects how different neuron populations interact. In addition, the nonlinearity can be shown by means of the magnitude and phase plots. Figures 3.7 and 3.8 demonstrate this phenomenon.

- EMG signals, particularly those recorded in PSG are recorded for over 8 hours, which necessitates a large memory for storage as well as analysis. As can be seen in both Figures 3.5 and 3.6, the x axis shows the time samples. As the sampling frequency is 256 Hz for these signals, the time axis corresponds to 9 hours and 8 minutes as well as 6 hours and 51 minutes, respectively.

- EMG signals are non-gaussian. Non-gaussian signals have excitations as a sequence of
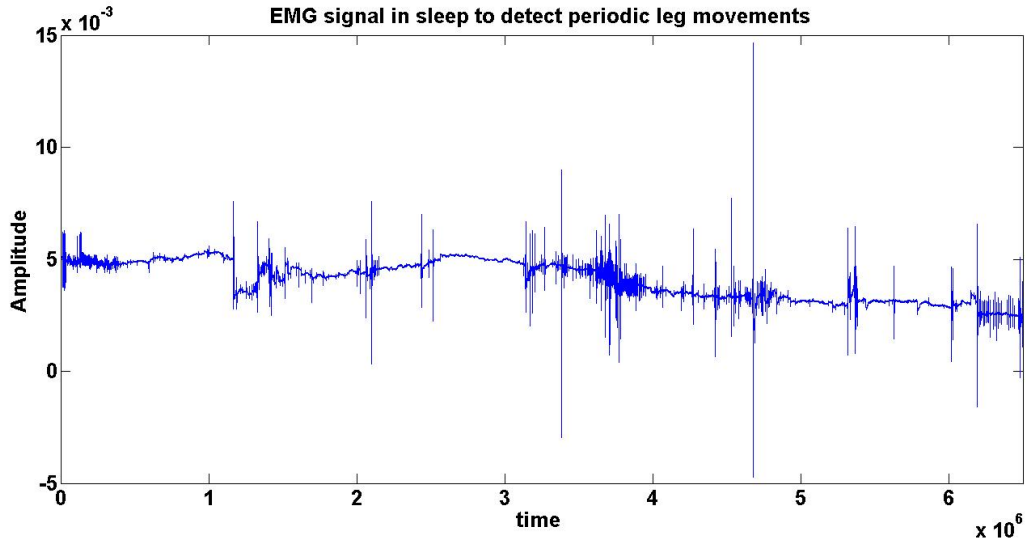
Figure 3.6: An example of the EMG signal recorded from a left leg for a subject high number of PLMs.

symbols with possibly additive noise that makes them complex to analyze. In addition, EMG results from summation of a large number of individual oscillators, which allows for Central Limit Theorem (CLT). However, since these oscillators are not independent, which is one of the requirements of the CLT, EMG signals are considered non-gaussian.

- EMG signals are expected to be multi-component signals due to the possibility that multiple sources of vibration are provided during movements.

- Non-availability of a significant biological marker is another challenge that EMG signals present. This is due to the fact that arriving at a diagnosis regarding the condition of the patient will most likely require the analysis of certain other clinical information.

From the characteristics of the EMG signal listed above, it is evident that EMG is a complex system, and that studying its behaviours promotes development of real-world complex

63
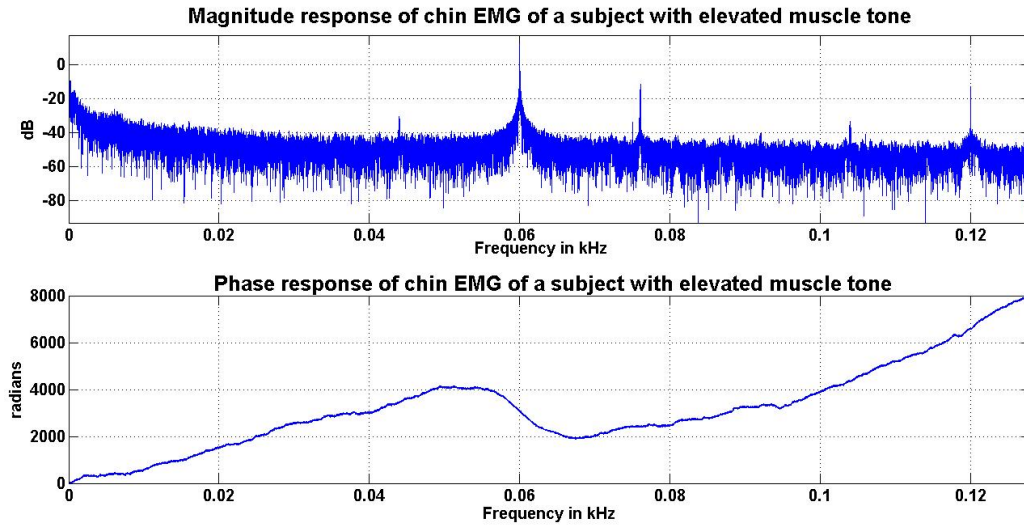
Figure 3.7: An example of how the EMG signal is non-linear by means of its magnitude and phase responses for an abnormal chin EMG subject in REM sleep.

technological systems, as many of the classical signal processing approaches are not suitable for this application. In addition, these characteristics pose a major challenge to the signal processing community. This is the driving force for proposing and developing algorithms based on generation of advanced signal processing.
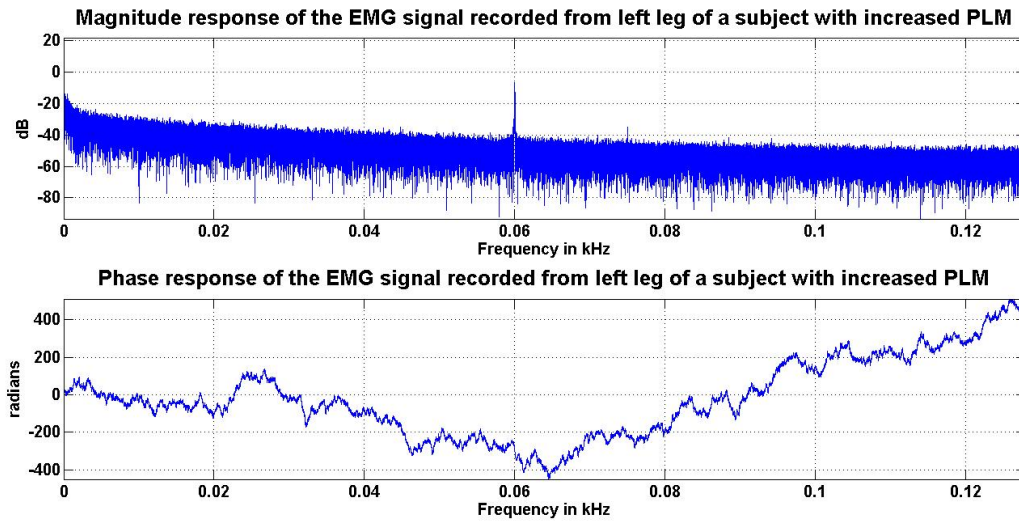
Figure 3.8: An example of how the EMG signal is non-linear by means of its magnitude and phase responses from a left leg for a subject with high number of PLMs.

3.8   Chapter Summary
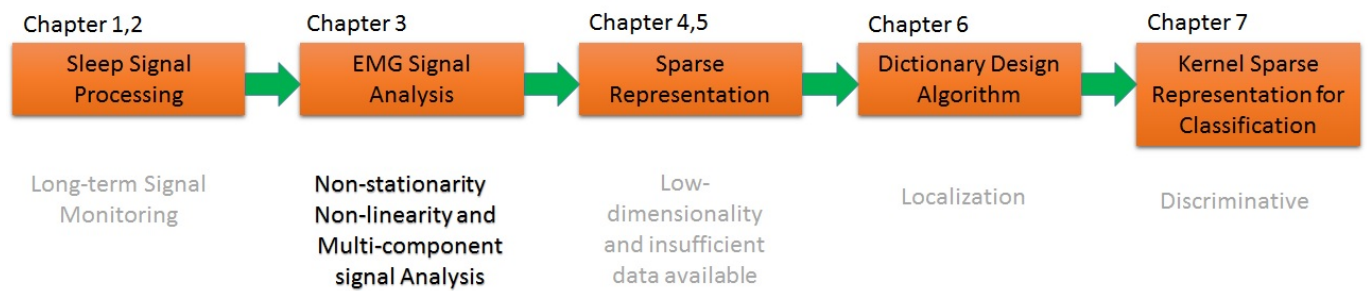
Figure 3.9: Flowchart of the proposed contributions

The contribution flowchart of this chapter is shown in Figure 3.9. This chapter presented a detailed review of the EMG signal analysis along with the history of signal processing, in particular on EMG signals. Regardless, if the EMG is recorded during sleep or not, EMG

signals are non-stationary that have varying spectral and temporal structures over time. In addition, other physiological parameters can affect these EMG signals, in particular ECG. However, EMG signals, particularly those recorded in PSG are recorded for over 8 hours, which requires a large memory for storage as well as analysis. Although traditional signal processing approaches can partially address these issues, state-of-the-art techniques have to be implemented to investigate four different implications as follows: i) long-term sleep signals; ii) non-stationarity iii)localizing to events of interest; iv) identification and classification of sleep signals. This lead us to our subsequent chapters that will emphasize on each of these properties and our developed algorithms for EMG signal processing in sleep. Next chapter describes a new method that was developed for analyzing lower dimensional signals based on sparse representation. The proposed approach uses sparse representation for EMG signals that are recorded during sleep for detecting sleep movement disorders.

Chapter 4

SPARSE REPRESENTATION

Screening of Sleep Movement Disorders Via EMG Signal Analysis
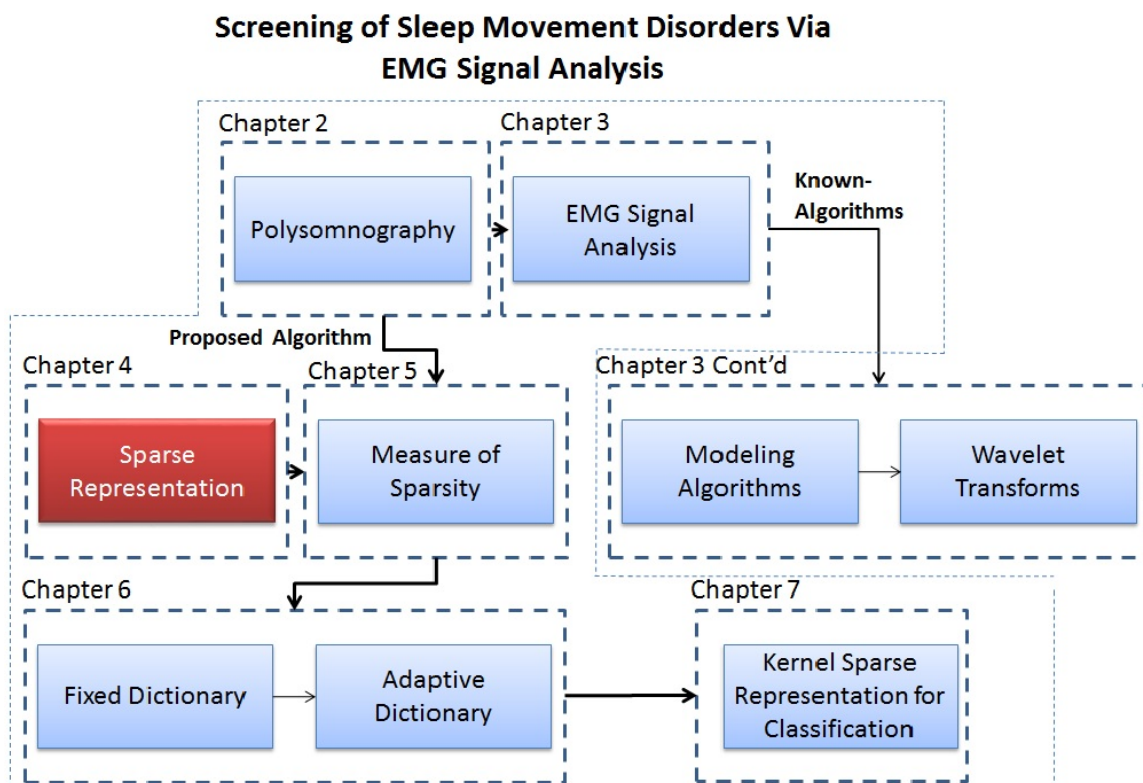
T HIS chapter describes an attempt to develop a new method for analyzing lower dimensional signals based on a advanced signal processing technique. The proposed approach uses sparse representation for real biomedical signals, that is, EMG recorded during sleep for detecting sleep movement disorders. Although sparse representation has been extensively applied on applications for reconstruction as well as classification for higher dimensional signals, it has not been fully explored for applications such as classification of 1-D biomedical signals.

4.1   Motivation

We live in an information age. We have seen a dramatic increase in computational power and data gathering mechanisms, as well as an ever-increasing demand for finer data analysis in applications that rely on scientific and geometric modelling. Each day, millions of large datasets are generated in medical applications, surveillance, and scientific acquisitions. The term "big data" has attracted a lot of attention in recent years. Big data is a collection of large and complex information that traditional data processing techniques are not capable of unravelling their properties. But no matter how big the data is or what insight one can glean from it, it is still just a snapshot: a moment in time. This is the reason we need to shift our thinking to the term "long data". Although, long data may have several interpretations, in the context of this dissertation it means data that may be recorded continuously over a period of time, typically more than 8 hours. Regardless of the definition, the effectiveness of these "long" datasets depends on our ability to efficiently process them, whether it is for storage, transmission, visual display, fast on-line graphical query, correlation, or registration

against data from other modalities. In order to choose the employed signal algorithm, first the characteristics of the signal should be considered. As our modern technology-driven civilization acquires and exploits ever-increasing amounts of data, "everyone" knows that most of the data we acquire can be thrown away with almost no perceptual loss. However, it is very important to choose between those data that we keep and those that we can throw out without actually affecting the performance. In other words, it is important to detect the portion of a signal that has all the information we are looking for and to discard the rest; or, to detect the signature that makes a given signal distinguishable among others and subsequently employ those features for other tasks. One of the techniques that has been attracted a lot of attention in the past decade is the application of sparse representations. This chapter aims to explain the generation of sparse representation and our proposed modification to this algorithm for classification purposes. First, we explain signal approximation. Next, sparse approximation and sparse representation are explored. Our proposed technique to quantify 1-D biomedical signals is reviewed. Finally, in order to evaluate the proposed classification algorithm, we apply them to our EMG signals in sleep, and study the results of the obtained algorithm on real biomedical signals.

4.2   Signal Approximation

One of the most important tools in modern operational systems is the use of optimal approximation (or estimation) methods. Although physical systems can usually be described very precisely according to a certain theoretical model, their actual realization through discrete observations requires additional optimal estimation procedures. A wide variety of such

procedures has been developed over the years, ranging from classic linear methods to more complicated non-linear estimation schemes. The optimal approximation is usually given by a vector space of functions X, and a set of samples from a function $x(t) \epsilon X$. The goal is to find an element $\hat{x} \epsilon X$ that is the optimal approximation to $x(t)$ according to a minimum error criterion [95]. These methods are all convolution-based in the sense that they use an interpolation model of the form:

$$(I_h s)(x) = \sum_{k \in Z} c_h(k) \varphi(\frac{x}{h} - k) \tag{4.1}$$

where h is the sampling stem and $\varphi(x)$ is the basic interpolation kernel [95]. The expansion coefficients in Eq. 4.1 typically correspond to the samples of the input function $s(x)$ taken on a uniform grid: $c_h(k) = s(hk)$. The simplest signal approximation is to represent the signal $x(t)$ as a weighted sum of functions $g_n(t)$ taken from the so called over-complete dictionary D:

$$x(t) = \sum_{n=1}^{\infty} a_n g_n(t) \tag{4.2}$$

Eq. 4.2 often called a Taylor series expansion or a Fourier series, however in engineering it is called an over-complete dictionary. In signal processing applications, it is desirable to find a representation where most of the coefficients in the sum are close to 0, i.e. $a_n = 0$. This over-complete representation is usually discussed as a property of a set of non-zero vectors

$\{\Phi_j(t)\}_{j\in J}$ such that for an arbitrary $x(t) \; \epsilon \; X$,

$$B_1\|x\|^2 \leq \sum_{i\in J} |< x, \Phi_i(t) >|^2 \leq B_2\|x\|^2 \tag{4.3}$$

where $<,>$ denotes the inner product, and B_1, B_2 are positive constant bounds. Eq. 4.3 represents the energy of a sequence in the base functions. Signal approximation using wavelet transform is an example of approximating a signal.

4.3   Sparse Approximation

Complex signals, can consist of contrasting content. For such signals, it is impossible to efficiently represent or model them using a linear transform over a single basis of functions, as done in a short time Fourier transform, or dyadic wavelet transform. A Fourier basis is highly inefficient for modelling transient phenomena, and a wavelet basis is highly inefficient for modelling regular oscillatory phenomena, representing biomedical signals using a union of the two bases promises to greatly increase the efficiency of a model of such signals. This can be done by means of an under-determined system of linear equations depicted in Figure 4.1.

Let us consider $y = Ax$ where $A\epsilon R^{m\times n}$ is a matrix with $m << n$. i.e. there are more variables than equations, often called an under-determined set of equations. In this case, x is under-specified because it can have many solutions that can lead to the same y. In other words, if A is full rank (m), for each $y\epsilon m$, there is solution that is of the form:
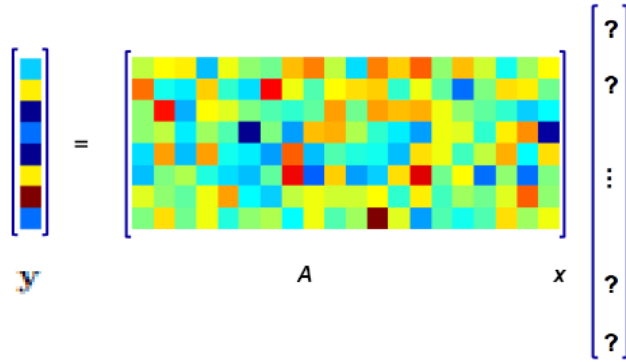
Figure 4.1: Under-determined Linear Systems

$$\{x|ax = y\} = \{x_p + z|z\epsilon\mathcal{N}\} \tag{4.4}$$

where x_p is any particular solution $(Ax_p = y)$. Another particular solution is

$$x_1 = A^T(AA^T)^{-1}y \tag{4.5}$$

and since A is full rank, AA^T is invertible. In fact, x_1 is the solution of $y = Ax$ that minimizes $\|x\|$, i.e., x_1 is solution of optimization problem

$$\text{minimize} \;\; \|x\| \;\; \text{subject to } Ax = y \tag{4.6}$$

Recent work of approximation theory in applied mathematics, and sparse approximation in signal processing, has examined the possibilities of generating very sparse and efficient linear signal models by using a union of several function bases, or any over-complete set of functions [95–98]. This work will demonstrate the benefits of such approaches for many

applications in signal processing and data analysis.

A sparse approximation models, or decomposes, a signal over a dictionary of atoms that need not be linearly independent, in an attempt to create a sparse and efficient representation at any specified precision. The freedom to choose the contents of a dictionary provides considerable flexibility compared with standard analysis methods, such as the Fourier transform and the family of wavelet transforms. Thus, sparse approximation is the problem of finding a signal or vector estimate with a sparseness property; that is, having a small number of nonzero elements, which approximately satisfy a system of equations. The problem of finding the sparse representation of a signal in a given over-complete dictionary can be formulated as follows: Given a $N \times M$ matrix $\mathbf{A}$ containing the elements of an over-complete dictionary in its columns, with $M > N$ and a signal $y \, \epsilon \, R^N$, the problem of sparse representation is to find an $M \times 1$ coefficient vector $\mathbf{x}$, such that $\mathbf{y} = A\mathbf{x}$ and $\|x\|_0$ is minimized, i.e.

$$x = \min \|x\|_0, \quad \text{subject to } A\mathbf{x} = \mathbf{y} \tag{4.7}$$

where $\|x\|_0$ is the ℓ^0 norm and is equivalent to the number of non-zero components in the vector $\mathbf{x}$. Finding the solution to Eq. 4.7 is NP hard, numerically unstable due to its nature of combinational optimization, and therefore it is rarely used in practical applications. Suboptimal solutions to this problem can be found by iterative methods like the matching pursuit and orthogonal matching pursuit. An approximate solution is obtained by replacing

the ℓ^0 norm in Eq. 4.7 with the ℓ^1 norm as follows [96–98]:

$$x = \min \|x\|_1, \quad \text{subject to } A\mathbf{x} = \mathbf{y} \tag{4.8}$$

where $\|x\|_1$ is the ℓ^1 norm. In case, the solution is sparse enough, the solution of Eq. 4.7 is equivalent to the solution of Eq. 4.8. A generalized version of Eq. 4.8, which allows for certain degree of noise, is to find $\mathbf{x}$ such that the following objective function is minimized:

$$J_1(x; \lambda) = \|y - Ax\|_2^2 + \lambda \|x\|_1 \tag{4.9}$$

where the parameter $\lambda > 0$ is a scalar regularization parameter that balances the tradeoff between reconstruction error and sparsity [96]. On the other hand, the sparsity is a penalty to the minimization of reconstruction error, therefore the larger the value of λ, the bigger the penalty.

4.4   Sparse Approximation in Regression

Simple sparse approximation problems originally arose in the study of linear regression. In this setting, a data vector using a linear combination of regressors are approximated. However, the number of regressors has to be controlled to avoid fitting noise in the data. Some of the existing popular regression methods include, Artificial Neural Networks (ANN) [99] or Kernel Machines [100].

Sparse theory brought with itself many approximation tools that were designed for specific

tasks [101]. The main principle behind these techniques is that for a given dataset, the representation that corresponds to minimal error and offers the most compact solution is considered the best approximation and could be used for further analysis [102]. Sparse representation has many applications including compression [103], face recognition [104], gene expression [105], and classification [106]. The association of sparse representation with classification has been relatively underexplored for biomedical signals.

The use of SRC was first proposed by Huang *et al.* in 2007 [107]. In their proposed algorithm, they incorporated reconstruction properties, discrimination power, and measure of sparsity during the approximation process. For the classification process, they used a set of empirical weight vectors combined with a non-linear Fisher information score to optimize the overall performance. Following this, Wright *et al.* [104] proposed a general algorithm for automatic image-based object recognition and considered that by properly harnessing the sparsity, choice of features is no longer critical in assessing the performance of the recognition. Using gene expression data, the sparse technique was successfully used in cancer detection [105], proving its usability for real-life signals. In their work, the authors performed rigorous validation approaches to compare the performance of their algorithm with a variety of Support Vector Machines (SVM). In another work on face recognition applications, Lan *et al.* investigated the discriminative information embodied in the sparse representation on two sets of face databases [108].

All of these cited works serve to illustrate a common theme: they use parsimony as a principle for choosing a limited subset of features or models from the training data, rather than directly using the data for representing or classifying an input (testing) signal. However, to our knowledge, this dissertation is one of the first known works that studies the sparse rep-

resentation for classification approaches to calculate sparse coefficients using all the training samples in addition to the testing samples that are not also members of the training samples. We propose to use sparse signal representation and LOO validation, an optimal validation approach with a least biased estimate to analyze EMG in sleep [106]. This work attempts to address the issue associated with lack of robust generalization techniques for sparse approximation when using a small dataset. We suggest that for a given classification problem, lack of adequate samples tends to bias the training and testing, which not only limits performance but also produces inconsistency for higher-order problems [106, 109]. Below is the brief explanation of sparse representation by ℓ^1 minimization.

4.5 Sparse Representation by ℓ^1 minimization

For a given dataset, $A_i = [a_1^i, a_2^i, ..., a_n^i] \in \Re^{dXn_i}$ for $i = 1, ..., k$, is a matrix where each of its columns are the training samples from the ith class and $\mathbf{y} \in \Re^{dX1}$ is a new test sample. A new matrix A for the entire training set is obtained by concatenating the training samples of all k classes, where a_{lj}^c and y_j are the jth entry of a_l^i and $\mathbf{y}$ respectively. In other words:

$$A = [A_1, A_2, ... A_k] = [A_1^1, A_2^1, ..., A_{n_k}^k] \in \Re^{dXn} \tag{4.10}$$

By rearranging the terms, a linear representation of $\mathbf{y}$ can be rewritten in terms of all training samples as

$$\mathbf{y} = A\mathbf{x}_0 \in \Re^d \tag{4.11}$$

where ideally

$$\mathbf{x}_0 = [0, ...0, \alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,n_i}, 0, ...0]^T \in \Re^n \tag{4.12}$$

is a coefficient matrix whose entries are zero except those associated with the ith class. In other words, the nonzero entries in the estimate $\mathbf{x}_0$ will all be associated with the columns of A from a single object class i and the test sample $\mathbf{y}$ can easily be assigned to that class. If the number of object classes k is reasonably large, the representation of $\mathbf{y}$ is naturally sparse. So the problem can be converted into finding a column vector $\mathbf{x}$ such that $\mathbf{y} = A\mathbf{x}$ and $\|\mathbf{x}\|_0$ is minimized. The so-called sparse representation $\|\mathbf{x}\|_0$ is ℓ^0 -norm, and it is equivalent to the number of nonzero components in the vector $\mathbf{x}$. The following equation is an expression of the ℓ^0 -norm optimization problem [110]:

$$\hat{\mathbf{x}}_0 = argmin\|\mathbf{x}\|_0 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y} \tag{4.13}$$

Due to the nature of combinational optimization, finding the solution to the above expression is NP-hard [104, 111]. Nevertheless, recent developments in the theory of sparse representation and compressed sensing [107, 112–114] have shown that if the solution x_0 is sparse enough, the solution of ℓ^0 minimization is equivalent to the following ℓ^1 minimization problem

$$\hat{\mathbf{x}}_1 = argmin\|\mathbf{x}\|_1 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y} \tag{4.14}$$

This problem can be solved in polynomial time by standard linear programming (LP) methods [106, 114]. The next section continues to explain the solution offered to solve the classification problems based on sparse representation algorithms.

4.6   Critical Analysis of the State-of-the-Art

Lately, there have been numerous attempts to exploit sparse representation, in particular for higher dimensional signals [67, 104, 115–118]. In general, these will be explained below:

4.6.1   Sparse Representation for Higher Dimensional Signals

High-dimensional signals are omnipresent in many areas of science and engineering, in particular in applications such as machine learning, signal and image processing, computer vision, and pattern recognition. In many of these data types, the signals are not distributed uniformly in the immediate surrounding space. As a matter of fact, they lie close to a union of low-dimensional manifolds [104, 119]. Analyzing such low-dimensional signals facilitate the reduction of computational cost in addition to memory requirements. Although, this technique involves many challenges that limit the applicability of these data types, many approaches have been proposed in the literature about machine learning that involves dimensionality reduction, clustering, and classification. Sparse representation for classification is one of the techniques that has shown promising results in high-dimensional signal classification, in particular when the number of classes is high. The SRC uses the training data in addition to assuming that a test sample can be represented based on those training data across all classes. However, based on the definition of sparse representation, a test sample would correspond to a linear combination of a few training samples from the right class. In other words, with sparse representation one can determine the class of a given test sample [120]. However, SRC works under the assumption that the algorithm succeeds

regardless of the modelling and condition of the data. While SRC works well in practical problems, it is important to consider the structural relationships among the data in each class, which often plays an important role in classification applications [107, 120].

4.6.2 Image Processing

The study of sparse representation has become a major field of research in image processing. This has led to state-of-the-art results in image enhancement [104, 119, 121]. While, sparse representation is originally trained to contain sufficient information for reconstruction, from a classification application it is considered a reconstructive approach. Huang *et al.* proposed for the first time that discrimination and reconstruction can work hand in hand and that introducing discrimination properties in to the classical reconstructive energy formulation of sparse coding enhances the classification performance [96, 107]. Their approach proved to yield robust and discriminant image representations by means of an inherent dimensionality reduction.

4.6.3 Face Recognition

Wright *et al.* published their work confined to human frontal face recognition [104]. They used the sparse representation to determine the class label of a new test sample using labeled training samples from k distinct classes. They arranged the given n_i training samples from the ith class as columns of a matrix $A_i = [v_{i,1}, v_{i,2}, ..., v_{i,n_i}]\epsilon\mathbb{R}^{m \times n_i}$. In the context of face recognition, a $w \times h$ is a gray scale image with the vector $\mathbf{v}\epsilon\mathbb{R}^m(m = wh)$ given stacking its columns and the columns of A_i are then the training face images of the ith subject. In their

proposed work, they claimed that if sparsity is properly harnessed, the choice of features is no longer critical i.e. number of features (how large the sample may be) does not have a direct correlation with the performance as long as the sparseness criterion is satisfied. Sparse techniques have also been successfully used for cancer diagnosis applications using gene expression data [105], proving their usability for real-life signals. The authors performed a rigorous validation approach whereby the performance of the algorithm was compared with a variety of SVMs. Lan *et al.* [108] investigated the discriminative information embodied in the sparse representation for two face databases. The original goal of these works was reconstruction using sparse representation in face database, where the limitation of the sample size is not an issue.

All of these mentioned works serve to illustrate a common theme: they use parsimony as a principle for choosing a limited subset of features or models from the training data, rather than directly using the data for representing or classifying an input (testing) signal. In addition, all these efforts have been applied on higher dimensional signals with applications such as face recognition and gene expression. Irrespective of the availability of these techniques for classification applications, there still exist certain open problems, such as:

1. Incompleteness of sparse theory

2. Non-availability of techniques that can accommodate lower-dimensional signals (1-D signals)

3. Non-availability of a robust technique (or generalization) that can accommodate small datasets.

In addition to the above, for a given classification problem, lack of adequate samples tends to bias the training and testing results, limiting their performance consistency for higher-order problems. The next section tends to demonstrate our modification of the SRC algorithm to prepare it for applying sparse representation in the $1-$D signal.

4.6.4   Lower Dimensional Signals

Can sparse representation be applied to lower dimensional signals? What if the dataset available does not have enough samples? These are the questions that inspired this research. The effectiveness of SRC is restricted by an important assumption: that data points from different classes are not distributed along the same radius direction. In other words, the magnitudes of feature vectors are not considered as discriminative information in face recognition applications since images of the same subject under different intensity levels are still considered to be of the same-class [111]. This assumption is reasonable, even if data from different classes are well separated in terms of Euclidean distance. However, due to non-stationarity behaviour, two-class classification problems and low dimensionality of biomedical signals, sparse representation has been relatively unexplored in the context of signal classification. In this work, to achieve optimum performance and yet prepare the signal for sparse representation, modification of SRC is required. Yet, to overcome this problem we propose to use sparse signal representation and LOO validation, an optimal validation approach with a least biased estimate, to analyze EMG in sleep [106]. We have attempted to address the issue associated with lack of robust generalization techniques for sparse approximation when using a small dataset and have suggested that for a given classification

problem, lack of adequate samples tends to bias the training and testing, which not only limits performance but also produces inconsistency for higher-order problems [106, 109] . Figure 4.2 shows the block diagram associated with this chapter. In addition, below is the brief explanation of our modification on sparse representation for signal classification:

Figure 4.2: The block diagram explaining the procedure while using sparse representation

4.6.5   Classification Based on Sparse Representation

Given a new test sample $\mathbf{y}$, the sparse representation $\hat{\mathbf{x}}_1$ is computed using Eq. (4.14). In an ideal case, the nonzero entries in the estimate $\hat{\mathbf{x}}_1$ will all be associated with the columns of A from a single object class i and $\mathbf{y}$ can simply be assigned to the object class with the single largest entry in $\hat{\mathbf{x}}_1$. Noise and modelling errors may lead to small nonzero entries associated with multiple object classes [104]. To overcome this problem, a new characteristic function, $\delta_i : \Re^n \longrightarrow \Re^n$ is introduced for each class, which selects only coefficients associated with that class and forces other coefficients to be zero. In other words, to approximate the given test sample $\mathbf{y}$, $\hat{\mathbf{y}}_i = A\delta_i(\hat{\mathbf{x}}_1)$, is estimated using only the coefficients associated with the ith class and the other values are set to zero [104, 106]. To achieve a generalized sparse representation, training samples are assigned to the object class that minimizes the residual

between $\mathbf{y}$ and $\hat{\mathbf{y}}_i$ [106] as given in Eq. 4.15.

$$\underset{i}{min} \; r_i(\mathbf{y}) = \|\mathbf{y} - A\delta_i(\hat{\mathbf{x}}_1)\|_2 \tag{4.15}$$

Figure 4.3 is an example of how the training and testing samples are generated along with the sparse coefficients corresponding to that sample.
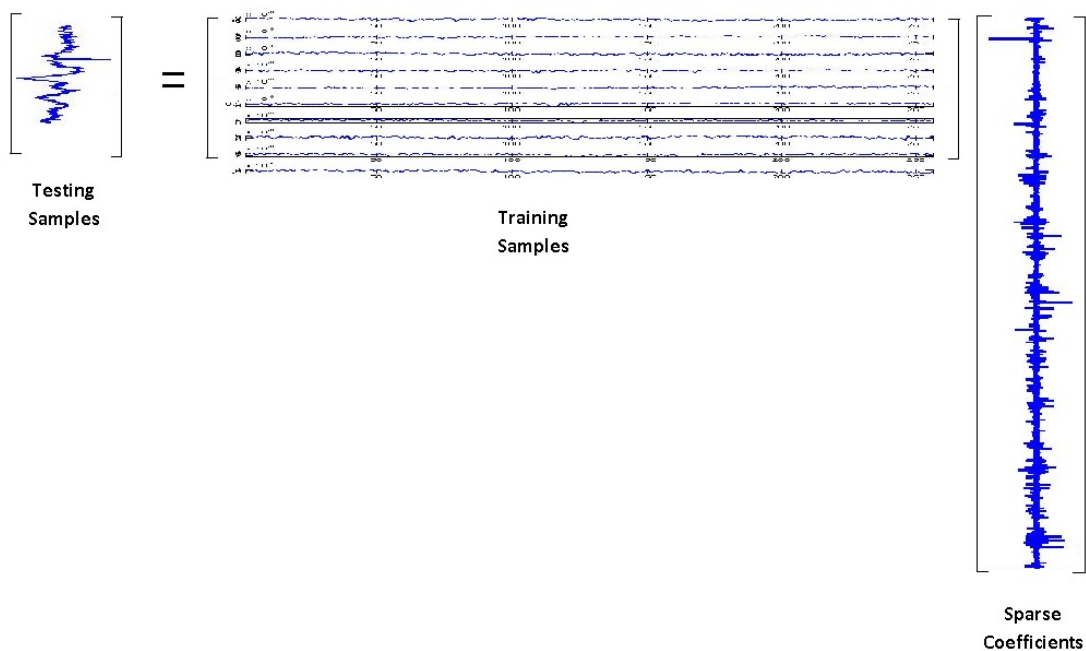
Figure 4.3: The proposed algorithm with the real dataset illustrating the SRC algorithm.

4.6.6 Cross-Validation

Cross-validation is a model validation technique for assessing the results of a statistical analysis. In other words, cross-validation is a statistical method for evaluating a learning algorithm by dividing data into two segments: one used to learn and train the model, and the other to test and validate the model. Cross-validation is assumed to be the best learner/predictor for dataset not already seen by the system [122]. In other words, cross-validation does not use the entire dataset when training the system. Some of the data is removed before the training process and would later be used to test the performance of the learned model on "new" data.

In a typical cross-validation, the training and validation sets must cross over in successive rounds such that all data points have a chance of being validated against each other [109,123]. There are two possible goals in cross-validation:

1. To estimate performance of the learned model from available data using one algorithm or to estimate the generalizability of an algorithm.

2. To compare the performance of two or more different algorithms and find out the best algorithm for the available data, or alternatively to compare the performance of two or more variants of a parameterized model.

4.6.6.1 Holdout Method

The holdout method is considered one of the simplest kinds of cross-validation [124]. The dataset is divided into two sets: training and testing sets. The training set is used to find

the approximation function and the testing set is used to predict the output values for the data. Although this method is advantageous to the residual methods, the evaluation depends heavily on which data points are chosen for training and which ones for testing. Thus the evaluation may be significantly different depending on how this division is made.

4.6.6.2 k-fold cross-validation

The k-fold cross-validation is an improved version of the holdout method. In this algorithm, the dataset is divided into k subsets. To evaluate the performance, each time, one of the k subsets is used for testing and the rest of $k-1$ sets are used for training, with the average error being calculated across the whole dataset. The advantage of such an algorithm is that all the data points will be in the test set exactly once and will get to be in the training set $k-1$ times. The variance of the resulting estimate is reduced as the number of k increases. However, this is said to be an expensive algorithm because of the computational time. Figure 4.4 shows the performance of a 3-fold cross-validation.

4.6.6.3 Leave-M-Out Cross-Validation

Leave-M-out involves using M observations as the validation set and the remaining observations as the training set. This is repeated on the original sample and a validation set of M' observations and a training set. LMO cross-validation requires to learn and validate C_n^m times (where n is the number of observation in the original sample).
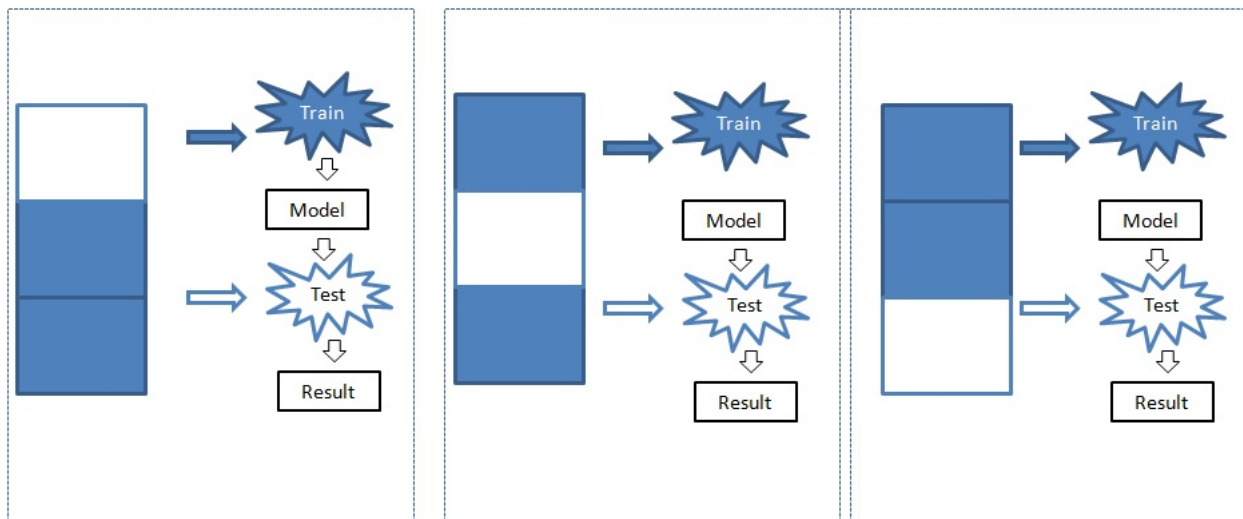
Figure 4.4: Typical 3-fold Cross-validation procedure.

4.6.6.4   Leave-One-Out

Leave-one-out cross validation is a Leave-M-Out cross-validation algorithm taken to its logical extreme with M equal to the number of data points (N) in the set. In other words, for N separate times, the function approximator uses all the samples except for a single observation and a prediction is made for that single observation. This step is repeated for all the data points available in the set, and the average error is computed and used to evaluate the model. Although one may assume this is a very expensive algorithm to compute, LOO is the most unbiased performance estimation and the evaluation error is good.

The algorithm given below summarizes the complete classification procedure. Our implementation minimized the ℓ^1-norm using the LP algorithm based on [125, 126]. In this algorithm our focus is to attain a generalized method using LOO and LMO for classification

of the samples.

Algorithm : Generalized algorithm based on Sparse Representation using LMO Cross-Validation

1. **Inputs**: Data matrix, M number of samples to be out for testing.

2. Use cross validation to randomly split the feature set into a distinct matrix of training samples $A = [A_1, A_2, ...A_k] \in \Re^{dX(n-m)}$ and test sample $\mathbf{y} \in \Re^{dXm}$.

3. Solve the ℓ^1 minimization problem

   $\hat{\mathbf{x}}_1 = argmin \|\mathbf{x}\|_1$ Subject To $A\mathbf{x} = \mathbf{y}$.

4. Compute the sparsity of the sparse coefficients.

5. Compute the residuals: $r_i(\mathbf{y}) = \|\mathbf{y} - A\delta_i(\hat{\mathbf{x}}_1)\|_2$.

6. Identify$(\mathbf{y}) = argmin_i \ r_i(\mathbf{y})$.

7. Cross-validate to check the correctness of the results.

8. **Output**: The average of the correctly identified.

The choice of LMO and in general LOO in this algorithm, refers to cases where the number of available samples is limited. In these cases the unbiased approximation and variability of the estimate on the entire training set is varied. In other words, the difference between the expected LOO error and the true error is approximately zero, where the expectation is over random training sets of samples from the same underlying distribution. To illustrate

the performance and robustness of the algorithm, the next section will evaluate by means of the real world 1-D EMG signals and synthetic dataset.

4.7   Experimental Analysis of the EMG Signals in Sleep

In this section, we present experiments on the two available datasets for sleep movement disorder detection, and classification, which serve to demonstrate the efficacy of the proposed classification algorithm. As explained before, most of the sparse representation algorithms do not use the testing and training samples simultaneously. In other words, they randomly select half of their images (in most of the cases) as the training set and the rest for testing. On the other hand, we use all our samples for both training and testing purposes. In our work, we consider the formulation of our algorithm from the point of view of testing and training all the samples in our dataset. We seek the sparse representation of the EMG signals and to the authors' knowledge, this perspective has not been considered elsewhere. Further, we propose a measure of sparsity as a stopping criteria that automatically selects only a subset of the samples. This will be explained in more detail in Chapter 5. In this section, we perform a set of experiments to verify the accuracy of the developed sparse representation for classification of EMG where the characteristics such as the non-stationarity makes them one of the most challenging signal type. Due to the uncertainties in these signals and their generation process, the spectral and temporal structures of them are varying over time. As well, EMG signal is expected to be a multicomponent signal due to the possibility that multiple sources of vibration is provided during movements. In addition, EMG signals in sleep are exceptionally lengthy.

4.7.1   Problem Statement

Given a one-dimensional EMG signal $y(t)$, we divide this into segments $\mathbf{y}_k$, each of length L samples. Hence, the kth segments $\mathbf{y}_k$ is given by

$$\mathbf{y}_k = [y(1), y(2), ..., y(L-1)]^T \tag{4.16}$$

where $L\epsilon\{128, 256, 512, 1024\}$. Then we construct a new matrix $Y\epsilon\mathbb{R}^{L,k}$ where $k\epsilon\{1,..,K\}$ and whose k^{th} column corresponds to the signal segment y_k.

4.7.1.1   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 [127]. It has been noted that RBD is a precursor to clinically evident Parkinson's disease, although medications may also affect this measure. 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 neuro-degenerative conditions [72]. This may also provide a useful measure for assessing the response to neuro-protective drugs. This is the main reason in using the chin EMG to analyze the behaviour of the muscle tone in REM sleep. Figure 4.9 shows both the normal and abnormal REM EMG signal. The next subsection will explain our procedure in this regard.

(a) Normal EMG Signal in REM Sleep

(b) Abnormal EMG Signal in REM Sleep
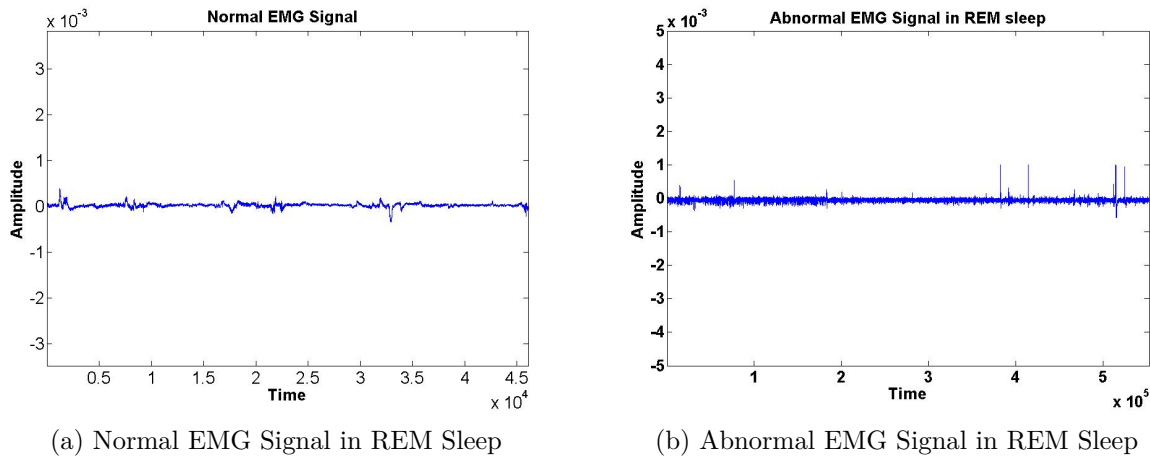
Figure 4.5: (a) The EMG signal of a normal subject. (b) The EMG signal of a subject with elevated muscle tone in REM Sleep.

4.7.1.2 Chin EMG Signal

The chin EMG dataset consists of 36 signals where 16 were normal behaviour subjects and 20 had elevated muscle tone. All the signals were recorded under various laboratory conditions over a nighttime sleep recording session at the Sunnybrook Health Sciences Centre. We computed the recognition rates with the training sample length of 128, 256, 512 and, 1024. Those numbers correspond to 0.5, 1, 2, and 4 seconds respectively. In addition, the classification performance is based on the LMO where the value of M is chosen randomly starting from 1 to 50% of the training space dimension. While sparse representation schemes provide effective means for achieving optimal data reduction by comparing the input with pre-formulated dictionaries, especially for huge datasets, this work intends to prove the usability of SRC by means of a robust generalization technique when the dataset is small. Figure 4.6 shows the recognition performance for various signal segments with three different feature extraction techniques: AR modelling, cepstrum analysis, and wavelet transform.

Figure 4.7 on the other hand uses visual inspection on some examples of the atoms learned by the SRC algorithm and the original data points.
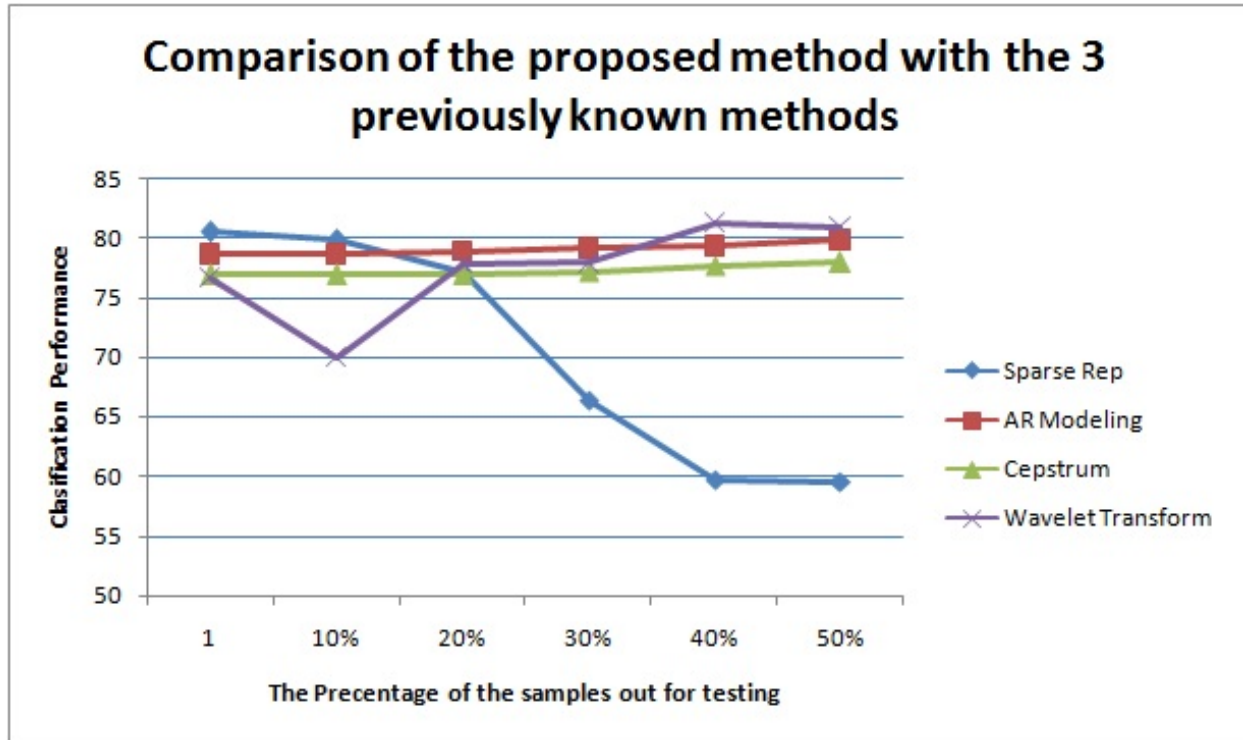
Figure 4.6: Comparison of the proposed method with the 3 previously known methods dataset(A).

In Table 7.1, we present the classification accuracy of the sparse representation for different values of M.

4.7.1.3 EMG and PLM

PLM of sleep is a physiological signal of rhythmic movement of the lower extremity with considerable clinical relevance during sleep. By convention, EMG from bilateral anterior tibialis muscles defines the presence of PLM of sleep. The bilateral anterior tibialis muscle is

(a) Original Data Blocks

(b) Sparse Representation Atoms
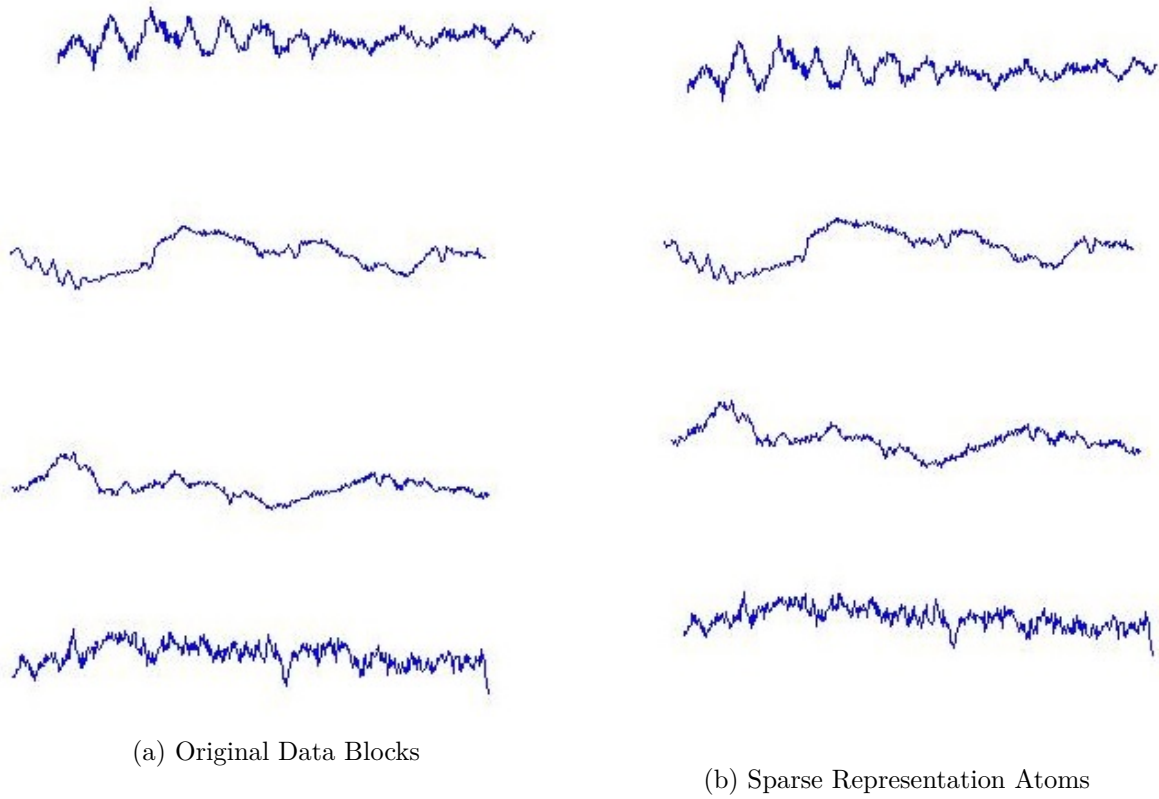
Figure 4.7: (a) Examples of the frames of the original EMG signals, (b) and of the atoms learned by our proposed sparse representation.

Table 4.1: The classification accuracy of the EMG signals fed into a sparse representation algorithm for different M of Leave-M-out.

M	Accuracy of the Proposed Method (%)
1	80.5
10%	79.8
20%	77.1
30%	66.4
40%	59.7
50%	59.5

one of the most robust muscles in the body, and researchers have made a standard recording from it for many years. PLM of sleep is defined as a burst of EMG activity in the anterior tibialis muscle for 0.5 seconds to no longer than 5 seconds with amplitude typically $> 25\%$ of that of the baseline. The periodicity is defined as four consecutive movements with a delay of inter-movement intervals of 4 to 90 seconds. While, traditionally a mean total of five or move PLM per hour of sleep has been considered abnormal, recently researchers have shown PLM per hour, which is known as the PLM index could be up to 25. Therefore, changing a mean total of 30 or more PLM index is considered to be abnormal.

4.7.1.4   LEG EMG Signals

The present analysis included patients who were recruited from a cerebrovascular disease research program at a tertiary stroke centre and also underwent both magnetic resonance imaging of the brain and polysomnography. Sixty five patients were included in the present analysis (mean age 63.5 years, 52% male). 22 patients (34%) presented with stroke, 19 (29%) with TIA ("mini-stroke" happens when a clot stops blood from flowing to the brain for a short time) , and 24 (37%) with other diagnoses.

After the polysomnography and neuro-imaging data were acquired, data segmentation was performed using the hypnogram, a graph of the sleep stages over time. Each sleep staging decision was based on a 30-second window of physiological signals called an Epoch. EMG signals from the first 20 epochs of Stage 2 (where most PLMs occur) were used in our analyses. Similar to the chin EMG, we compute the recognition rates with the training sample dimensions of 128, 256, 512 and, 1024. Those numbers correspond to 0.5, 1, 2, and

4 seconds respectively. Table 4.2 shows the results of the leg EMG data. In this analysis
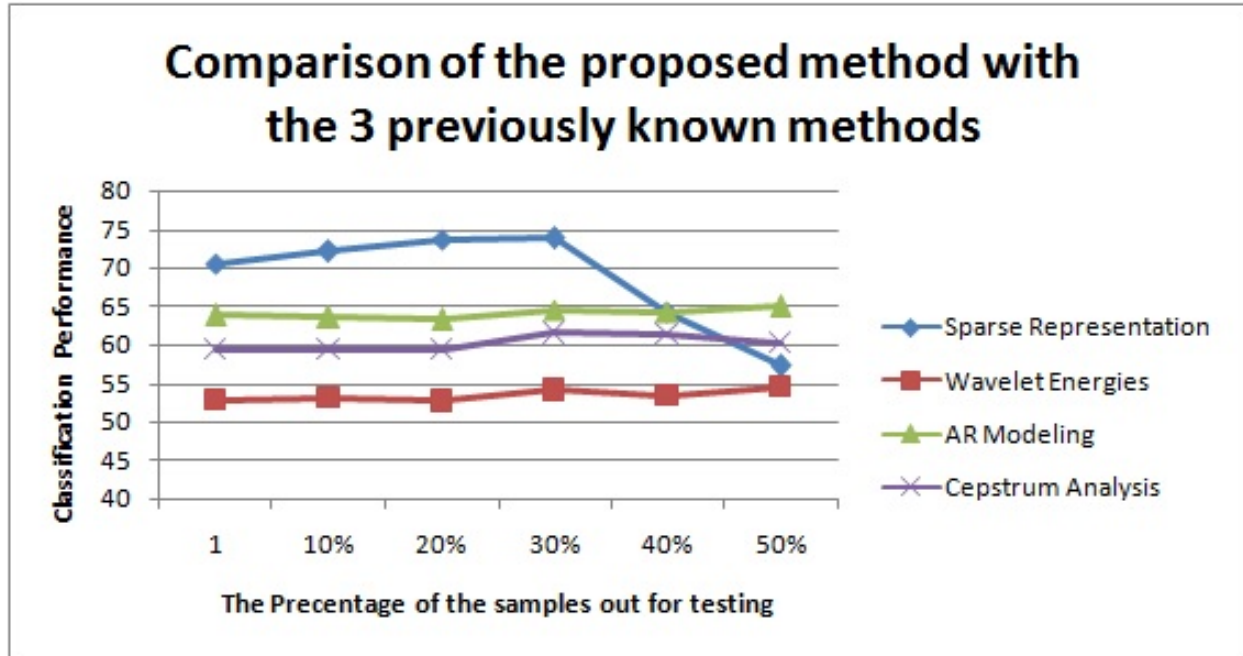
Figure 4.8: Comparison of the proposed method with the 3 previously known methods dataset(B).

we used two classes of subjects: those with PLM index of > 30 and those with PLM index of < 30. This number is the threshold used for developing our class labels. Those with higher PLM index are categorized into abnormal whereas those with lower PLM index are considered normal. In addition, the comparison of the results of our proposed algorithm with the three other known algorithms is illustrated in Figure 4.11.

4.7.2   Evaluation Assessment on Surrogate Synthetic Signal

In addition to analyzing real biomedical signals, we tested the algorithm using synthetic non-stationary signals. The proposed SRC technique with the sparsity measures facilitates the

Table 4.2: The classification accuracy of the EMG signals for PLM subjects fed into a sparse representation algorithm.

Length of the Signal	SRC using LOO
128	71.68
256	70.56
512	68.84
1024	66.31

adaptation of the LOO method to any of the signals available in automating the identification process of the SRC application. The surrogate synthetic signals consist of two non-stationary signals with spikes and signals with different local correlations, and two stationary signals with linear and power-law trends. Figures 4.9 to 4.12 show these signals respectively.
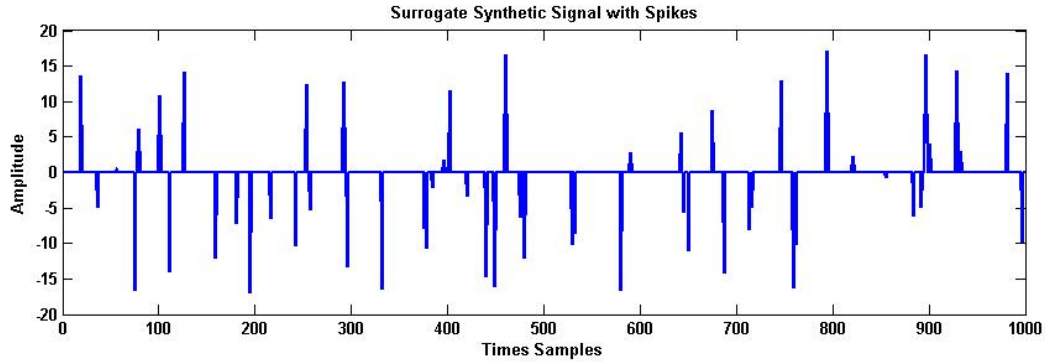
Figure 4.9: Non-stationary synthetic signal composed with spikes probability $p = 0.05$, Amplitude $A_{sp} = 10$.

We then calculated the SRC performance for different M. Table 4.3 summarizes the results we achieved for different scenarios.

Although results achieved by the surrogate synthetic signals are not promising, it is consistent with our results achieved by the real EMG signals. In all the cases, we achieve a better performance for LOO in comparison with LMO.
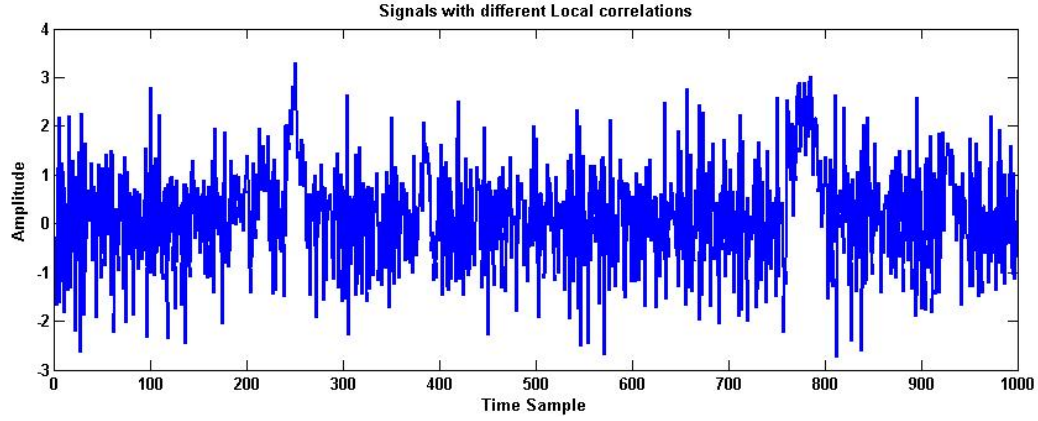
Figure 4.10: Non-stationary synthetic signal with different local correlations composed of mixed signals.
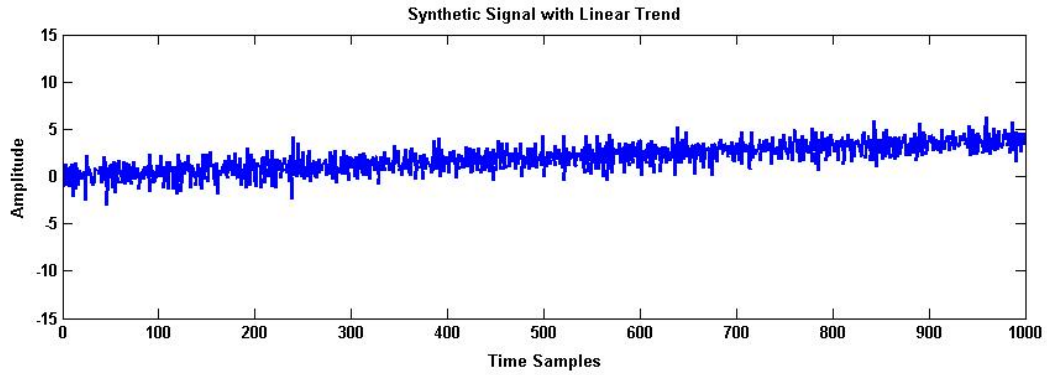
Figure 4.11: Surrogate signals with linear trends. $A = s^{-8}/index$.

Table 4.3: The performance accuracy of the SRC algorithm on surrogate synthetic signals

Signal	M=1	M=50%
Linear/Power-law	75%	62.6%
Spikes/Correlations	52%	45%

96
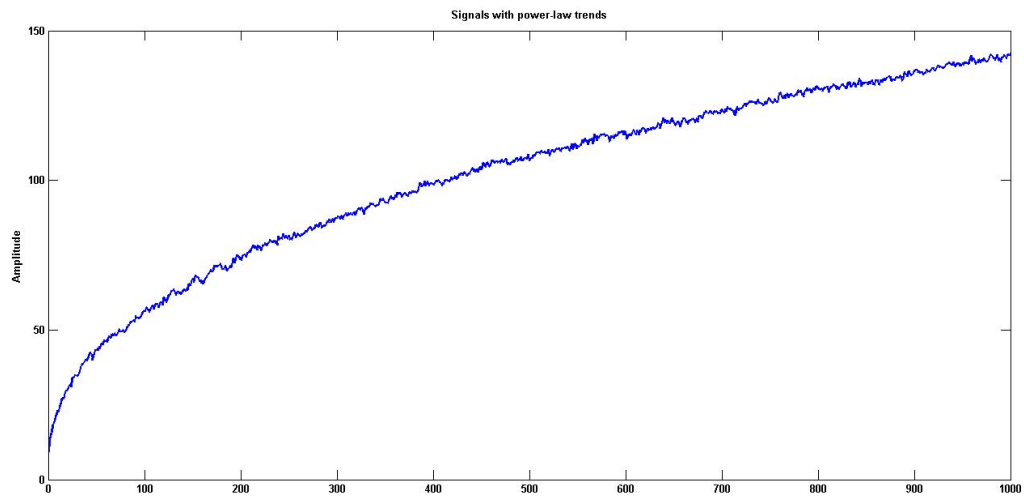
Figure 4.12: Surrogate signals with power-law trends.

4.8   Summary

In summary, we applied the proposed sparse representation for classification of EMG signals both in REM and Stage 2 of sleep for detection of RBD and PLM disorders respectively. In this algorithm, we segmented each signal into equal length blocks and used each block to construct the training samples. Our testing algorithm was based on LOO and LMO as well as the ℓ^1 minimization. To achieve a generalized sparse representation, testing samples were assigned to the object class that minimizes the residual between those samples and the reconstructed signal. As an example, Figure 4.13 shows the residual values chosen in random from the training and testing samples of the chin EMG signal. The effectiveness of our proposed algorithm was compared with the three known algorithms of AR modelling, cepstrum analysis and wavelet transforms. Modelling was used because it allowed us to separate the signal into possibly correlated components based on the frequency and time behaviours in the signal. Specifically, AR modelling was chosen because it has been widely used and proven to model and predict various types of natural phenomena, such as earthquakes and neural activity. AR modelling expresses a particular point in the data as having a weighted sum of the past values along with a random component. These weighted values are parameters of the model and are solely based on the order of the model. A problem arises when determining the optimal model order to use. Higher order models are affected by noise while lower models do not detect high frequency components. Additionally cepstral modelling was used to monitor the rate of change in the frequency spectrum. Although AR modelling and cepstrum analysis have been found to provide a sufficient and accurate representation for many different types of signals in many different applications, they require

the fitted signal to be stationary over a given interval. For such algorithms, adaptive segmentation techniques could possibly segment the signal into stationary components, which makes them problematical and complicated algorithms that may not be suitable for real-life signal applications.

Wavelet transforms are commonly used in EEG processing because they decompose the data into components with a series of fixed frequency bands. Obtaining these bands enabled us to obtain frequency changes at a specific time, which meant we could compare frequency bandwidths (for example alpha, beta, etc) of different signals. Wavelet transforms are also designed to specifically work with signals that are non-stationary. While, wavelet transforms have good ability to compress signals, in particular non-stationary signals, they work best when the signal length is $= 2^{Some_i nteger}$. Otherwise wavelets approximate the left side of the signal at the expense of the right side [128]. In addition, wavelet transforms cannot support weighted distance measures [128].

Regardless of the advantages and the disadvantages of these three algorithms, from Figures 4.6 and 4.11 is it apparent the proposed algorithm produces higher accuracy in particular for the LOO algorithm. However, the performance decreases as we decrease the number of validations (i.e. by increasing M for testing). This is because, sparse representation fails when there is occlusion in the system. In other words, when we do not use enough samples to train the system, sparse representation can not adapt to the changes of the non-stationary signals and will in turn produce lower classification performance.
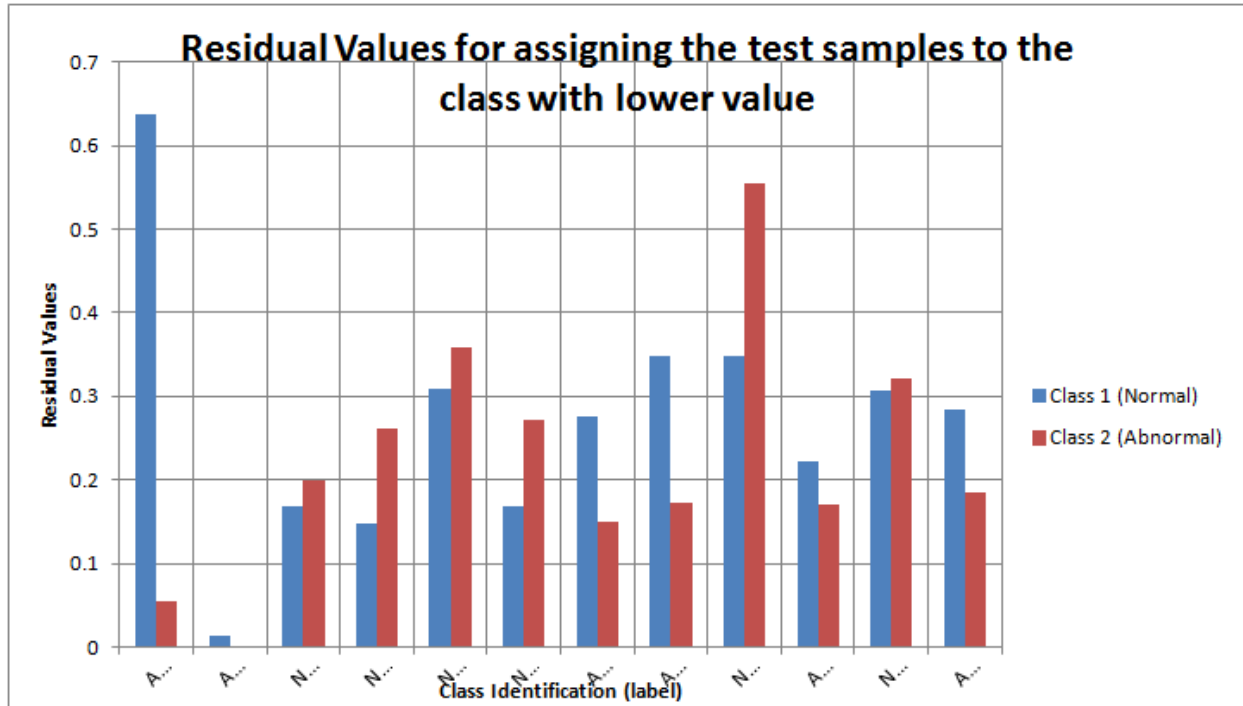
Figure 4.13: The two residual value for every sample. The algorithm assign the test sample to the class with lower residual value.

4.9   Challenges

The challenges we faced using sparse representation for classification are as follows:

- Dealing with real data that contain missing observations, corruptions, and outliers.

- Handling insufficient data.

- Expensive computation for long-term signals such as sleep.

- Need of real-time implementation for pattern recognition system.

- Need of a robust feature extraction techniques by means of sparse representation.

4.10   Summary of the Contributions

In this chapter, we have designed for the first time a generalized sparse representation algorithm for 1-D signals that is suitable for classification applications. The properties of such an algorithm are listed in the following criteria:

Long-term analysis: We used long-term EMG signals to analyze the behaviour of such signals for the two chin and leg EMG.

Non-stationary compatibility: This algorithm does not use any assumptions in regards to the dynamic changes in a given signal, as it can detect those changes by means of evaluating the sparse coefficients.

Real-life signal monitoring: For the first time, we used the sparse representation for classification of the long-term EMG signal for detecting and classifying the signal into normal and abnormal classes.

Segment-based classification: It has come to our attention that many articles have based their classification rates on a signal by signal basis, meaning that they classify the signals by whether the entire signal is normal or abnormal. With our proposed algorithm, we decided on the normality and abnormality of the EMG signals based on the majority ruling. In other words, a subject may have had an abnormal EMG movements during sleep for a very short period of time, but may have been a normal subject, and vice versa.

4.11   Chapter Summary

Chapter 1,2	Chapter 3	Chapter 4,5	Chapter 6	Chapter 7
Sleep Signal Processing	EMG Signal Analysis	Sparse Representation	Dictionary Design Algorithm	Kernel Sparse Representation for Classification
Long-term Signal Monitoring	Non-stationarity Non-linearity and Multi-component signal Analysis	**Low-dimensionality and insufficient data available**	Localization	Discriminative

Figure 4.14: Flowchart of the proposed contributions

The contribution flowchart of this chapter is shown in Figure 4.14. This chapter presented a classification approach based on sparse representation. We proposed novel generalization algorithms based on the linear programming problem, using the LOO and LMO approaches. In this chapter, we focused on quantification and detection of sleep abnormality. The performance measures obtained using long term sleep data have validated the usefulness of the technique for real time data processing. This representation was successfully used to quantify EMG signal in both RBD and PLM disorder subjects.

Since this is the first work reported in the area of EMG in sleep movement disorders, the results are promising and show great potential for applications such as therapy and the prediction of fatal diseases. This is especially true since all techniques were done without any assistance or interventions from the user as the procedure was fully automated. As such, systems with this capability could in fact be used not only as tools to detect abnormality behaviour but also as earlier detection and diagnostic tools for different diseases such as Parkinson's or stroke. Nevertheless, this chapter mainly uses a fixed matrix for training

the samples that mainly refers to dictionary selection. Dictionary selection and dictionary learning are the two main methods that emerged to determine a dictionary within a sparse decomposition. Dictionary selection requires choosing a pre-existing dictionary. These dictionaries could be constructed using the Fourier basis, wavelet basis, or modified discrete cosine basis. As well, it can be constructed by a redundant or over-complete dictionary by forming a union of bases (for example the Fourier and wavelet bases) while allowing a representation of different properties of the signal [129, 130]. On the other, hand dictionary learning algorithms tend to capture the specific features of the signal or a set of signals to update the elements of the dictionary atoms. Dictionary learning methods are often based on an alternating optimization strategy, in which the dictionary is fixed, and a sparse signal decomposition is found; then the dictionary elements are learned, while the signal representation is fixed [129, 131].

For real life signals such as sleep signals, analysis is difficult when an event of interest is only present for a short period of time. Algorithms designed to detect these small changes and localize the events of interest would be of great importance. As well, designing dictionaries that could be adapted to the signal of interest is of significant value. In addition, it has been shown that with sparsity properly harnessed, the inconsistency and lack of adequate sample size decreases when LOO method is considered. As such, in the next chapters, we further investigate the implications and different methods used to measure sparsity, as well as dictionary learning algorithms in addition to our modification and their implementation and results.

Chapter 5

Measure of Sparsity

Screening of Sleep Movement Disorders Via EMG Signal Analysis
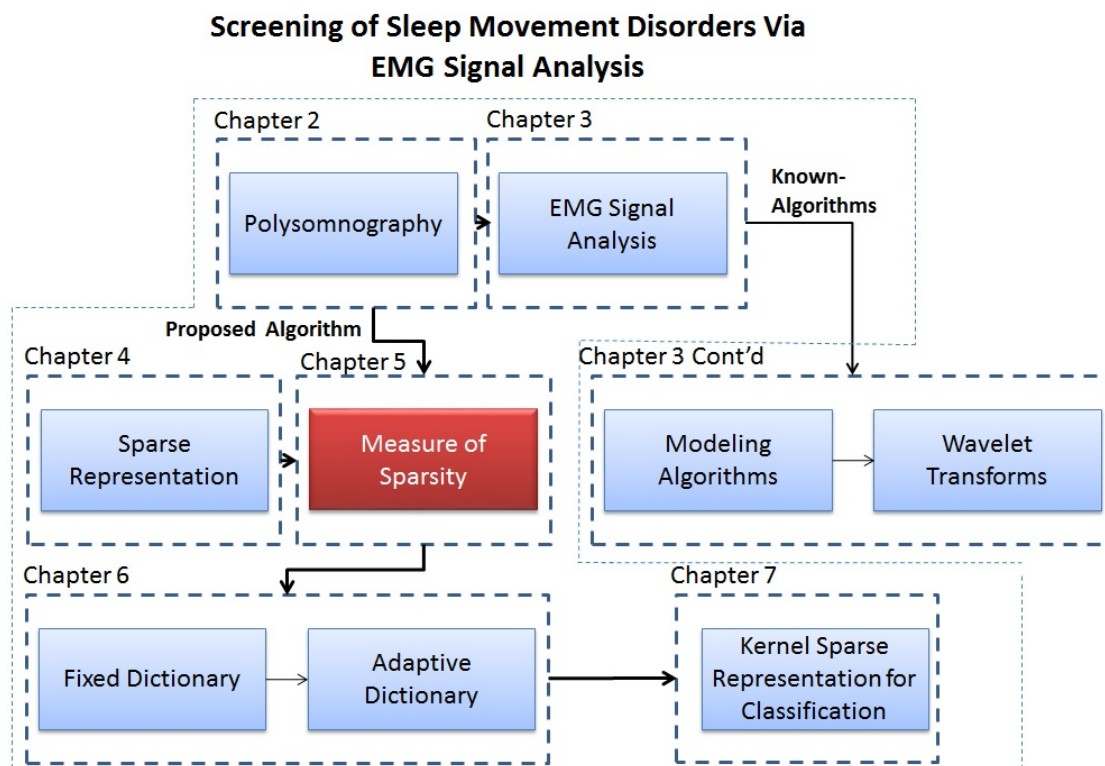
T HIS chapter will explain one of the key concepts of fundamental importance in sparse representation, called the measure of sparsity, by means of different sparsity criteria and measures. The aim of this chapter is to compare several commonly-used sparsity measures and to show our proposed method to measure this phenomenon.

5.1 Motivation

In the past several years, there have been exciting breakthroughs in the study of sparse representation, and the concept of sparsity has been readily used in diverse areas of research such as face recognition [104], image processing [132], medical imaging [121], and antennas [119]. In addition, sparsity has shown a great deal of success in many machine learning algorithms, and techniques such as matrix factorization [133, 134], feature extraction [135], compressed sensing [136], dictionary learning and designed algorithms [131], and signal representation [137]. There are many measures of sparsity that lead to an interpretation of the definition of the term. Intuitively, a sparse representation is one in which a small number of coefficients have all the energy, so that the signal can be represented by only those few coefficients. However, the question one should ask is which sparsity would result in a better and more robust representation. As mentioned, the best sparsity is a distribution of energy on one coefficient and zero elsewhere, and the least sparsity is the even spread of energy over all coefficients. That said, it is reasonable to assume that any "good" measure of sparsity should increase as we move to the coefficient with the highest energy. In other words, if we multiply the two extreme cases of sparsity, the best sparse and the least sparse, with a

constant, we still get the same energy distribution (the first has the energy distributed only on one coefficient and the latter has the energy distributed evenly throughout all coefficients). Thus, it is reasonable to assume that changing the energy value, e.g., doubling the energy, would not affect the sparsity; in summary sparsity is invariant under multiplication by a factor. This chapter aims to find that the best sparsity is invariant to change. To achieve this goal, we will first review the criteria that have been introduced in the literature [138].

5.2   The Sparsity Criteria

The following are the 6 main criteria that contribute to the measure of sparsity [138], as introduced by Hurley *et al.* in their paper. Although most of them have been applied to financial settings (distribution of wealth), they can be used interchangeably for distribution of energy of the coefficients. In general, the best sparsity is often referred to as inequality of distribution, and the least sparsity is referred to as equitable distribution [138–140]. If S corresponds to sparsity and $\mathbf{x} = [x_1 \ x_2 \ ...]$ are the coefficients magnitudes and $\alpha \epsilon \mathbb{R}$ is the constant, we have the criteria as follows [138–140]:

1. **Robin Hood-** Robin Hood decreases sparsity. This is in direct relationship with the definition of sparsity, which is a distribution of energy in only a few of the coefficients. In other words, robbing α from the rich (x_i) and giving to the poor (x_j) decreases the inequality of wealth distribution. Mathematically we have: $S([x_1, \ ... \ x_i - \alpha, \ ... \ x_j + \alpha, \ ...]) < S(\mathbf{x})$ for all α, x_i, x_j such that $x_i < x_j$.

2. **Scaling-** As mentioned before, sparsity is invariant to scale changes. Doubling the

energy distribution does not alter the overall coefficients' energy. In other words, the relative energy distribution is of importance and not the absolute value, i.e. making everyone 10 times richer does not affect the distribution of wealth. Mathematically:

$S(\alpha\mathbf{x}) = S(\mathbf{x}), \forall\alpha\epsilon\mathbb{R}$

3. **Rising Tide-** Although multiplying the constants with a coefficient does not affect the overall distribution of energy, adding a constant would decrease sparsity, in particular when this value is very large i.e., giving everyone a large amount of money, say 2 million dollars, will make everyone equally wealthy, which in turn reduces the sparsity. Alternatively, $S(\alpha + \mathbf{x}) < S(\mathbf{x})$ where $\alpha\epsilon\mathbb{R}$ and $\alpha > 0$.

4. **Cloning-** Sparsity is invariant under cloning. In other words, sparsity of two identical samples would still be same even if the two samples are added. Mathematically $S(\mathbf{x}) = S(\mathbf{x}\|\mathbf{x}) = S(\mathbf{x}\|\mathbf{x}\|\mathbf{x}) = S(\mathbf{x}\|\mathbf{x}\dots\|\mathbf{x})$ where$\|$ defines concatenation.

5. **Bill Gates-** If the energy distribution of one of the samples increases, the set is said to be more sparse. Similar to wealth distribution, if one individual becomes infinitely wealthy, the wealth distribution becomes very sparse. Alternatively, $\forall i, \exists\beta = \beta_i > 0$ such that $\forall\alpha > 0$:

$S[x_1 \dots x_i + \beta + \alpha \dots]) > S([x_1 \ x_i + \beta \dots])$

6. **Babies-** Babies increase sparsity. In a signal that has some sort of energy distribution, adding zeros increases the sparseness of the distribution. In other words, $S(\mathbf{x}\|0) > S(\mathbf{x})$.

5.3 Sparsity

In this section we discuss a number of popular sparsity measures. A vector $\mathbf{x} \epsilon \mathbb{R}^n$ is sparse if only a few entries are nonzero. The number of nonzero is called ℓ^0-**norm** of $\mathbf{x}$, which is one of the most traditional sparsity measures. Mathematically, it is as follows:

$$\|x\|_0 \doteq \#\{i | x_i \neq 0\} \tag{5.1}$$

However, ℓ^0 norm has been found not to be suitable for many practical scenarios, mainly because its derivative contains no information (derivative is zero): hence it cannot be used in any optimization problem. Additionally, the presence of noise makes the ℓ^0 norm ineffective. For this reason, researchers have defined ℓ^0_ϵ in noisy settings where we are interested in some coefficients greater than a threshold ϵ. However, similar to ℓ^0, ℓ^0_ϵ has a gradient with no information and hence it is not useful for optimization applications. To yield a better performance, ℓ^p norm is often used in place with $0 < p < 1$. The ℓ^1 measure which is ℓ^p with $p = 1$ can easily be calculated via linear programming and can be used to approximate ℓ^0 norm. In general:

$$\|x\|_p = (\sum_i |x_i|^p)^{1/p} \tag{5.2}$$

$$\|x\|_0 = \lim_{p \to 0} \|x\|_p^p \tag{5.3}$$

where the geometrical interpretation of ℓ_p norm is demonstrated in Figure 6.3. Other less

popular sparsity measures are available which approximate ℓ^0 measures but emphasize on different properties. One is the $\tanh_{a,b}$ which is used instead of ℓ^p since it is limited to the range of $(0,1)$. One benefit of ℓ^p and $\tanh_{a,b}$ is that they preserve the sparsity even if the large component has been divided into smaller components [138]. The log measure on the other hand enforces sparsity outside some range. This is very effective, in particular when low energy coefficients are present, as it spreads the energy of the small components. The $pq-$mean is the ratio of the generalized p mean and q mean. The most popular of these is the Gini Index [141]. Given a vector $\mathbf{x} = [x_1 \; x_2 \; x_3 \; ... \;]$ is a sorted vector, arranged in ascending order, $x_{(1)} \leq x_{(2)} \leq ... \; x_{(N)}$ where $(1), (2), ...(N)$ are the new indices after the sorting operation. Thus the Gini Index is of the form given by Eq.5.4:
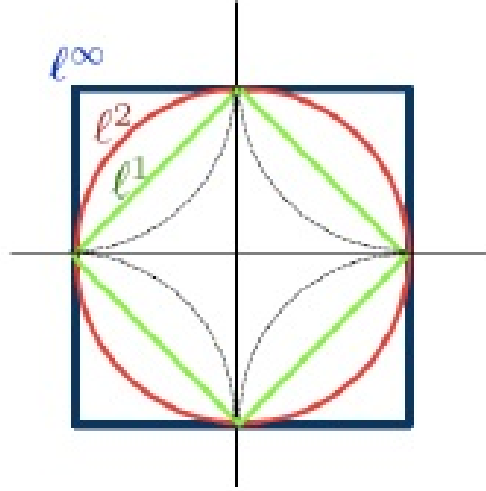
Figure 5.1: Geometrical representation of the ℓ_p-**norm**

$$S(\mathbf{x}) = 1 - 2\sum_{k=1}^{N} \frac{x(k)}{\|\mathbf{x}\|_1} (\frac{N - k + \frac{1}{2}}{N}) \tag{5.4}$$

This is an interesting measurement since as we increase the number of coefficients, the measure of sparsity calculated using the Gini Index will converge. The Gini Index was originally proposed by economists as a measure of the inequality of wealth. When the Gini Index is used in signal processing, the notations change to "*efficiency of representation*" or sparsity. Table 5.1 lists some of these popular measures of sparsity and Table 5.2 compares different measures of sparsity with the criteria defined above.

Table 5.1: The commonly used sparsity measures

Sparsity Measures	Definition
ℓ^0	$\#\{j, x_j = 0\}$
ℓ^0_ϵ	$\#\{j, x_j \leq \epsilon\}$
ℓ^1	$\sum_j x_j$
ℓ^p	$(\sum_j x_j^p)^{1/p} \quad 0 < p < 1$
$\tanh_{a,b}$	$\sum_j \tanh((ax_j)^b)$
log	$\sum_j \log(1 + x_j^2)$
$pq-$ mean	$(\frac{q}{N}\sum_j^N x_j^p)^{\frac{1}{p}}(\frac{1}{N}\sum_j^N x_j^q)^{\frac{-1}{q}} \quad p < q$
Gini	$1 - 2\sum_{k=1}^N \frac{x_{(k)}}{\|\mathbf{x}\|_1}(\frac{N-k+\frac{1}{2}}{N})$

Table 5.2: Comparison of Different Sparsity Measures using the 6 different Criteria

Sparsity Measures	Robbin Hood	Scaling	Rising Tide	Cloning	Bill Gates	Babies
ℓ^0	$\checkmark$					$\checkmark$
ℓ^0_ϵ						$\checkmark$
ℓ^1			$\checkmark$			
ℓ^p	$\checkmark$		$\checkmark$			
$\tanh_{a,b}$	$\checkmark$		$\checkmark$			
log			$\checkmark$			
$pq-$ mean	$\checkmark$	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$
Gini	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

The next section will explain our proposed measure of sparsity along with the results driven from the REM EMG datasets.

5.4   Sparsity Measure

There are many studies suggesting that random features perform well for sparse signal recovery [142, 143]. If any signal x is sparse in some known orthogonal basis, the properties of sparse recovery from random measurements show that the signal can be reconstructed using those coefficients. How much sparsity is needed to efficiently reconstruct the signal is of great importance. Recent research has demonstrated that in a more practical situation, if the signal is less sparse and A does not have good properties for sparse recovery using ℓ^1 minimization, then the signal cannot be recovered [144, 145]. However, the measure of sparsity has not been fully analyzed for signal classification. In this work, the measure of sparsity is used as a feature to perform classification rather than relying only on a single statistic for both validation and identification [106]. In our work, we used residual error to measure how well the representation approximates the test sample, and we used the degree of sparsity to evaluate how accurate the representation was. Our proposed measure of sparsity is as follows:

$$S_p(x) = (\frac{1}{C_p})(\frac{(\frac{1}{n-1})\sum_{i=1}^{n}|x_i - m|^p}{\frac{1}{n}\sum_{i=1}^{n}|x_i|^p})^{\frac{1}{p}} \tag{5.5}$$

where

$$m = (\frac{1}{n})\sum_{i=1}^{n}x_i \tag{5.6}$$

and

$$C_p = (\frac{(n-1)^{p-1} + 1}{n^{p-1}})^{\frac{1}{p}} \tag{5.7}$$

The default value for p is chosen to be one and n is the length of the feature set. The sparsity is one if and only if a vector contains a single non-zero component, and is zero if and only if all components are equal. Evaluating the 6 criteria for this measure of sparsity, it is apparent that it satisfies all criteria except the cloning criterion. This is shown in Table 5.3. Unlike the Gini Index it does not satisfy all the criteria, but it uses fewer computational resources than the Gini Index as it does not need the samples to be in ascending order. The next section evaluates this measure of sparsity with the algorithm we proposed in the previous chapter.

Table 5.3: The commonly used sparsity measures.

Property	Samples with criterion	Desired Outcome
Robin Hood	$[1,0,0,5]$ vs $[1,-2,0,7]$	$S[1,0,0,5] > S[1,-2,0,7]$
Scaling	$[1,0,0,5]$ vs $[2,0,0,10]$	$S[1,0,0,5] = S[2,0,0,10]$
Rising Tide	$[1,0,0,5]$ vs $[3,2,2,7]$	$S[1,0,0,5] < S[3,2,2,7]$
Cloning	$[1,0,0,5]$ vs $[1,0,0,5,1,0,0,5]$	$S[1,0,0,5] \neq S[1,0,0,5,1,0,0,5]$
Bill Gates	$[1,0,0,5]$ vs $[1,0,0,100]$	$S[1,0,0,5] < S[1,0,0,100]$
Babies	$[1,0,0,5]$ vs $[1,0,0,0,0,0,5]$	$S[1,0,0,5] < S[1,0,0,0,0,0,5]$

5.4.1   Evaluation Assessment on Chin EMG Dataset

In this section, we compute the classification rates with the feature space dimensions of $956X26$ [72]. The feature space for this section contains the AR coefficients for just 8 subjects of EMG signals with 4 being normal and 4 with elevated muscle tone. These 8 signals were first segmented into stationary components using adaptive signal processing

techniques. 956 samples were extracted, of which 547 were normal segments and 409 were abnormal segments. On each of these stationary segments, modelling techniques were applied [72]. To evaluate the performance of our proposed algorithm, SRC was applied to this feature set. The value of M is chosen randomly starting from 1% to 50% of the feature space. In Table 7.1, we present the classification accuracy of the AR coefficients using sparse representation for different M, as well as the degree of sparsity of the sparse approximation coefficients. The degree of sparsity for all sparse approximations of this feature set is calculated using Eq. 6.4. From this table, it is apparent that the generalization scheme in practical applications strongly depends on the value of M. It is also obvious that sparsity measures reduce as we increase the value of M; the higher the value of M the lower the accuracy. Although the LOO cross-validation method generally incurs a high computational cost, it is the least biased estimate, since rigorous validation is performed compared to LMO cross-validation, which exhibits a comparatively higher variance. In this case, the algorithm reaches its maximum learning capacity (stability) by the LOO method, since the perturbation induced by LOO is small and therefore the classifier is stable. As we increase the size of the perturbation, stability is less likely to hold, and the sparse approximation is less sparse; thus the accuracy decreases by about 20%. To prove the generality of the algorithm, we analyzed the algorithm using different number of segments. This is shown in Figure 6.1, where we used 10%, 20%,...100% of the segments and applied LMO for both $M = 1$ and $M = 50\%$ of the original feature space. The overall performance of all the cases when $M = 1$ is about 80.5%; however, this value reduces by about 20% to 60% as we increase M. The same analysis was applied to a different random feature of cepstrum coefficients [72] with overall average accuracy of 72% for $M = 1$ and a 10% reduction when

Table 5.4: The classification accuracy of AR coefficients fed into a sparse representation algorithm for different M of Leave-M-out, as well as the degree of sparsity of sparse coefficients.

M	Accuracy of AR Coefficients $(\%)$	Sparsity of the Sparse AR Coefficients
1	80.5	0~1 (avg = 0.79)
10	79.8	0.18~0.78 (avg = 0.67)
100	77.1	0.37~0.67(avg = 0.55)
200	66.4	0.4~0.56 (avg = 0.5)
400	59.7	0.39~0.58 (avg = 0.5)
478	59.5	0.41~0.56 (avg = 0.49)

$M = 50\%$ of the feature set (shown in Figure 5.3). Compared to the previous work [72], the results show that sparse representation for classification using LOO increases the overall performance accuracy by about 10%.

Figure. 5.2 illustrates an example of sparsity of the sparse coefficients. As can be seen from the graph, many of samples are very close to zero while there are some non-zero values as well.
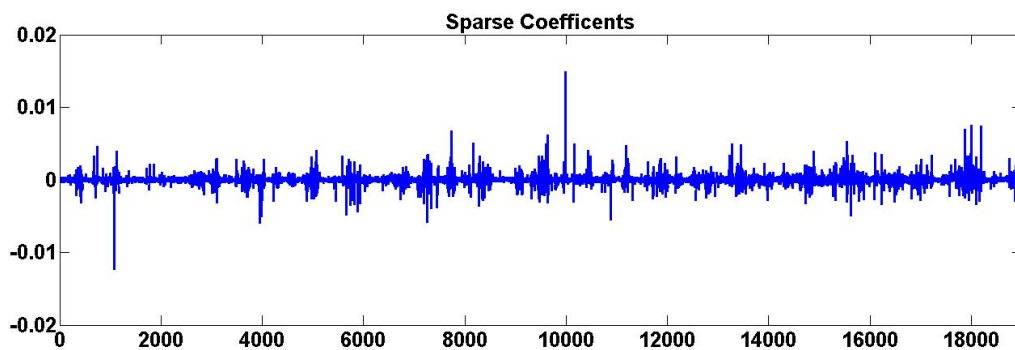
Figure 5.2: An example of sparsity and sparse coefficients.

114

(a) Feature space using AR coefficients.

(b) Feature space using cepstrum coefficients.
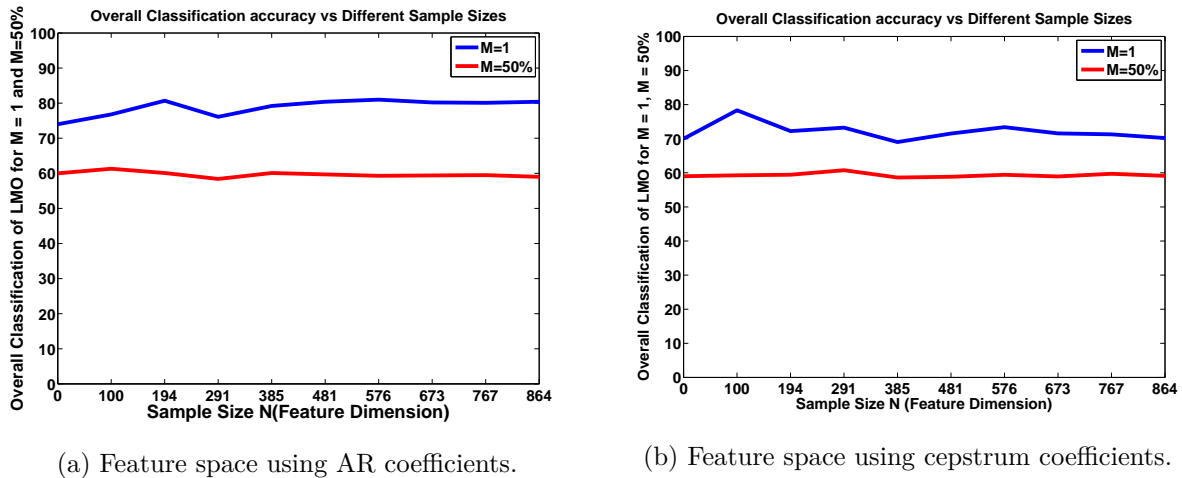
Figure 5.3: **Overall Classification of Leave-M-Out** Vertical axis: overall classification for $M = 1$ and $M = 50\%$ of the feature space. Horizontal axis: different segment length using (a) AR coefficients and (b) Cepstrum coefficients as feature space.

5.4.2   Evaluation Assessment on Surrogate Synthetic Signal

In addition to analyzing real biomedical signals, we tested the algorithm using synthetic non-stationary signals. The proposed SRC technique with the sparsity measures facilitates the adaptation of the LOO method to any of the signals available in automating the identification process of the SRC application. The surrogate synthetic signals consist of two non-stationary signals with spikes and signals with different local correlations, and two stationary signals with linear and power-law trends. Figures 5.4 to 5.7 show these signals respectively.

We then calculated the SRC performance for different M in addition to calculating the sparsity of these signals. Table 5.5 summarizes the results we achieved for different scenarios.

From the results achieved, it is evident that sparsity has a direct link to classification performance. If sparsity is properly harnessed, the choice of features is no longer critical

115
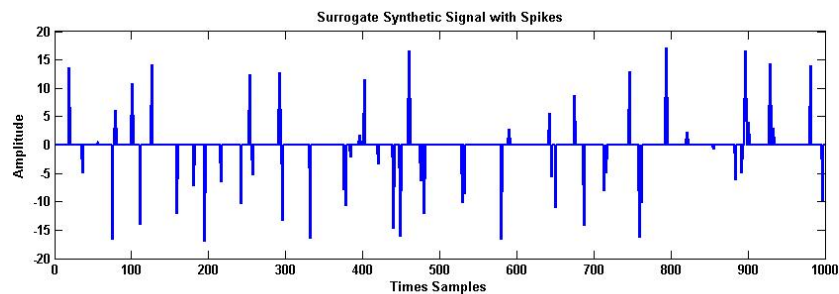
Figure 5.4: Non-stationary synthetic signal composed with spikes probability $p = 0.05$, Amplitude $A_{sp} = 10$.
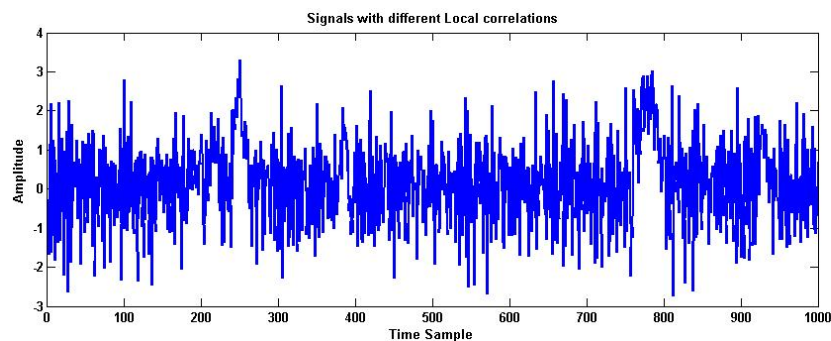
Figure 5.5: Non-stationary synthetic signal with different local correlations composed of mixed signals.
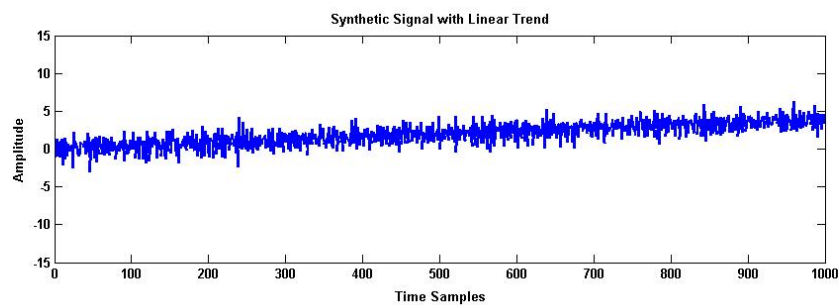
Figure 5.6: Surrogate signals with linear trends. $A = s^{-8}/index$.

i.e., the number of features (however large they may be) does not have a direct correlation with performance as long as the sparseness criterion is satisfied. This is true for both real

116
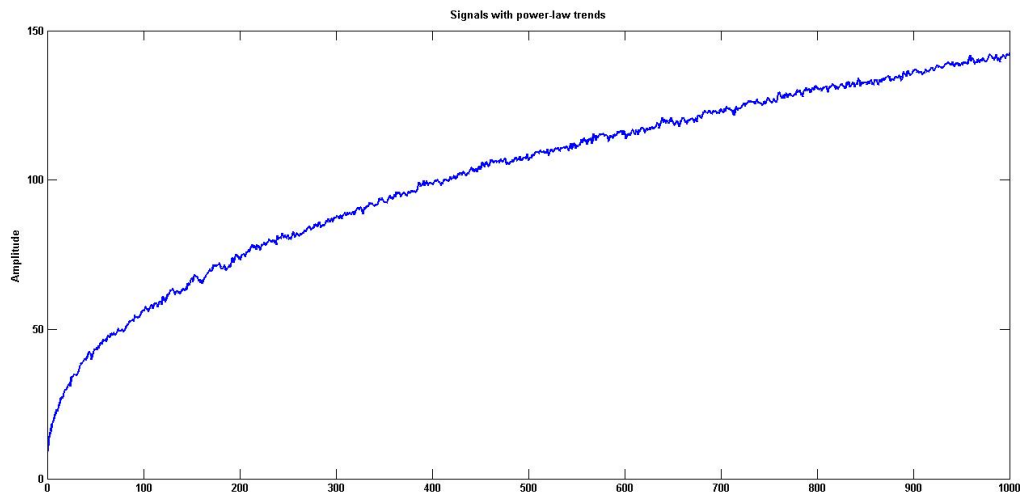
Figure 5.7: Surrogate signals with power-law trends.

Table 5.5: The performance accuracy of the SRC algorithm and the measure of sparsity for these signals.

Signal	M=1	M=50%	Sparsity Measure
Linear/Power-law	75%	62.6%	0.57
Spikes/Correlations	52%	45%	0.39

and synthetic datasets. For those that are sparser compared to other ones, the classification performance seems to be higher.

5.5   Chapter Summary

Saying a signal is sparse means that its energy is concentrated in some transform domain. A measure of sparsity could be used to search for such a domain. In this chapter, the goal was to measure the sparsity of the coefficients. We later found that our proposed measure of sparsity not only resulted in faster convergence, in comparison to Gini index that needs

to sort the samples prior to analysis, but could also lead to better classification performance as long as the value of that measure is close to the best sparsity. In addition, we applied our proposed measure to both real and synthetic signals. In the next section, we analyze different designed dictionaries to evaluate and enhance the performance of our classification algorithm.

Chapter 6

Dictionary Design Algorithms

Screening of Sleep Movement Disorders Via EMG Signal Analysis

Chapter 2

Chapter 3

Polysomnography

EMG Signal Analysis

Known-Algorithms

Proposed Algorithm

Chapter 4

Chapter 5

Chapter 3 Cont'd

Sparse Representation

Measure of Sparsity

Modeling Algorithms

Wavelet Transforms

Chapter 6

Chapter 7

Fixed Dictionary

Adaptive Dictionary

Kernel Sparse Representation for Classification

T HIS chapter presents the development of alternative dictionary design algorithms that have been shown to have a link between sparsity in the dictionary and sparsity in decomposition. Sparsity in the dictionary allows for faster and more efficient dictionary learning algorithms, whereas the sparsity in decomposition provides a sparser representation and a decomposition that matches the signal features well.

6.1   Motivation

Sparse and redundant representation modelling of data requires an ability to facilitate signals as linear combinations of a few atoms from pre-defined dictionary elements. As such, it is crucial to choose a dictionary that sparsifies the signals. In general, choosing a proper dictionary is done in two ways: i) using predefined mathematical models of the data to build a sparsifying dictionary, or ii) learning which dictionary performs best on a training set. In this chapter, a novel framework for dictionary learning based on NMF is presented, with an application to signal classification. We first describe the evolution of these two hypotheses, in addition to our proposed algorithm for dictionary learning.

6.2   Dictionary Algorithms

The process of digitally sampling a signal is to represent a function as the sum of delta functions in space or time. While this representation is suitable for visual display, it is mostly ineffective for analysis tasks. In signal processing in general, depending on the task, more meaningful representations should be used that can capture the useful characteristics

of the signal. For example, for classification algorithms, the representation has to highlight more prominent features, whereas for compression, the representation should capture a large part of the signal with only a few coefficients. To represent a signal, one must choose a dictionary, which is the set of elementary signals or atoms used to decompose a given signal. If the dictionary elements form a basis, then all the signals can be uniquely represented as a linear combination of the dictionary atoms. In case of an orthogonal dictionary, which forms the simplest case, the representation tends to be computed as the inner products of the signal and the dictionary atoms. In the non-orthogonal case, however, the coefficients are constructed by the inner products of the signal with the inverse of the dictionary atoms, which is often referred to as the bi-orthogonal dictionary. Although the orthogonal and bi-orthogonal dictionaries are mathematically very simple and have been extensively used by researchers they are considered as weak dictionaries [67]. This has led researchers to develop newer over-complete dictionaries that have more atoms than the dimensions of the signal and could lead to a wider range of signal representation. Although research on over-complete dictionaries has been commenced in the past decade [67, 119, 146], certain open problems still exist. Such dictionaries have the potential to introduce an ambiguity in the definition of a signal representation. Consider the dictionary $A = [A_1 A_2...A_n]\epsilon\mathbb{R}^{m \times n}$, where the columns contain the dictionary atoms and $n >> m$. Representing a signal $y\epsilon\mathbb{R}^m$ using the elements of this dictionary can take two paths: the *analysis* path or the *synthesis* path [67]. In the analysis path the signal is represented via its inner products with the atoms as follows:

$$\mathbf{x} = A^T\mathbf{y} \tag{6.1}$$

whereas for the synthesis path, the signal is represented as a linear combination of the atoms,

$$\mathbf{y} = A\mathbf{x} \tag{6.2}$$

If the dictionary is complete (i.e. $m = n$), the analysis and the synthesis definitions coincide for bi-orthogonal dictionaries. In other general cases, the two definitions differ dramatically. However, when A is over-complete, the family of representations $\mathbf{x}$ is actually infinitely large. This allows us to find the most informative representation of the form

$$\text{minimize} \ \ \|x\| \ \ \text{subject to} \ \mathbf{y} = A\mathbf{x} \tag{6.3}$$

Eq.6.3 promotes sparsity of the representation, which in turn sorts the coefficients to decay more rapidly. Thus, Eq.6.3 is often referred to as *sparse coding.* As mentioned in the previous chapters, sparsity can be solved by various cost functions such as the ℓ^p with $0 \leq p \leq 1$. Although the options of 6.1, 6.2, and 6.3 have been extensively studied, depending on the algorithm, choosing the proper dictionary for a given task must be addressed. Earlier works made use of the traditional dictionaries such as wavelet transform and Fourier dictionaries, which are well suited for 1-D stationary signals. However, for more complicated signals, in particular the natural biomedical 1-D signals, these dictionaries are not well equipped. This is due to the complexity such as, non-stationarity and non-linearity of such signals.

Different types of dictionaries have been developed in response to the growing need; these have emerged from one of two sources: a mathematical model of the data, or a set of realizations of the data. Dictionaries of the first type require choosing pre-existing mathematical

models. These dictionaries could be constructed using the Fourier basis, the wavelet basis, or the modified discrete cosine basis. As well, they can be constructed from a redundant or over-complete dictionary by forming a union of bases (for example the Fourier and wavelet bases) while allowing a representation of different properties of the signal [129, 130]. Dictionaries of the second type tend to capture the specific features of the signal or a set of signals to update the elements of the dictionary atoms. Dictionary learning methods are often based on an alternating optimization strategy, in which the dictionary is fixed and a sparse signal decomposition is found; then the dictionary elements are learned while the signal representation is fixed [129, 131]. Most recently there has been growing interest in the second type, which can deliver increased flexibility and has the ability to adapt to a specific signal type. This is the motivation for describing the evolution of these dictionary types and our proposed algorithm.

6.3   Elements of Modern Dictionary Design

In the 1980s, statisticians saw the rise of a new powerful approach, widely known as robust statistics, which advocates sparsity as the key to tasks such as recovery and analysis. This led the researchers to seek sparser and more efficient representations, which had to be done by shifting from the linear model towards a more flexible non-linear formulation. To achieve the best approximation in a non-linear model, each signal is allowed to use a different set of atoms from the dictionary elements. Thus the approximation is as follows:

$$y \approx \sum_{n \epsilon I_K(y)} c_n \phi_n \tag{6.4}$$

where $I_K(y)$ is an index set adapted to each signal individually. This led the researchers to design newer, more efficient transforms in addition to developing fundamental concepts guiding modern dictionary design algorithms. One of the key concepts in achieving sparsity in modern dictionary design is to allow for better localization. Atoms with higher sparsity agree to more flexible representations based on the local signal characteristics, which in turn limit the effects of irregularities observed in large coefficients [67]. One of the very first algorithms designed, achieving the sparsity property, is the Short Time Fourier Transform (STFT), which is the natural extension of the Fourier transform [147]. In STFT, the Fourier transform is applied locally to portions of the signals that are possibly overlapping. This reveals a time-frequency description of the signal. The Gabor transform was another form of STFT suggested by Dennis Gabor back in 1940 [148]. This transform was rediscovered by researchers in around 1980 [149].

A basic 1-D Gabor dictionary consists of windowed waveforms as follows:

$$G = \{\phi_{n,m}(x) = w(x - \beta m)e^{i2\pi\alpha nx}\}_{n,m\epsilon\mathbb{Z}} \tag{6.5}$$

In this equation, $w(.)$ is a low-pass window function which is typically a Gaussian localized at 0. α and β are the time and frequency resolutions of the transform. These works were extended to higher dimensional signals with more complex Gabor structures, which had directionality property by means of varying the orientation of the sinusoidal waves [150]. This concept has been extensively used in fields of visual cortex and image processing

tasks. Today, the Gabor transform is mainly applied in analysis and detection tasks, as a collection of directional filters. Another main conceptual advancement was achieved with the rise of multi-scale analysis. It has been shown that natural signals exhibit meaningful structures over different scales. In addition, they can be analyzed and described efficiently by multi-scale constructions [67]. To achieve these properties, researchers had to abandon the orthogonality property for over-completeness. This allowed them to have an efficient number of atoms in an orthogonal transform and led to a fundamental move from transforms to over-complete dictionaries for sparse signal representations. This brought with itself a diverse set of applications, such as signal and image compression, de-noising and face recognition [151,152]. In addition, considerable research for over-complete dictionary design was developed in the context of signal and image classification [96,153–155]. All these works use sparse representation to formulate a signal classification problem given over-complete dictionaries. These over-complete dictionaries first merged the existing transforms such as the Fourier, wavelet, and STFT, which are the unions of several similar bases. They provide guidelines to increase the variety of features representable by the dictionary. In the next section, we will explain the dictionaries that were modelled by a more simple class of mathematical functions and our proposed sparse representation for classification results using these simple mathematical models as elements of the dictionary. We will then focus sparse representation that is able to characterize locality relations. These mathematical models can convey valuable information for classification [104]. One of these algorithms is NMF, which has been extensively used for localization. For natural signals, specifically sleep signals that are long in duration, detecting an event that happens in a very short period of time is a tedious task and requires extensive signal processing algorithms. NMF is one of

the decomposition techniques that has become popular in research for extracting features in machine learning, computer vision, and signal processing [156].

6.3.1 Autoregressive Modelling

Representing the signal with fewer coefficients was a major driving force for the development of more efficient representations. In the late 1960s Akiake used AR models to estimate the spectral characteristics of time series [68]. This soon became the common approach for researchers to identify noisy feedback systems, which could not be solved by direct application of frequency approaches [68]. This approach was even used for feature extraction techniques in late 1980s and early 1990s. AR modelling is an all-pole model that provides parameters which are correlated with the physiological system producing it. The general form of AR modelling is as follows:

$$x(n) = -\sum kPa_k x(n-k) + Gu(n) \qquad (6.6)$$

where a_k are the AR coefficients and P is the model order. In evaluating this model, two critical parameters have to be considered. First, the model order has to be selected such that it is sufficient to replicate the signal. Second, the signal has to be stationary in order for AR coefficients to be effective.

In this work, we computed the classification rates with the feature space dimensions of 956×26. This feature space was originally calculated from [72]. The dataset used in this work consisted of signal segments from 8 chin EMG signals (4 with normal behaviour and

126

4 with elevated muscle tone) in REM sleep from participants who had undergone the sleep test. Since the EMG signals have non-stationary behaviours, the signals were fed to an adaptive signal processing algorithm to partition the signal into stationary segments. Each segment was then used to calculate the AR coefficients of model order 26. The order 26 was calculated using

$$AIC(P) = \ln \hat{\sigma}_{wP}^2 + 2P/N \qquad (6.7)$$

where $\hat{\sigma}_{wP}^2+$ is the estimated variance of the linear prediction error. As the order increases $\hat{\sigma}_{wP}^2$ decreases and therefore $\ln \hat{\sigma}_{wP}^2$ decreases. However, $2P/N$ increases by increasing P. Yet, at some value of P, the minimum value could be evaluated [8, 64, 157].

The degree of sparsity for all sparse approximations of this feature set is calculated using Eq. 5.5. From this Table, it is apparent that the generalization scheme in practical applications strongly depends on the value of M. It is also obvious that sparsity measures reduce as we increase the value of M; the higher the value of M, the lower the accuracy. Although the LOO cross-validation method generally incurs a high computational cost, it is the least biased estimate, since rigorous validation is performed compared to LMO cross-validation, which exhibits a comparatively higher variance. In this case, the algorithm reaches its maximum learning capacity (stability) by the LOO method, since the perturbation induced by LOO is small and therefore the classifier is stable. As we increase the size of the perturbation, stability is less likely to hold, and the sparse approximation is less sparse; thus, the accuracy is reduced by about 20%. To prove the generality of the algorithm, we analyzed it using different numbers of segments. This is shown in Figure 6.1a, where we used 10%, 20%,...100% of the segments and applied LMO for both $M = 1$ and $M = 50\%$ of the original

(a) Feature Space using AR coefficients

(b) Feature space using cepstrum Coefficients
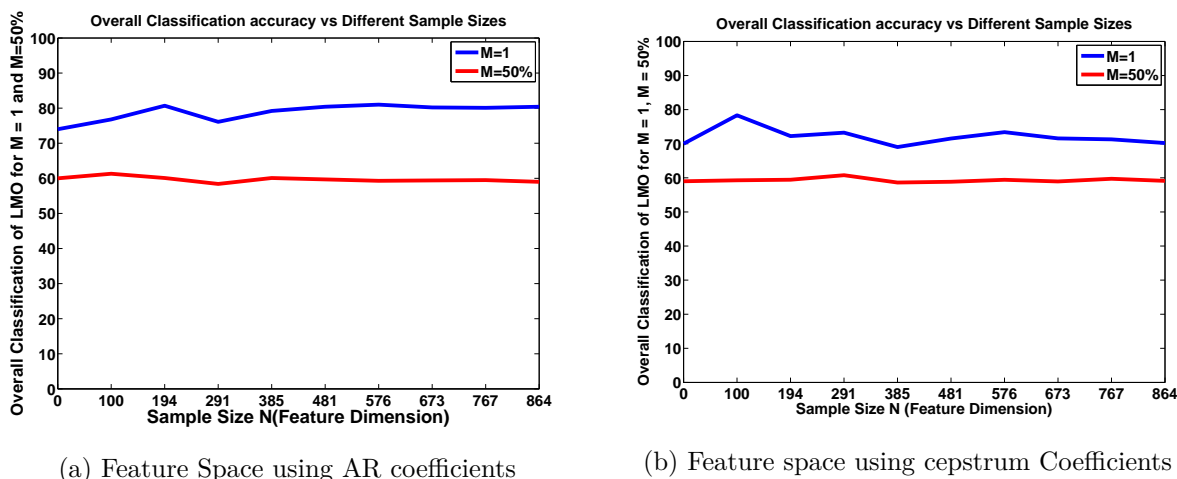
Figure 6.1: Overall Classification of Leave-M-Out Vertical axis: overall classification for $M = 1$ and $M = 50\%$ of the feature space. Horizontal axis: different segment length using (a) AR coefficients and (b) cepstrum coefficients as feature space.

feature space. The overall performance of all the cases when $M = 1$ is about 80.5%, however this value reduces by about 20% to 60% as we increase M. The same analysis was applied to a different random feature of cepstrum coefficients [72] with overall average accuracy of 72% for $M = 1$ and a 10% reduction when $M = 50\%$ of the feature set (shown in Figure 6.1b). The results compared to the previous work [72] shows that sparse representation for classification using LOO increases the overall performance accuracy by about 10%. The overall MATLAB implementation of the algorithm takes only few seconds on a typical 3 GHz PC.

6.3.2   Non-negative Matrix Factorization (NMF)

Another procedure we used to evaluate our sparse representation for classification by means of using existing mathematical models is NMF. There are many matrix decomposition techniques available including Independent Component Analysis (ICA), Principal Component

Table 6.1: The classification accuracy of AR coefficients fed into a sparse representation algorithm for different M of Leave-M-out as well as the degree of sparsity of sparse coefficients.

M	Accuracy of AR Coefficients (%)	Sparsity of the Sparse AR Coefficients
1	80.5	0~1 (avg = 0.79)
10	79.8	0.18~0.78 (avg = 0.67)
100	77.1	0.37~0.67(avg = 0.55)
200	66.4	0.4~0.56 (avg = 0.5)
400	59.7	0.39~0.58 (avg = 0.5)
478	59.5	0.41~0.56 (avg = 0.49)

Analysis (PCA), SVD, and NMF. These have become an active area of research for extracting features in machine learning, computer vision, and signal processing [156]; however, depending on the application and the data, different results have been reported. Some works demonstrated the advantages of ICA over NMF and PCA. Lee *et al.* compared the result of PCA, NMF and ICA for feature extraction from multiple video frames [158] and showed that ICA-based features result in better video representation than the PCA- or NMF-based features. In another study the authors compared the techniques and showed that ICA method outperforms the other methods in occluded face recognition application [159]. In another study, the authors demonstrated that NMF is more stable for larger basis compared to ICA and PCA [160].

With respect to the application used, NMF compared to ICA and PCA has more advantages, as listed below:

- NMF is applied to a non-negative matrix and constrains the matrix factors W and H, which are also non-negative;

- NMF decomposed factors promise a higher Time-Frequency (TF) representation and

localization;

- NMF codes naturally favour sparse, parts-based representations which in the context of recognition can be more robust than non-sparse, global features [156].

The problem with this technique is that the extracted feature vectors have a very high dimension. This is because the length of each feature vector is proportional to the signal's sampling frequency, and as a result they are not very appealing for classification. However, some variations of NMF is subject to additional constraints that allow particularly accurate control over sparseness and, indirectly, over the localization of features.

In general, if the matrix is of the form $V_{M \times N}$ the decomposed signal is $V_{M \times N} = W_{M \times r} H_{r \times N} = \sum_{i=1}^{r} w_i h_i$, where V is the TF matrix, W is the base and H is the coefficient vectors and r is the rank of the NMF decomposition. This decomposition is such that W represents the signal structure and the H represents the location of the corresponding base vectors in time. The algorithm that was used in this work is as follows [106]: let A be the matrix each column of which are the training samples from the ith class. The elements of this matrix are a set of feature vectors that have been produced from transforming the signal into TF plane and from there to an NMF algorithm with a rank of $r = 50$. y is the new testing samples, in which a linear representation of y can be rewritten in terms of all training samples as $y = Ax$. We again use cross validation to randomly split the data into distinct matrix for training and testing. Once the system is trained with the training signals we test it using the ℓ_1 minimization problem for sparse representation as in Eq. 6.8

$$\hat{\mathbf{x}}_1 = argmin \|\mathbf{x}\|_1 \quad \text{subject to} A\mathbf{x} = \mathbf{y} \tag{6.8}$$

6.3.3   chin EMG Dataset

6.3.3.1   Algorithm

In this section, we present a method for classifying non-stationary time series utilizing a feature set based on TF representation followed by NMF. This data was further analyzed using sparsity measure and sparse representation for classification. The proposed scheme for real biomedical signal analysis and discrimination can be briefly characterized by the following steps and the block diagram of the proposed method is given in Figure 6.2:
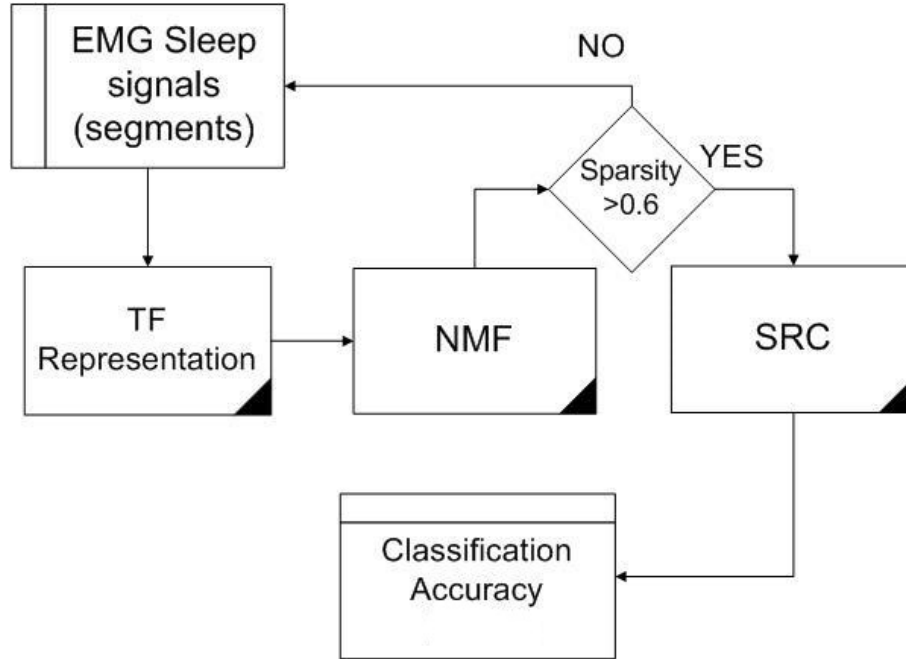
Figure 6.2: Block Diagram of the Proposed method

Preprocessing: In this study we utilized the sleep dataset that was provided to us by the Sunnybrook Health Sciences Centre. The sleep dataset consisted of signal segments

from 8 chin EMG signals (4 with normal behaviour and 4 with RBD from patients who had undergone the sleep test. A traditional scoring system for sleep was established [10], with the electrophysiological parameters of EEG, EOG and EMG. The system used for recording chin EMG signals during sleep included 3 relatively midline electrodes, one above the jaw line, one below the jaw line and one back-up electrode (shown in Figure 1.9). The two electrodes were typically subtracted from another to eliminate artifacts shared by both electrodes. In this study, a subject was defined as historically normal if there was no history of any violent behaviour, including sleep talking, shouting, screaming, hitting or punching, or flying out of bed during the night sleep; otherwise, he or she was considered as abnormal.

Higher Dimensional Mapping: The preprocessed signal was then mapped to the TF representation using the wavelet scalogram, and a transformed signal was obtained. The frequency resolution of this transformed function was 150 (scale is from 1 : 150) and the time samples consisted of 1000 elements. The advantage of using wavelet scalogram is that it uses an adaptive varying time width defined by the scaling parameter. Another important advantage of the scalogram is that it provides positive and cross-term free TF representation.

Feature Extraction: The transformed signal matrix was then decomposed into two matrices, W the bases matrix and H the coefficient matrix. In this work we used W as the feature vector since it is good for local feature extraction and classification. The rank of the matrix was chosen to be $r = 50$, as the results show that NMF yields the lowest recognition rate for decomposition dimensions of less than 50 ($r < 50$). Because of the high dimensionality of the NMF, we proposed using the degree of sparseness. Sparse NMFs are useful when the degree of sparseness in the non-negative basis matrix or the non-negative coefficient matrix in an NMF needs to be controlled when approximating high-dimensional data in a lower

dimensional space. This process is based on degree of sparsity. If the measure of sparsity is more than 60% the segmented sleep signal will discarded, else that segment will be used for classification.

Classification: Based on the feature space, the normal and abnormal RBD classes were defined. The feature vectors along with the class definition together were fed into sparse representation for classification. In this step the influence of both training and testing sample size on the design and performance of pattern recognition systems was investigated by using sparse representation incorporated with the LOO approach. As stated, this method is believed to be one of the most optimized validation approaches with least biased estimate [109, 124]. The presentation of the underlying theory has been complemented with examples with of real time sleep signals, and results of the application of these data were validated based on the degree of sparsity measure and classification accuracy.

For comparison, we have compared our proposed method with the other previous methods [87, 106]. This has been shown in Figure 6.3. Although our proposed method is unstable when leaving half of the samples out ($M = 50\%$), it shows a better classification accuracy for LOO.

With the demand for obtaining and describing even more complicated and complex phenomena, the urge to design algorithms from which the optimal representation can be constructed has increased. Recent research in sparse approximation involves extension of these techniques towards obtaining optimal signal representation, based on the data itself, which leads to optimal signal discriminations. To achieve this, researchers have proposed obtaining robustness by adapting the dictionary atoms themselves to the signal content. Dictionaries
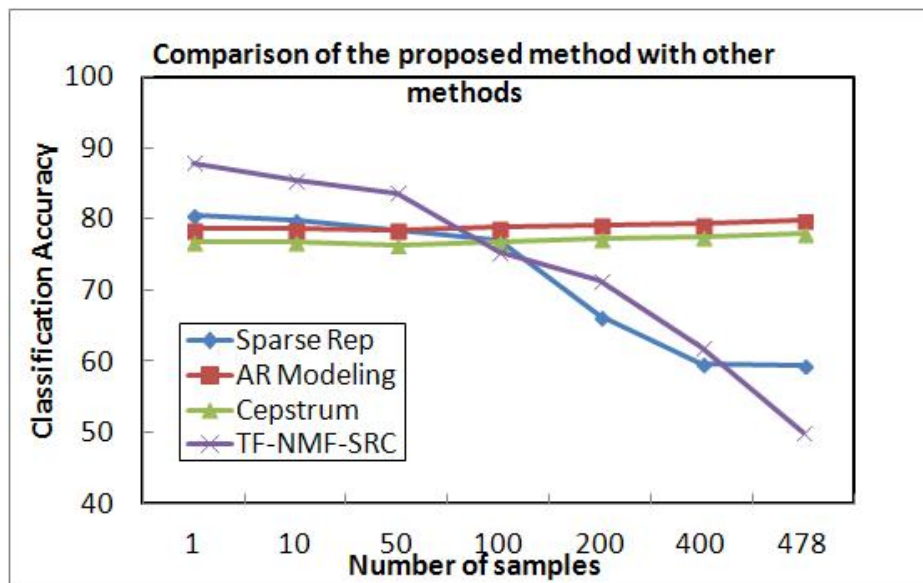
Figure 6.3: Comparison of our proposed method using the non-negative matrix factorization and other methods. This graph shows that our proposed algorithm give the highest accuracy possible when using NMF dictionary based algorithms combined with the LOO algorithm.

that are mainly trained from very large sets of examples are found to accelerate convergence and improve training results. One of the main advantages of trained dictionaries is that they lead to distinctive results in many practical signal processing applications. This led Engan *et al.* in 1999 and Aharon *et al.* in 2005 to efficiently train a generic dictionary for sparse signal representation called Method of Optimal Directions (MOD) and K-SVD respectively [161,162]. In the next section, these algorithms are explained. It is worth mentioning that throughout this dissertation, we refer to dictionary selection as A and dictionary learning algorithms (adaptive dictionaries) as D.

134

6.3.4   Method of Optimal Directions

MOD is one of the very first sparsification processes that attempted to find a dictionary D and sparse matrix x that minimizes representation error:

$$\text{argmin}||y - Dx||_F^2 \quad \text{subject to } ||x_i|| \leq T \tag{6.9}$$

γ_i represents the columns of x, and ℓ^0 sparsity counts the number of zeros in the representation. This optimization problem consists of 2 steps: sparse coding and dictionary updates. Using any standard technique, the sparse coding step is performed on the individual signals, and the dictionary updates use the analytic solution of quadratic problems which involves the Moore-Penrose pseudo-inverse [67, 163]. Although MOD requires only a few iterations to converge, the relatively complex inversion matrix does not provide an efficient method, and other works had to be implemented to produce more efficient and less complex algorithms.

6.3.5   K-SVD Algorithm

Similar to MOD, K-SVD is intended to efficiently train a generic dictionary for sparse signal representation using sparse coding and dictionary update steps; the difference is in the updating of dictionary elements. Unlike MOD, which uses an inversion matrix to update the dictionary elements, K-SVD is performed atom by atom in a simple and efficient process [67, 131]. The K-SVD algorithm takes its name from the SVD process that repeatedly

updates the atoms K times. For a given atom k, the quadratic term is of the form:

$$||y - \sum d_j x_j^T - d_k x_k^T||_F^2 = ||E_k - d_k x_k^T||_F^2 \qquad (6.10)$$

where x_j^T are the rows of X, and E_k is the residual matrix.

Although K-SVD provides a faster algorithm compared to the MOD algorithm by updating both the current atom and the sparse coefficients simultaneously, K-SVD suffers from few weaknesses, such as high non-convexity. To overcome this problem Aharon *et al.* proposed the non-negative K-SVD. For many applications, using dictionary learning algorithms and sparse representations simultaneously with non-negativity constraints on both the dictionary and the sparse coefficients can lead to revealing the behaviour of all training signals. This in turn results in sparser dictionary elements that lead to faster and more robust convergence [162]. Their algorithm was based on a variation of Basis Pursuit and allowed the non-negativity constraints on the coefficients under the least square sense [131, 162]. Although this algorithm produces faster convergence, in order to reach a local minima an iterative technique is required, which produces the same complexity as the original SVD step. This leads to our next contribution in this dissertation: designing an adaptive dictionary that uses sparsity in both the dictionary and the decomposition to create a fast and efficient dictionary learning algorithm. This algorithm is similar to K-SVD, except that it uses NMF for matrix decomposition instead of SVD. The next section explains our proposed algorithm in more detail with a comparison of the NMF and SVD algorithms depicted in Table. 6.2.

6.3.6   K-NMF Algorithm

Similar to K-SVD, K-NMF is an iterative algorithm that alternates between sparse coding of the examples based on the current dictionary atoms to better fit the data. In this algorithm the update of the dictionary columns is combined with the update of the sparse representation, which results in accelerated convergence. Basically the K-NMF consists of two parts: first it updates the non-negative dictionary elements, and then it calculates the sparse coefficients using the NMF decomposition algorithm. The K-NMF algorithm is comparably faster than the K-SVD, and it could be used for local feature extraction and classification. For this algorithm, a typical sample has a sparse representation with respect to such (possibly learned) bases. Generally, this type of dictionary is well localized, is well adapted to the signal's local structure, and provides an interpretation of the inherent non-stationary structures of the signal. Additionally, extracting regularities and structural patterns from the EMG using NMF will project all the EMG signals that have the same spectral shape onto a single basis. This is because that the EMG's mixing model is defined in the magnitude spectrum, and magnitude spectra have phase-invariant properties. This will train the dictionary elements and lead the training signals to have sparse representation, which can be further used in classification. The K-NMF algorithm could be well adapted to the signal's local structure, and it provides an interpretation of the inherent non-stationary structures of the signal. This will allow us to train the dictionary elements based on data which leads to better discrimination performance.

To design the optimum dictionary for EMG signals, we propose using sparse representation and NMF. One reason that NMF is a powerful tool for extracting regularities or structural

patterns from EMG signals may be that the mixing model is defined in the magnitude spectrum domain. Because of the phase-invariant nature of magnitude spectra, NMF is able to project all signals that have the same spectral shape onto a single basis. This allows us to represent a variety of EMG phenomena efficiently using a very compact set of spectrum bases.

The details of this algorithm are as follows:

Input: TF representation matrix $\{V_i\}_{N \times M}$ where N and M are the time and frequency resolutions with d number of dictionary atoms

Initialization: Random dictionary matrix $D^{(0)} \epsilon \Re^{n \times m}$. Set $J = 1$.

Procedure: Repeat until convergence (or maximum number of iterations)

1. **Sparse Coding Step:** For each $i = 1, \ldots, n$

   $$\min_x \|\mathbf{y} - D(\hat{\mathbf{x}}_1)\|_2^2 \text{ s.t. } \left\|\mathbf{x}^{(i)}\right\|_0 \leq T$$

2. **Dictionary Update Step:** for each $m = 1, \ldots, M$

   - $w_k \leftarrow \{i | 1 \leq i \leq n, \ \mathbf{x}_k^{(i)} \neq 0\}$

   - Compute $\mathbf{E}_k = \mathbf{Y} - (\mathbf{DX} - d_k \mathbf{x}^k)$ where, $\mathbf{d}(i) = \begin{Bmatrix} 0 & \mathbf{u}_1(i) < 0 \\ u_1(i) & Otherwise \end{Bmatrix}$,

     $\mathbf{x}(i) = \begin{Bmatrix} 0 & \mathbf{v}_1(i) < 0 \\ v_1(i) & Otherwise \end{Bmatrix}$
     $\mathbf{u}_1$ and $\mathbf{v}_1$ are the first singular vectors E_k.

   - restrict E_k by choosing only the columns corresponding to w_k, and obtain $E_k^{w_k}$

   - Apply NMF decomposition to $E_k^{w_k} = WH$ and update $d_k = W$ and $\mathbf{x}^k = H$

- $j \leftarrow j + 1$

Output: Element representation for adaptive non-negative dictionary D.

A part-based representation such as NMF is based on the low-rank approximation and will produce dimensionality reduction to the long data. Although many of these dictionaries are good for representation, classification and recognition, using them is somehow not feasible. Even though many algorithms have been proposed for face recognition applications, training dictionaries for classification of lower dimensional signals has only been partially explored. In other words, an algorithm that could possibly transfer the linear inseparable samples to a space where they become linearly/nonlinearly separable is needed. One possible solution is to incorporate kernel functions with sparse representation [164]. In order to achieve the best separability, one may need to modify sparse approximation methods such that the objective function is enhanced with a discrimination term that represents the separability properties of the signal.

Table 6.2: This table compares the two algorithms of SVD and NMF together. In general NMF reduces the dimension of the samples without orthogonal restriction, and clusters the correlated variables and describes more details of the dataset compared to SVD algorithm

SVD	NMF
$A = U_k \sum_k V_k^T$	$A = WH$
U is $M \times M$, $\sum$ is $M \times N$ and V is $N \times N$	W is $M \times k$ (basis matrix) and H is $K \times N$ (coefficient matrix)
• The left-singular vectors of A are eigenvectors of AA^* • The right-singular vectors of A are eigenvectors of A^*A • The non-zero-singular values of A (found on the diagonal entries of $\sum$) are the square roots of the non-zero eigenvalues of both A^*A and AA^*	Use any optimization method to calculate W and H, i.e. $\min f(W, H) = \frac{1}{2}\|A - WH\|_F^2$ such that $W, H \geq 0$
• Transferring correlated variables into a set of uncorrelated ones that better expose the various relationship amount the original data items • Identifying and ordering the dimensions along which data points exhibit the most variations. • Finding the best approximation of the original data points using fewer dimension	Factorization of matrices is generally non-unique, and a number of different methods of doing so have been developed (e.g. principal component analysis and singular value decomposition) by incorporating different constraints NMF differs from these methods in that it enforces the constraint that the factors W and H must be non-negative, i.e. all elements must be equal to or greater than zero
Unique Factorization	Decompose the signal into two factors with non-negative entries
Optimality property, Orthogonality	Sparsity and non-negativity
Speedy and Robust computation	Reduction in Storage
Applications: Gene expression Data Image Processing Information retrieval	Application: Data mining Dimension Reduction Clustering
Used mostly for reduction	Could be used for classification
Good for Global Feature extraction and classification	Good for local feature extraction and classification
	NMF achieves sparseness as a result of a side effect caused by non-negativity constraints

6.4   Application to EMG Signals

We carried out several experiments on natural EMG signals with the aim of showing the practicality of the proposed algorithm and the general sparse coding scheme. The training data was constructed following the TF-representation of different segment length size of 128, 256, 512 and 1024.

The TF matrix was fed to an adaptive dictionary, which was constructed based on the K-NMF algorithm. This method finds the best dictionary elements to represent the test sample $\{y_i\}_{i=1}^{N}$. The elements of the dictionary D were first generated at random with size of $n \times m$ with i.i.d. entries. The values of the dictionary atoms were later updated by every data signal using NMF algorithms. The updates were only applied on the non-negative data points and were moving forward for a maximum of 20 iterations. The sparse coding stage was based on linear programming to calculate the sparse coefficients. Please note that different performances could have been obtained if we had switched to other pursuit algorithms. We concentrated on linear programming for two reasons. First we wanted to be consistent throughout this dissertation on the choice of the pursuit algorithm; second, ℓ^1 algorithms are simple and have faster execution time. Using NMF instead of SVD allowed for less computational time as well as better localization, which is very beneficial in applications such as classification and recognition. A total of 65 subjects were used for this analysis. Each had a left and right leg recording for a total of 130 signals. The classification performance was evaluated similar to the previous chapters using LOO and LMO where M is total number of samples out for testing starting from 1 to 50% of the total sample length

size. The results are demonstrated in Figure 6.4.

Table 6.3: This table compares the performance time of the K-SVD and K-NMF algorithm in seconds. It can be seen that the performance time decreases for our proposed algorithm

Signal Size	K-SVD	K-NMF	K-SVD Performance	K-NMF Performance
128	120 s	115 s	71.8%	72.2%
256	245 s	238 s	73%	73.5%
512	310 s	298 s	73.1%	73.7%
1024	365 s	342 s	74.5%	75.9%

Table 6.3 tabulates the time it takes to execute both algorithms in addition to the performance of the two algorithm for comparison purposes. As can be seen from this table, the execution time is lower in K-NMF compared to K-SVD and the performance of K-NMF is slightly higher in comparison to K-SVD Figure 6.5 compares the performance of K-SVD and K-NMF algorithm for the same dataset with segment length of 1024. The classification is based on LOO.
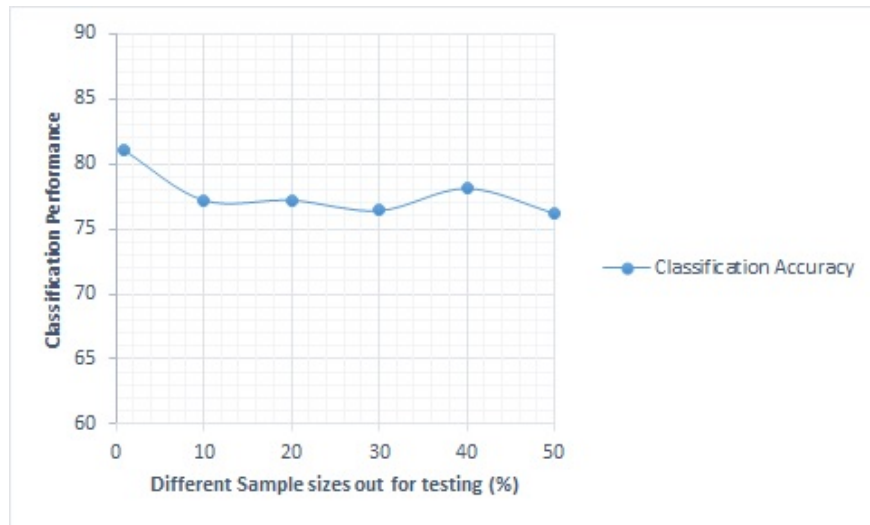
Figure 6.4: The classification performance while using the K-NMF algorithm for designing the dictionary. The x-axis represents the number of samples out for testing in terms of percentages.
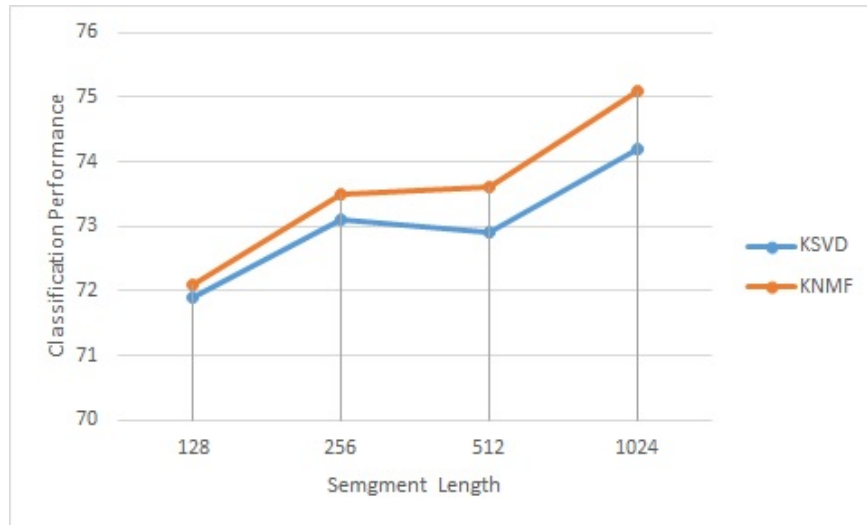
Figure 6.5: Comparison of K-NMF and K-SVD algorithm using the 130 leg EMG database. The K-NMF results into better classification performance compared to K-SVD.

6.5   Chapter Summary
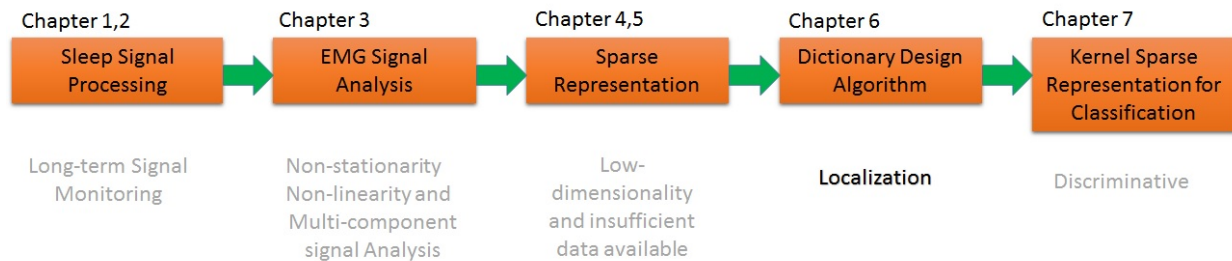
Figure 6.6: Flowchart of the proposed contributions

The contribution flowchart of this chapter is shown in Figure 6.6. This chapter has presented the development of an alternative dictionary design algorithm based on existing mathematical models and dictionary design algorithm. We have shown that there is a link between sparsity in the dictionary and sparsity in the decomposition. Sparsity in the dictionary

143

allows for faster and more efficient dictionary learning algorithms, whereas the sparsity in the decomposition provides a sparser representation and a decomposition that successfully matches the signal features. The focus of this chapter is also to promote NMF by means of sparsity and dictionary learning. NMF is a great algorithm for local features which in turn enhances the classification performance. In addition, NMF works well for non-stationary signals such the EMG signals in sleep. By the results achieved in this work, we showed that by means of NMF, more robust and consistent performance is produced for different numbers of sample length for testing and training, which makes this algorithm superior to those previously mentioned. However, to enhance performance and to achieve the best separability, the sparse approximation methods should be modified such that the objective function is enhanced with a discrimination term representing the separability properties of the signal. In the next chapter, we will explain our proposed algorithm for enhancing the performance in addition to presenting results from the two EMG datasets.

Chapter 7

KERNEL SPARSE REPRESENTATIONS

Screening of Sleep Movement Disorders Via EMG Signal Analysis

Chapter 2

Polysomnography

Chapter 3

EMG Signal Analysis

Known-Algorithms

Proposed Algorithm

Chapter 4

Sparse Representation

Chapter 5

Measure of Sparsity

Chapter 3 Cont'd

Modeling Algorithms

Wavelet Transforms

Chapter 6

Fixed Dictionary

Adaptive Dictionary

Chapter 7

Kernel Sparse Representation for Classification

I N this chapter, we describe how an advanced classification algorithm was carried out by means of a more sophisticated machine learning algorithm to achieve the best performance. In other words, to achieve the best separability, the sparse approximation methods have been modified such that the objective function is enhanced with a discrimination term representing the separability properties of the signal. For this reason, an algorithm is needed to transfer nonlinear, inseparable samples to a space where they become linearly separable. This is achieved by incorporating kernel functions with sparse representation.

7.1   Motivation

As shown in the previous chapters, EMG signal characteristics in sleep pose severe challenges in terms of the analysis and extraction of discriminant features. As explained in Chapter 3, EMG signals are non-stationary, non-linear, and have multiple components. These properties bring up the necessity to design novel algorithms that can discriminate and automatically detect an abnormality. In Chapter 4, we designed an algorithm based on sparse models to classify such signals. Although this classification algorithm resulted in high performance in comparison with other known algorithms for LOO applications, it did not have robust performance when the number of samples out for testing were more than 1. In Chapter 6 we explained how dictionary design performance will help make our proposed algorithm more robust by incorporating localization; nevertheless, the performance of the classification was lowered. To achieve a balance between the robustness and classification performance, in this chapter we aim to exploit "deep learning" and study the discriminant features of the signals by means of sparse representation to design a robust algorithm in

addition to high classification performance.

7.2   Machine Learning

Machine learning is a subset of artificial intelligence using intelligent algorithms to recognize, classify, and perform regression. These tasks are performed on a set of input patterns, which are then mapped to a given output or some category labels. Such types of machine leaning are trained to learn from experience and then generate probabilistic outputs or mapping patterns. Figure 7.1 demonstrates a typical machine learning algorithm procedure. This chapter aims to enhance the classification performance and increase the robustness of sparse representation algorithms for a complex biomedical signal such as EMG signal in sleep.

In many cases, extensive use of computers for solving complex problems is not an easy task. As a matter of fact, this process requires the use of certain domain-specified training samples that can give some information regarding the mathematical model functioning behind the behaviour of the complex systems. This constitutes the theory of machine learning. In machine learning and pattern classification, the decision function is the learning methodology applied to the target function and the learning algorithm is an algorithm that uses a set of training samples and maps them to the appropriate outputs. The machine learning that comprises of algorithms which are able to understand and decipher patterns from non-stationary signals using iterative multiple transformation with a tendency to adaptively change ways according to the signal characteristics is called "Deep Learning". Deep learning is about the knowledge of representing multiple levels of non-linear operations to help make sense of the complex data. This chapter aims to consider these approaches and apply them
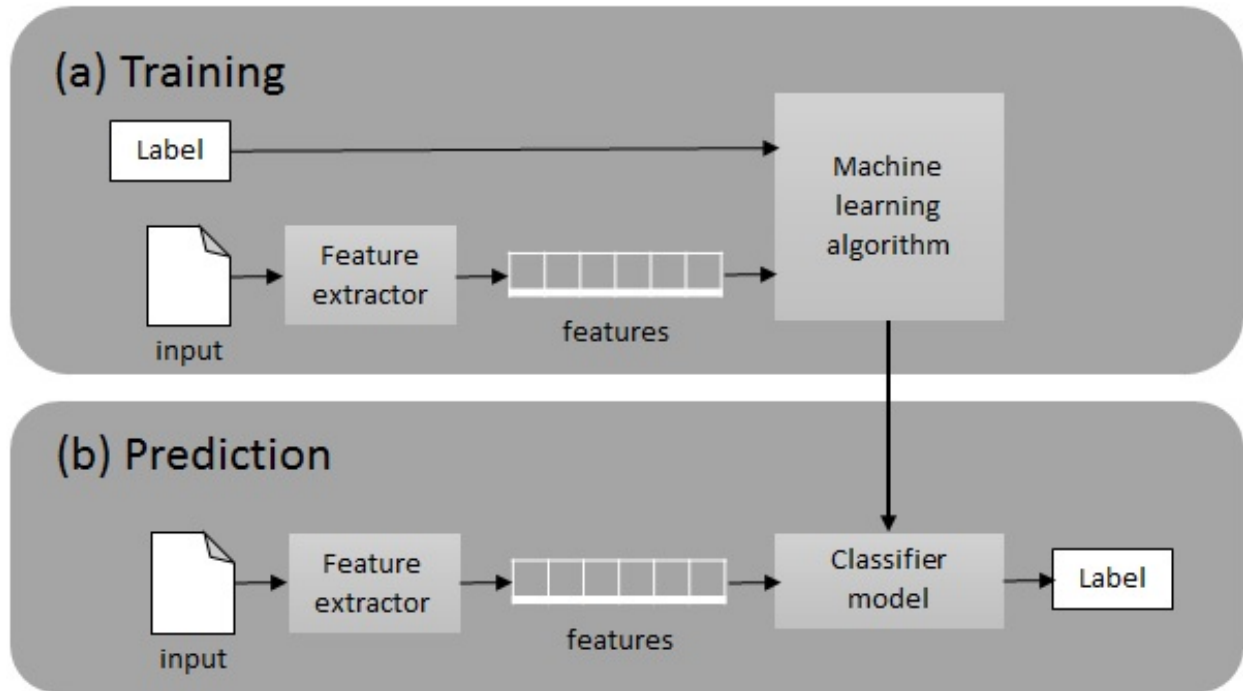
Figure 7.1: This graph shows a supervised classification. (a) During training a feature extraction is used to extract the features from the input signals and produce the feature set. This feature set will later be used to classify the signal. Pairs of feature sets and labels are fed into the machine learning algorithm to generate a model. (b) The same feature extraction scheme is used to predict the label for the unknown samples (testing samples).

on the complex EMG database in sleep. Figure 7.2 illustrates the structure of deep learning algorithms by using multiple levels for deciding on the data labels.
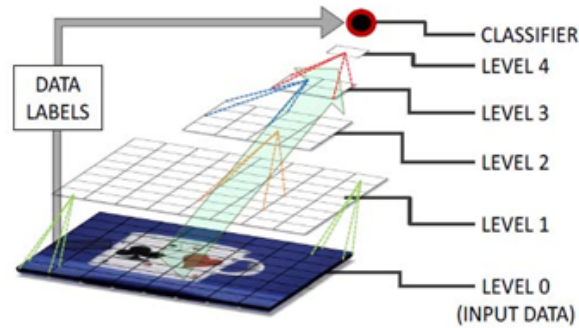
Figure 7.2: Deep learning structures

7.3   Kernel-based Machine Learning

To appropriately select and extract features from the data for pattern recognition applications, it is desirable to detect linear relationships in the input data. This, however, is not feasible for complex datasets involving random variations in feature space or the data itself. While applying machine learning methods, non-linear separation of classes occurs. To overcome this issue, kernels can be applied to the input data, that in turn transform the input to a higher-dimensional space with better separability of the classes. Kernels tend to map the non-linear higher-dimensional space to a space in which the separation of classes would be linear in feature space. Figure. 7.3 illustrates how kernel functions transform the data into higher dimensional space to make the separation of the output possible. Any kernel-based machine learning method employed for classification, includes two major steps [165, 166]:

- Mapping the non-linearly separable input data to linearly separable data (possibly by means of feature space) and,

- Learning algorithm applications that aim for finding patterns in the feature space.

149
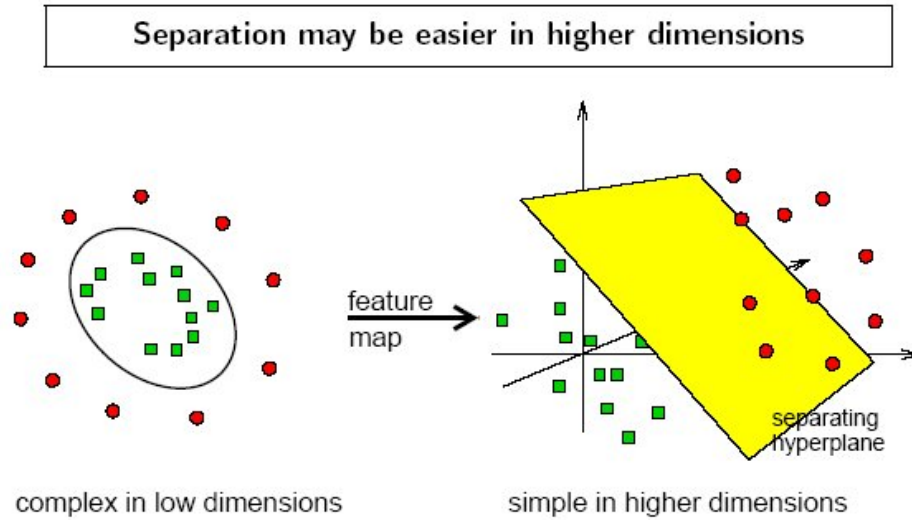
Figure 7.3: Kernel functions transform the data into a higher dimensional space to make separation possible.

In order to prepare the system for classification, the training set is constructed from the initial input data. Using the learning algorithm and the training set, the system learns to classify the data accordingly. To compute a kernel function, one must evaluate the inner product of the training samples. Let the training samples be of the form:

$$Y = y_1, y_2, ..., y_n \tag{7.1}$$

Then, using a kernel function ϕ, the input dataset Y is mapped into a higher dimensional space as follows:

$$Y = y_1, y_2, ..., y_n \longrightarrow \phi(Y) = \phi_1(Y), \phi_2(Y), ..., \phi_n(Y) \tag{7.2}$$

The mapping of Y can also be into a new space F which is known as *feature space* that allows the input features to be classified to the correct category label. In addition, some attributes of the input quantities can be considered in the feature space while performing classification tasks. Figure 7.4 illustrates how feature space can map the features to a space that the samples can be categorized.
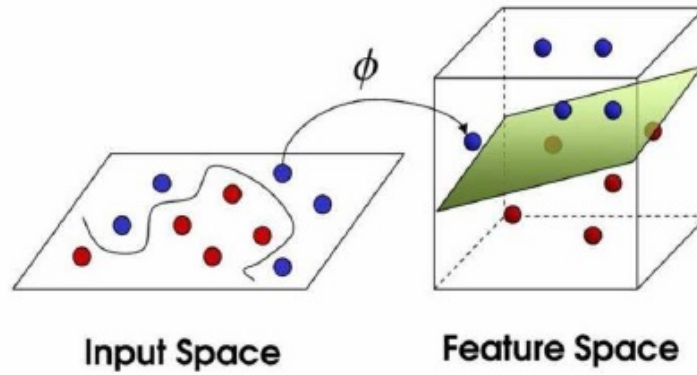
Figure 7.4: Transformation of the input data in to Feature Space, using the kernel function ϕ such that the samples are well separated

One of the benefits of kernel function is that it leads to dimensionality reduction of the input dataset in the feature space. In other words, by selecting only those features that are discriminatory out of a number of attributes, the input data can be classified in the best possible way. This is true since many of the features are redundant and can be eliminated for evaluating the performance. Therefore, a good feature selection involves not only dimensionality reduction, but also removal of irrelevant features. To formulate the pattern classification based on kernel functions let us assume the following:

$$f(Y) = \sum_{i=1}^{N} w_i \phi_i(Y) + b \tag{7.3}$$

In this equation, w_i is the weight associated with each of the input data points, b is the bias of the system and $\phi : Y \longrightarrow F$ is the non-linear mapping from the input space to the feature space. This, as well, shows that linear learning machines provide a dual representation for the inputs as follows:

$$f(Y) = \sum_{i=1}^{d} \alpha_i x_i < \phi(y).\phi(z) > +b \tag{7.4}$$

α_i is the transformed weight vector in the feature spaces, x_i is the output corresponding to each input data point and d is the number of training points. The kernel function K can be defined as $K(y,z) =< \phi(y).\phi(z) >$ for all $y, z \epsilon Y$. In other words, kernel functions can be used to implicitly map the data points into the feature space by defining the inner products between the input samples and the mapped features. The kernel matrix is constituted as follows:

$$K = (< y_i, y_j >)_{i,j=1}^{d} \tag{7.5}$$

There are different kernel functions that have a lot of popularity, such as linear kernels, Fisher kernels, and polynomial kernels. For this research we used Fisher kernels, which are widely used in generative, discriminative pattern classification and recognition applications [167–170].

7.4   Fisher Kernels

To assess the discriminative models in machine learning algorithms, the dependency of the unobserved random variable on the observed random variable is done by means of the conditional probability distribution function. The Fisher scores, which are the features

obtained after applying the Fisher kernels, will give the probabilistic model and show how the information is distributed of the data items within the model. The motivation in using Fisher kernels in this study are:

- Fisher kernels can be applied to different length time series signals

- Discriminative methods such as sparse representation for classification will produce better results when combined with Fisher kernels.

The next section will explain kernel sparse representation and our modifications.

7.4.1 Kernel Sparse Representation for Classification

This section will explain sparse representation for classification and our modification using kernel methods. SRC first encodes a query sample as a linear combination of a few functions from a pre-defined dictionary. It then identifies the class label by evaluating the class that results in minimum reconstruction error. The effectiveness of SRC is restricted by the important assumption that data points from different classes are not distributed along the same radial direction. Otherwise, this approach would lose its discrimination ability even though data from different classes are well separated in terms of Euclidean distance [104]. This assumption is reasonable for face recognition applications, as images of the same subject under different intensity levels are still considered to be of the same class [171]. However, due to non-stationary behaviour, no access to multiple classes (a two class classification problem), and low dimensionality of biomedical signals, sparse representation has not been used for such applications. To achieve optimum performance and prepare the signal for

sparse representation, modification of SRC is required. This modification will map the signal to a space with higher dimensions. This allows elements of the dictionary to adaptively update and use kernel methods to transfer samples to a nonlinear separable space, thereby increasing the classification performance.

Let $A_i = [a_1^i, a_2^i, ..., a_n^i] \in \mathbf{R}^{dXn_i}$ be the matrix where each column is the training sample from the ith class calculated using the K-NMF algorithm, $i = 1, ..., k$, and $\mathbf{y} \in \mathbf{R}^{dX1}$ be a new test sample. If a_{lj}^c and y_j are the jth entry of a_l^i and $\mathbf{y}$ respectively, the result is matrix A for the entire training set, by concatenating the training samples of all k classes:

$$A = [A_1, A_2, ...A_k] = [A_1^1, A_2^1, ..., A_{n_k}^k] \in \mathbf{R}^{dXn} \tag{7.6}$$

The sparse coefficients can be calculated based on the elements of dictionary A and the test samples of $\mathbf{y}$. Since most of the sparse representation applications are used in reconstruction of the training signals, such problems can be solved in polynomial time by standard Linear Programming (LP) methods, shown in Eq. 7.7 [114]. However, the conventional methods are not discriminatory and may not be suitable for applications such as classification.

$$\hat{\mathbf{x}}_1 = argmin\|\mathbf{x}\|_1 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y} \tag{7.7}$$

Our modification of the algorithm is based on kernel sparse representation. Let us assume the nonhomogeneous equation of [164]:

$$\mathbf{x}\epsilon K^n, \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y} \tag{7.8}$$

where $\mathbf{y}$ denotes the vector with m components. Suppose that the test sample y from the kth class can be approximately represented as the linear combination of the training samples of matrix A and $\mathbf{x}$, described as

$$Ax = y \Leftrightarrow \begin{matrix} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \\ \vdots \qquad \vdots \qquad \quad \vdots \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \end{matrix} \qquad (7.9)$$

To apply the kernel method, samples are mapped from the original space to another space by a nonlinear mapping. Given a signal $\mathbf{y} \epsilon \mathbf{R}^N$, and dictionary A, we seek a kernel sparse representation with the least number of nonzero coefficients such that $\mathbf{y} = \phi(A\mathbf{x})$, where ϕ is the embedded nonlinear mapping related to the kernel functions. With an appropriate kernel function, the sparse coefficients can be computed as a regular ℓ^2 norm in the kernel space and the sparse representation formulation of Eq. 7.7 can be written as:

$$\hat{\mathbf{x}}_1 = argmin \|x\|_2 \quad \text{subject to} \quad \|\phi(\mathbf{y}) - \phi(A\mathbf{x})\|_2 \leq \varepsilon \qquad (7.10)$$

where ideally

$$\hat{\mathbf{x}}_1 = [0, ...0, \alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,n_i}, 0, ...0]^T \in \mathbf{R}^n \qquad (7.11)$$

is a kernel mapped coefficient matrix whose entries are zero except those associated with the ith class. In other words, the nonzero entries in estimate $\hat{\mathbf{x}}_1$ will be associated with the columns of $\phi(A)$ from a single object class i, which can easily assign test sample $\mathbf{y}$ to

the class. However, noise and modelling errors may lead to small nonzero entries associated with multiple object classes [104]. To evaluate the performance of the aforementioned methodology we used the two available EMG datasets for sleep movement behaviour detection and classification, which serve to demonstrate the efficacy of the proposed classification algorithm. As also explained in the previous chapters, the two datasets are as follows: the leg EMG along with the MRI images to detect PLM disorder and the chin EMG to detect RBD. Figure 7.5 demonstrates the algorithm in form of a block diagram for this chapter. The following sections explain these in more detail.
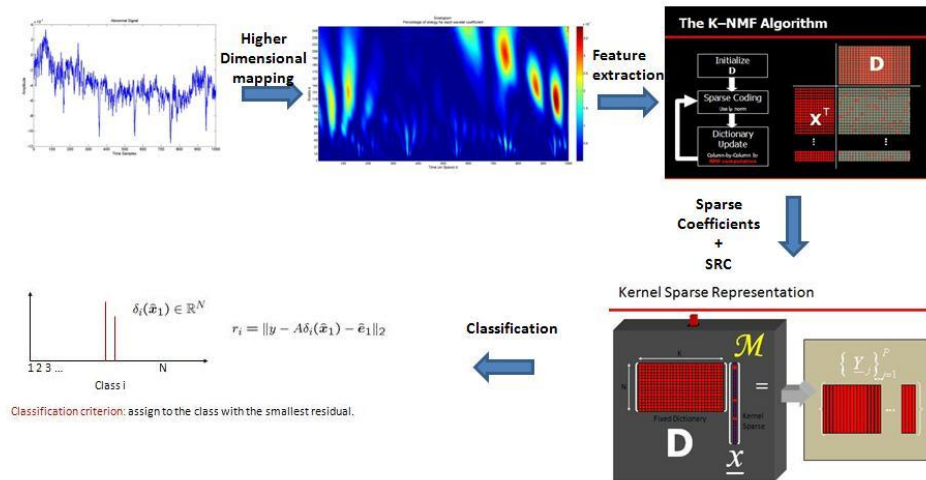
Figure 7.5: Algorithm block diagram

7.5   PLM Dataset

The extracted feature vectors $\hat{\mathbf{x}}_1$, which were calculated based on the kernel sparse repres-
entations, were used to discriminate the EMG signals into normal and abnormal classes with
the aid of an automatic classifier. LOO was used for testing the validity of the proposed
scheme and to assess the discrimination ability of the extracted features. The class labels
(class information) were WMH and PLM indices respectively.

In order to minimize the computational time based on the properties of the EMG signal,
only the first 20 epochs of the stage 2 EMG data were used for analysis (since most PLMs
occur during this stage of sleep). To further assess the proposed algorithm, REM sleep was
also used. Since movement is not expected for these patients during REM sleep (i.e., we
excluded participants with Parkinson's Disease or RBD), the analysis is expected to show
little to no improvement for these subjects.

Following the TF-representation, the TF matrix was fed to an adaptive dictionary, which
was constructed based on the K-NMF algorithm. This method finds the best dictionary
elements to represent the test sample $\{y_i\}_{i=1}^N$. Using NMF instead of SVD, allowed for less
computational time as well as better localization, which is very beneficial in applications
such as classification and recognition.

When the atom representation of adaptive dictionary elements D_i had been calculated for
each signal i where $i = 1, 2, ..., 65$, the new dictionary was based on concatenation of all these
elements: $A = [D_1, D_2, ..., D_{65}]$. This new dictionary was now a fixed dictionary and was
used to calculate the sparse coefficients. The algorithm for the kernel sparse used a kernel

function to calculate the sparse coefficients. In this method, the optimization problem of Eq. 7.7 was replaced by

$$\min \|x\|_1 \quad \text{subject to} \quad \phi(y) = \phi(Ax). \tag{7.12}$$

Solving Eq. 7.12 allowed us to calculate the sparse coefficients $\mathbf{x}$. Next $\mathbf{y}$ was classified in terms of $\mathbf{x}$. Similar to the SRC algorithm, the minimum residual between $\mathbf{y}$ and its reconstructed $\hat{\mathbf{y}}$ in the reduced subspace was used to determine the label c of $\mathbf{y}$. For each class (elevated PLM or WMH index, non-elevated PLM or WMH index), the characteristic function δ was defined, which will choose the coefficients corresponding to each class.

$$\delta_i(\hat{\mathbf{x}}_1) = \{ \begin{matrix} x_j & \text{if } c_j = i \\ 0 & \text{otherwise} \end{matrix} \tag{7.13}$$

The ith approximation to the test sample y can be expressed as $\phi(A\delta_i)$. Thereby, the estimated label $\hat{c}$ was used for $\mathbf{y}$ by minimizing the residual between the $\phi(y)$ and its approximation. Accordingly, the result is:

$$r_i(\mathbf{y}) = \|\Phi(\mathbf{y}) - \Phi(A\delta_i(\hat{\mathbf{x}}_1))\|_2 \tag{7.14}$$

A total of 65 subjects were used for this analysis. Each had a left and right leg recording, for a total of 130 signals. Figure 7.6 shows a comparison of left-leg EMG signals for two different patients. Based on medical records, Patient 40 experienced a very low number of PLMs and brain imaging details a very low concentration of WMH. Patient 140 was the

opposite, possessing a large amount of WMH and an abnormally high number of PLMs per hour. This can be seen in the sample taken from Stage 2 of the sleep cycle shown in Figure 7.6, as the signal from Patient 140 (blue) illustrated a greater number of PLMs. However, both patients did experience PLMs; therefore, deciphering a level of severity and eminent future risk in specific patients can be somewhat difficult.

In each of the assessments, the first 20 epochs of the signal were used totalling to $153,600$ samples (sampling rate of $256Hz$) per epoch. One trial signal out of the 130 signals, and the remaining 129 trial signals were used as training signals to produce classification accuracy. Tables 7.1 and 7.2 show the classification accuracy for the Stage 2 dataset using the PLM index (higher than 30) and WMH index (higher than 6) as class labels, respectively. The rationale for the specified threshold is that healthy middle-aged individuals without any sleep disturbance still have PLM indices > 10 per hour of sleep [172]. As well, the cut-off of ≥ 30 PLMs per hour has been shown to be clinically relevant with regards to the development of incident cardiovascular disease [39].

Table 7.1: Overall Classification performance of Stage 2 EMG using PLM index as the class label

Signal Name	Description of the Dataset
Higher PLM index	28 Left-Leg EMG with PLM index > 30
	28 Right-Leg EMG with PLM index > 30
	First 20 Epochs in Stage 2
Lower PLM index	37 Left-Leg EMG with PLM index < 30
	37 Right-Leg EMG with PLM index < 30
	First 20 Epochs in Stage 2
Overall accuracy of **97%** was achieved	

To further evaluate the proposed SRC method, we extended our validation to the REM stage of the equivalent EMG signals. Tables 7.3 and 7.4 show the results of this dataset.
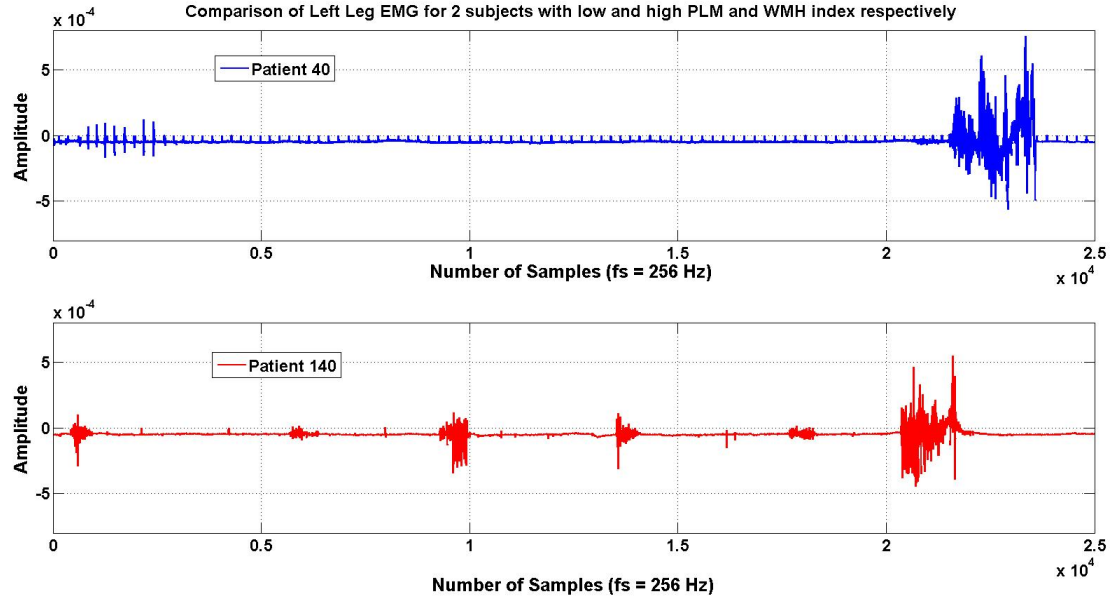
Figure 7.6: Comparison of two extreme cases of the leg EMG signals

Table 7.2: Overall Classification performance of Stage 2 EMG using WMH index as the class label

Signal Name	Description of the Dataset
Higher WMH Index	28 Left-Leg EMG with WMH index > 6 28 Right-Leg EMG with WMH index > 6 First 20 Epochs in Stage 2
Lower WMH Index	37 Left-Leg EMG with WMH index < 6 37 Right-Leg EMG with WMH index < 6 First 20 Epochs in Stage 2
An overall accuracy of **67%** was achieved	

Table 7.3: Overall Classification performance of Stage REM EMG using PLM index as the class label

Signal Name	Description of the Dataset
Higher PLM Index	12 Left-Leg EMG with PLM index > 30
	12 Right-Leg EMG with PLM index > 30
	First 20 Epochs in Stage REM
Lower PLM Index	49 Left-Leg EMG with PLM index < 30
	49 Right-Leg EMG with PLM index < 30
	First 20 Epochs in Stage REM
An overall accuracy of **44%** was achieved.	

Table 7.4: Overall Classification performance of Stage REM EMG using WMH index as the class label

Signal Name	Description of the Dataset
Higher WMH Index	23 Left-Leg EMG with WMH index > 6
	23 Right-Leg EMG with WMH index > 6
	First 20 Epochs in Stage REM
Lower WMH Index	38 Left-Leg EMG with WMH index < 6
	38 Right-Leg EMG with WMH index < 6
	First 20 Epochs in Stage REM
An overall accuracy of **52%** was achieved.	

A total of 61 subjects were analyzed (4 of the original 65 subjects did not have REM sleep recorded during their sleep studies). Overall, 122 signals were used for the analysis and the LOO method was used to evaluate the classification accuracy.

Although in the above two tables the overall classification accuracy is fairly low, this analysis validates our algorithm. REM sleep is categorized as a stage where there is little to no movement. Thus, there is an absence of PLMs in normal subjects during REM sleep. Generally speaking, only patients with Parkinson's disease, REM sleep behaviour disorder or other medical conditions / drug exposures will experience movements in REM sleep. The next section validates our second dataset on REM EMG to detect RBD for our modification of SRC algorithm based on kernel functions.

7.5.1   Chin EMG Dataset

The dataset consists of signal segments from 36 chin EMG signals (16 with normal behaviour and 20 with elevated muscle tone) from participants who had undergone the sleep test. The dataset includes 14 male subjects and 22 female subjects with an average age of 44 ± 20. 88% of these subjects were taking medications such as Zopiclone, Celebrex, Fluoxetine, Topiramate, or Maxalt for some sleep abnormalities.

The preprocessed signals, x, are then mapped to the TF representation of Eq 7.15.

$$V_{CWT}(t, s) = \frac{1}{\sqrt{s}} \int x(\tau) g(\frac{\tau - t}{s}) d\tau \tag{7.15}$$

In this equation the frequency resolution is 256 and the time samples varies from $128 - 1024$.
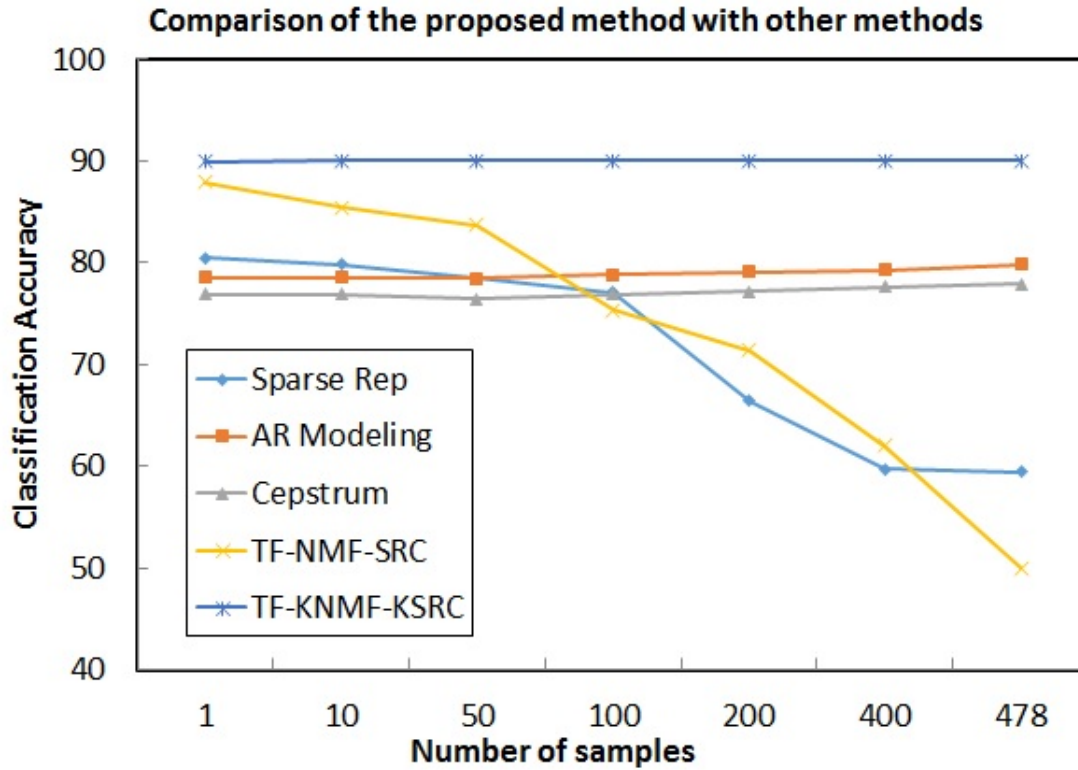
Figure 7.7: Comparison of the proposed method with the previously known methods dataset(A)

Following the TF-representation, the TF matrix is fed to the adaptive dictionary elements which are based on the K-NMF algorithm. This method finds the best dictionary elements to represent the test sample $\{y_i\}_{i=1}^{N}$, yet it is not the most discriminative information to be useful for classification. Thus we need algorithms that can transfer the linear inseparable samples to a space where they become linear separable. This will be achieved by incorporating kernel functions to sparse representations.

Once the updated elements of the dictionary are used to calculate the kernel sparse coefficients, we evaluate the classification performance, we assign **y** to the class with minimum

residual by attaining a generalized method using LOO. We evaluated the performance to be 92.4%, which is higher than the two previous methods developed of cepstrum and wavelet energy features (48.4% and 77.9% respectively) [72, 87].

The proposed technique not only allowed us to achieve higher accuracy compared to the other methods, but also increased the sparsity of the coefficients. As the complexity of the algorithm is proportional to the length of the signal, we achieved the highest accuracy at the expense of about $O(N^3)$ operations (including complexity of $O(N^2)$, $O(N^{3.5})$ and $O(N^3)$ for TF, K-NMF and the Kernel Sparse Representation for Classification (KSRC) respectively) .

7.6   Chapter Summary
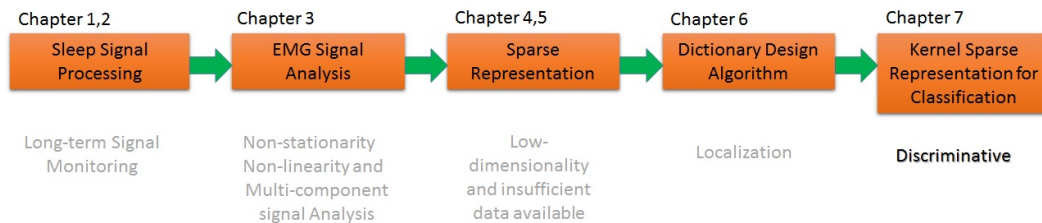
Figure 7.8: Flowchart of the proposed contributions

The contribution flowchart of this chapter is shown in Figure 7.8. In this chapter, a method was proposed to modify SRC based on a kernel algorithm, which classifies non-stationary EMG time-domain signals. This was accomplished using a non-traditional TF representation, alongside K-NMF. Three different analyses were applied to the signals. Depending on the signals, they were classified into differently distinguished categories for each analysis.

In one, the classes were elevated PLM and non-elevated PLM. In the other, high and low WMHs were assessed, while in the last one, elevated muscle tone and no muscle movement were used. Sparse representation was used to create the separation between the two classes. Modifications to conventional SRC methods were made to include the kernel SRC, which allowed for greater discrimination and more accurate classification of the test signals. KSRC is stronger than conventional methods, such as Linear Discriminant Analysis (LDA), when classifying extended time signals. LDA depends on a single decision line only. In comparison, KSRC aims to find a linear combination of the number of training signals that account for a given test signal, while achieving sparse representation properties. Adami *et al.* use a GMM for their method of classifying movements [173]. The movements evaluated were voluntary and were not measured during true sleep conditions, thus making a comparison more difficult. Nonetheless, the GMM algorithm achieved a classification performance of 84.6% [174]. Other manuscripts that included detailed algorithm descriptions were limited, and therefore the ability to compare comprehensively is hindered.

Even though dictionary learning algorithms play an important role in the classification task, SRC is essential in achieving the highest possible accuracy. Detection algorithms presented by Ferri *et al.*, Pittman *et al.*, and Wetter *et al.* do not include SRC and the overall classification average is lower [175–177]. The manuscripts mentioned above achieved overall classification percentages of 90%, 87%, and 94.3%, respectively [175–177]. To compare results more accurately, the specificity and sensitivity of the KSRC has been included for one of the studies. The proposed KSRC algorithm achieves a specificity and sensitivity of 96.05% and 98.15%, while Wetter *et al.* achieved 97%, and 92%, respectively [177]. It should be noted that the discussed research papers do include dictionary learning in some form.

When solely using dictionary learning algorithms, a fundamental limitation exists on the size of the learning dictionary; training a large dictionary requires a vast number of samples. The K-NMF constructed dictionary used alongside the proposed algorithm incorporated 122 signals from 65 subjects. The considerable number of signals used for training of the K-NMF dictionary likely had an effect on the increased accuracy. Research measurements performed by Ferri *et al.*, Pittman *et al.*, and Wetter *et al.* included a significantly smaller number of subjects, none of which involved even half as many samples [175–177] as the present work. Generally, the number of samples used in dictionary learning is expected to grow with the number of atoms to guarantee that sufficient information is present for the training process.

As shown in this chapter and the previous chapters, EMG signals are complex and long-term signals that are not easy to analyze. The structure and morphology is not well defined. Therefore a single-level feature extraction is not sufficient to analyze such signals. For this reason, in this chapter we proposed to use multilevel feature extraction approaches to tackle such signals. This is depicted in Figure 7.9.
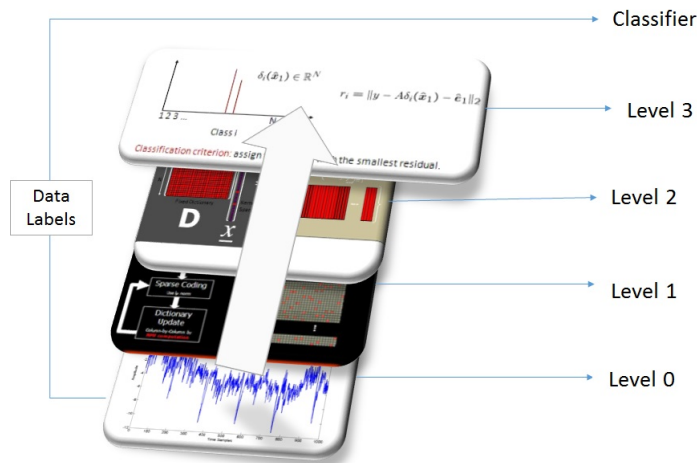
Figure 7.9: Multilayer Feature extraction by means of our proposed algorithm for EMG signals

Chapter 8

Conclusion

IN this dissertation, we presented a robust signal processing framework for efficiently analyzing and classifying non-stationary EMG signals. We evaluated the proposed method using synthetic and real EMG signals in different stages, and achieved desirable results. Figure 8.1 displays the contribution flowchart as they evolved throughout this dissertation.
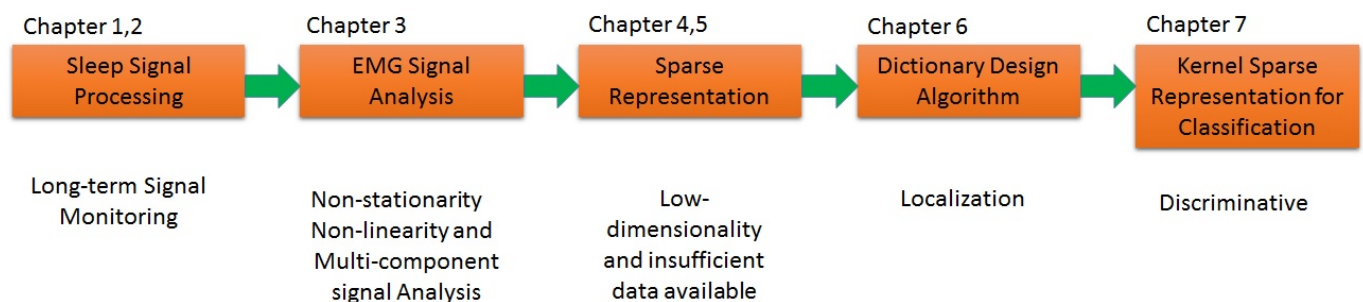
Figure 8.1: Flowchart of the proposed contributions

Chapter 1 presented a detailed introduction to sleep and sleep behaviours. Moreover, it

explained different sleep disorders and focused on sleep movement disorders. In Chapter 2, we described how sleep signals are acquired and explained the two main datasets used. EMG signal analysis was explained in Chapter 3, in addition to the challenges of analyzing sleep EMG signals. In Chapter 4, we presented SRC as an efficient tool for analyzing and classifying EMG signals, especially when insufficient samples are available. Chapter 5 explained different sparsity measures and our proposed measure of sparsity. In Chapter 6, we presented a novel adaptive learning algorithm that is used to design dictionary elements, based on NMF decomposition. The suitability for localization of long term signals was discussed with synthetic and real sleep signal examples. Chapter 7 focused on incorporating SRC with a discriminative term by means of kernel functions. This allowed us to enhance the performance and to produce robust results, which were evaluated using the real EMG signals in sleep.

8.1   Outcome of the proposed work

Table 8.1 summarizes the various solutions provided by the proposed sparse representation framework in efficiently analyzing and classifying non-stationary EMG signals. The proposed framework highlights how sparse representation can be used effectively when insufficient numbers of samples are available, in addition to localizing to an event of interest when it only happens in a short period of time. As well, sparse representation can be used for classification with robustness when incorporated with kernel functions. The outcome of the proposed work could be grouped into two core contributions, core theoretical contributions and core practical contributions.

Table 8.1: Summary of the proposed solutions and the requirements to efficiently analyze and classify the EMG signal in sleep

Requirements for efficient non-stationary EMG signal analysis in sleep	Solutions provided/suggeted by the proposed work	Chapter Reference
Sleep Analysis	Polysomnography	Chapter 2
Long-term signal processing	Sparse representation	Chapter 4
Insufficient number of samples	LOO	Chapter 4,5
Higher Dimensional Non-stationary Signal Representations	TF representation	Chapter 4, 6, and 7
Localization of the algorithm	NMF decomposition	Chapter 6
Increasing the robustness	K-NMF dictioanry design algorithm	Chapter 6
Increasing the classification performance	KSRC	Chapter 7
Feature selection	Measure of Sparsity	Chapter 5

8.1.1   Core theoretical Contribution

The following summarizes our core theoretical contributions in the overall area of EMG signal analysis and classification.

Sparse Representation Our main contribution is designing an algorithm based on sparse representation for classification which focuses attention on developing a long-term algorithm that works well when insufficient numbers of samples are available. To fulfill this objective, LOO has been used, which is the least biased estimate. Another key insight behind this approach was that this algorithm does not use any assumptions regarding to the dynamic changes in a given signal, as it can detect those changes by means of evaluating the sparse coefficients. In addition, this algorithm is capable of reducing the dimension of the signal without actually affecting the performance.

Measure of Sparsity A signal is sparse if its energy is concentrated in some transformed domain. A measure of sparsity can be used to search for such a domain. The key insight behind this approach was to measure the sparsity of the coefficients. The proposed method resulted in faster convergence as well as better classification performance, as long as the value of that measure is close to the best sparsity.

Fixed Dictionary Algorithms We presented the development of alternative dictionary algorithms based on existing mathematical models. We first used our previous algorithm based on AR modelling to design the dictionary. However, since we were interested in analyzing long-term signals we used NMF-based dictionaries. We developed a dictionary based on TF matrices and NMF decompositions.

Adaptive Dictionary Design Algorithms NMF codes naturally favour sparse, parts-based representations which in the context of recognition can be more robust than non-sparse, global features. This led to our main contribution, which was to design an adaptive dictionary based on NMF decomposition that alternates between sparse coding of the examples based on the current dictionary atoms to better fit the data, called the K-NMF. In this algorithm the update of the dictionary columns was combined with the update of the sparse representation, which resulted in accelerated convergence. The K-NMF consisted of two main parts: first it updated the non-negative dictionary elements, and then it calculated the sparse coefficients using the NMF decomposition algorithm.

Classification A new discriminant approach based on modification of sparse representation was offered to improve the classification accuracy in automated decision making

systems. A kernel function was added to the sparse representation algorithm that automatically identified the clusters of the signals which were then used for classification.

8.1.2 Core Practical Contributions

Real-world applications were employed to evaluate the proposed work. These applications are as follows:

Sleep Disorder Diagnosis Sleep and sleep-related problems play a role in a large number of human disorders, and affect every field of medicine. It is estimated that 50 to 70 million Americans suffer from a chronic sleep disorder, which hinders their daily life, affects their health, and incurs a significant economic burden to society. The negative public health consequences of sleep disorders are enormous and could have long-term effects, including increased risk of hypertension, diabetes, obesity, heart attack, stroke, and in some cases death. Although the skills acquired by medical practitioners 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 this dissertation we developed automatic detection of sleep movement disorder that could be used as an early detection and treatment of sleep disorders. These techniques cannot only help clinicians apply appropriate therapeutic procedures to control the abnormal behaviour, but can also help patients have a better quality of life. Since this is the first work reported in the area of automation of sleep disorders, the results are promising and show great potential for applications such as therapy and prediction

of fatal disease. This is especially true since all techniques were done without any assistance or interventions from the user, as this was a fully automated procedure. As such, systems like this one could in fact be used as a tool that could not only detect the abnormality behaviour but also as an earlier detection and diagnostic tool for different diseases such as Parkinson's and stroke.

EMG Signal Analysis in Sleep For the first time, we used sparse representation for classification of long-term EMG signals for detecting and classifying the signals into normal and abnormal classes.

Sleep Movement Disorder Motor activity in sleep is very common. In fact, any arousing event may cause movements, which is the primary reason in most of the abrupt changes in PSG. During sleep about 50 to 100 isolated movements of the arms and the legs typically occur. On average, axial changes of body position happen five to seven times during the night in blocks of 10 minutes occurring either before or after the REM sleep. However, during REM sleep, muscle twitching is common and may be recorded as a brief, spontaneous contraction causing a flicker of movements under the skin that can be recorded by means of the chin EMG recording. In general, any other movements in sleep are exacerbated by many causes, in particular as a physical pain or discomfort, psychological disturbances, environmental factors, and a wide variety of sleep movement disorders. The results obtained show promising solutions for such behaviours and have a high potential of technology transfer. This technology may lead to the development of novel implications in sleep disorder monitoring, which in turn will benefit global health care.

172

Periodic Leg Movement Disorder PLM are detected in approximately 80% of patients with RLS, but are also seen in individuals with obstructive sleep apnea (OSA), RBD congestive heart failure, Parkinson's disease, Multiple System Atrophy, as well as in users of antidepressant medications, and even in seemingly healthy subjects, particularly the elderly. To date, the largest epidemiological study reported simultaneous presence of PLM and sleep complaints in 3.9% of 18,980 subjects. The clinical significance of PLMs is unclear; however, emerging research suggests important associations with nocturnal sympathetic activation, inflammatory markers, congestive heart failure, incident cardiovascular disease, and mortality in specific patient populations. In addition, PLMs have recently been shown to be associated with WMH, which are independently associated with incidental stroke and death. Untreated PLM could lead to a three- to four-fold increased risk of hypertension, stroke, or, HF; a potentially increased risk of cancer mortality; and up to a four-fold increased risk of motor vehicle accidents. Our automated analysis of sleep signals has the potential to facilitate detection of clinically relevant nocturnal phenomena in a cost-effective and efficient manner. In this dissertation, we used 65 subjects who had undergone sleep recording as well as MRI to record their brain activity. We achieved a significant accuracy of 97%.

REM Sleep Behaviour Disorder RBD is characterized by elaborating movements correlated with dreaming during REM sleep. 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. 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 active 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 loss, which means loss of safe sleep. Serious injuries have the potential to occur as a result of dream acting behaviour. The overnight polysomnography of subjects also confirmed the presence of muscle tone in REM sleep. As well, most of the patients with RBD have significant abnormal REM sleep muscle activity. In addition, RBD can be an early warning for the emergence of Parkinson's and other neurodegenerative conditions antedating the illness by many years. Given numerous potentially relevant health-related linkages, our investigation in sleep disorders has the potential to facilitate detection and prevention of such diseases. In this work we used chin EMG signal to analyze subjects who had elevated muscle tones in their REM sleep compared to healthy subjects.

8.2   Limitations and Future Work

The following could be the directions for future work in applying and enhancing the proposed work with more intelligence and accuracy.

- With the suggested algorithm, discriminatory signal patterns are detected in a two-group classification problem. This can be extended to design a unique multidimen-

sional adaptive system to categorize the patterns in scenarios with more than two groups.

- The results can be tabulated such that the variations of the performance for each training set is observed for the LMO cross-validation approach.

- The developed SRC algorithm could be applied to different sleep abnormality detections, in particular sleep apnea. In general, the proposed algorithm can be used to develop an automated algorithm based on sparse representation to measure and estimate the Neck Fluid Volume and the changes in the tracheal sounds, a biomarker in sleep apnea detections.

- The NMF technique decomposed a TF matrix into r spectral and temporal components where the values of r were experimentally selected. The more accurately we decomposed the TF into its components, the more accurately we could extract the discriminatory features when designing our dictionary. In Chapter 6, we could extend the work to have an optimum order number for the decomposition of such a matrix.

- The proposed method was only applied to the portion of the signal depending on the application used. For example, for the Leg EMG we only used the first 20 epochs of Stage 2, where most PLMs occur or only during the REM sleep stage for chin EMG. This could later be used for all the EMG signals in sleep, i.e. all the stages.

- Although the computational expenses for such an algorithm are high, the advantages of the proposed technique outweigh the computational expenses. This is because in the near future, rapidly growing technological developments will significantly reduce

processing time so that the technique will eventually become a real-time processing tool. Additionally, the technique can be implemented on a dedicated device to be more easily marketable especially in sleep labs where technicians and sleep experts have to automatically score and detect the abnormalities that exists in these long-term signals.

- With the advent of wearable devices that can record sleep signals, the urge and tendency for sophisticated signal processing is on the rise. This makes the signal processing an ongoing never ending research.

References

[1] Mar'ia Mikhailovna Manaseina. *Sleep: its physiology, pathology, hygiene, and psychology*. W. Scott, ltd., 1897.

[2] Mark Peplow. Structure: The anatomy of sleep. *Nature*, 497(7450):S2–S3, 2013.

[3] Cameron D Harris. Neurophysiology of sleep and wakefulness. *Respir Care Clin N Am*, 11(4):567–86, 2005.

[4] Andrew Winokur, Keith A Gary, Shannon Rodner, Carole Rae-Red, Antonio T Fernando, and Martin P Szuba. Depression, sleep physiology, and antidepressant drugs. *Depression and anxiety*, 14(1):19–28, 2001.

[5] Anthony Kales. *Sleep: Physiology & Pathology: A Symposium*. Lippincott, 1969.

[6] National Sleep Foundation NSF. *National Sleep Foundation - The Latest Sleep News and Information*. Online, 2014.

[7] Joris C Verster, Seithikurippu R Pandi-Perumal, and David L Streiner. *Sleep and quality of life in clinical medicine*. Springer, 2008.

[8] Mehrnaz Shokrollahi. Analysis of electromyogram in rapid eye movenet sleep. Master's thesis, Ryerson University, 2009.

[9] CM Shapiro and MJ Flanigan. Abc of sleep disorders. function of sleep. *BMJ: British Medical Journal*, 306(6874):383, 1993.

[10] A. Rechetschaffen and A. Kale. Manual of standardized terminology techniques and scoring system for stages of human subjects 3rd. los angels: Brain information service. *Brain Research Institute of Uni. versity of California*, 1973.

[11] Colin M Shapiro, Ralph Bortz, Duncan Mitchell, Peter Bartel, and Pieter Jooste. Slow-wave sleep: a recovery period after exercise. *Science*, 214(4526):1253–1254, 1981.

[12] Claudio Bassetti and Michael S Aldrich. Narcolepsy. *Neurologic clinics*, 14(3):545–571, 1996.

[13] Farideh Ebrahimi, Mohammad Mikaeili, Edson Estrada, and Homer Nazeran. Automatic sleep stage classification based on eeg signals by using neural networks and wavelet packet coefficients. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 1151–1154. IEEE, 2008.

[14] Raymond Cooper, John Walkinshaw Osselton, and John Crossley Shaw. *EEG technology*. Butterworth-Heinemann, 1974.

[15] Rangaraj M Rangayyan. *Biomedical signal analysis*. IEEE Standards Office, 2001.

[16] Saeid Sanei and Jonathon A Chambers. *EEG signal processing.* John Wiley & Sons, 2008.

[17] Noboru Takeuchi, Naohisa Uchimura, Yuji Hashizume, Masaki Mukai, Yoshinori Etoh, Katsuyasu Yamamoto, Tatayu Kotorii, Hiroharu Ohshima, Masachika Ohshima, and Hisao Maeda. Melatonin therapy for REM sleep behavior disorder. *Psychiatry and clinical neurosciences*, 55(3):267–269, 2001.

[18] Maurice M Ohayon. Epidemiology of insomnia: what we know and what we still need to learn. *Sleep medicine reviews*, 6(2):97–111, 2002.

[19] Charles M Morin, M LeBlanc, M Daley, JP Gregoire, and C Merette. Epidemiology of insomnia: prevalence, self-help treatments, consultations, and determinants of help-seeking behaviors. *Sleep medicine*, 7(2):123–130, 2006.

[20] Raman Malhotra. Sleep disorders: Their impact on public health. *JAMA*, 299(13):1612–1613, 2008.

[21] DAVID W Hudgel and THERESA Harasick. Fluctuation in timing of upper airway and chest wall inspiratory muscle activity in obstructive sleep apnea. *J Appl Physiol*, 69(2):443–450, 1990.

[22] Sudhansu Chokroverty. *Sleep disorders medicine: basic science, technical considerations, and clinical aspects.* Butterworth-Heinemann, 1994.

[23] Yves Dauvilliers, Sylvie Rompré, Jean-Françis Gagnon, Mélanie Vendette, Dominique Petit, and Jacques Montplaisir. REM sleep characteristics in narcolepsy and REM sleep behavior disorder. *Sleep*, 30(7):844, 2007.

[24] Jacques Montplaisir, Sylvie Boucher, Gaétan Poirier, Gilles Lavigne, Odile Lapierre, and Paul Lespérance. Clinical, polysomnographic, and genetic characteristics of restless legs syndrome: a study of 133 patients diagnosed with new standard criteria. *Movement Disorders*, 12(1):61–65, 1997.

[25] Magdolna Hornyak, Bernd Feige, Dieter Riemann, and Ulrich Voderholzer. Periodic leg movements in sleep and periodic limb movement disorder: prevalence, clinical significance and treatment. *Sleep medicine reviews*, 10(3):169–177, 2006.

[26] ML Fantini, M Michaud, N Gosselin, G Lavigne, and J Montplaisir. Periodic leg movements in REM sleep behavior disorder and related autonomic and eeg activation. *Neurology*, 59(12):1889–1894, 2002.

[27] Patrick J Hanly and Naheed Zuberi-Khokhar. Periodic limb movements during sleep in patients with congestive heart failure. *CHEST Journal*, 109(6):1497–1502, 1996.

[28] Thomas C Wetter, Victor Collado-Seidel, Alexander Yassouridis, Claudia Trenkwalder, et al. Sleep and periodic leg movement patterns in drug-free patients with parkinson's disease and multiple system atrophy. *Sleep: Journal of Sleep Research & Sleep Medicine*, 2000.

[29] Changkook Yang, David P White, and John W Winkelman. Antidepressants and periodic leg movements of sleep. *Biological psychiatry*, 58(6):510–514, 2005.

[30] S Ancoli-Israel, DF Kripke, MR Klauber, WJ Mason, R Fell, and O Kaplan. Periodic limb movements in sleep in community-dwelling elderly. *Sleep*, 14(6):496–500, 1991.

[31] Martin Desseilles, Thanh Dang-Vu, Manuel Schabus, Virginie Sterpenich, Pierre Maquet, and Sophie Schwartz. Neuroimaging insights into the pathophysiology of sleep disorders. *Sleep*, 31(6):777, 2008.

[32] Conrad Iber, American Academy of Sleep Medicine, et al. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications.* American Academy of Sleep Medicine, 2007.

[33] Douglas M Wallace, Alberto R Ramos, and Tatjana Rundek. Sleep disorders and stroke. *International Journal of Stroke*, 7(3):231–242, 2012.

[34] Maria Alessandria and Federica Provini. Periodic limb movements during sleep: a new sleep-related cardiovascular risk factor? *Frontiers in neurology*, 4, 2013.

[35] Lynn Marie Trotti, David B Rye, Christine De Staercke, W Craig Hooper, Arshed Quyyumi, and Donald L Bliwise. Elevated c-reactive protein is associated with severe periodic leg movements of sleep in patients with restless legs syndrome. *Brain, behavior, and immunity*, 26(8):1239–1243, 2012.

[36] Taha Tahir Bekci, Mehmet Kayrak, Aysel Kiyici, Hatem Ari, Turgut Teke, Emin Maden, and Hakan Akilli. The relation between lp-pla2 levels with periodic limb movements. *Sleep and Breathing*, 16(1):117–122, 2012.

[37] Shahrokh Javaheri. Sleep disorders in systolic heart failure: a prospective study of 100 male patients. the final report. *International Journal of cardiology*, 106(1):21–28, 2006.

[38] Robert Skomro, Rogerio Silva, Rosana Alves, Adelaide Figueiredo, and Geraldo Lorenzi-Filho. The prevalence and significance of periodic leg movements during sleep in patients with congestive heart failure. *Sleep and Breathing*, 13(1):43–47, 2009.

[39] Brian B Koo, Terri Blackwell, Sonia Ancoli-Israel, Katie L Stone, Marcia L Stefanick, Susan Redline, et al. Association of incident cardiovascular disease with periodic limb movements during sleep in older men outcomes of sleep disorders in older men (mros) study. *Circulation*, 124(11):1223–1231, 2011.

[40] Mahek Mirza, Win-Kuang Shen, Aamir Sofi, Ahad Jahangir, Naoyo Mori, A Jamil Tajik, and Arshad Jahangir. Frequent periodic leg movement during sleep is associated with left ventricular hypertrophy and adverse cardiovascular outcomes. *Journal of the American Society of Echocardiography*, 26(7):783–790, 2013.

[41] Alexander Joseph Moszczynski, Anu Tandon, Fernando Morgadinho Santos Coelho, Lorne Zinman, and Brian Murray. Mortality associated with periodic limb movements during sleep in amyotrophic lateral sclerosis patients. *Einstein (São Paulo)*, 10(4):428–432, 2012.

[42] Carlos H Schenck and Mark W Mahowald. REM sleep behavior disorder: clinical, developmental, and neuroscience perspectives 16 years after its formal identification in sleep. *Sleep*, 25(2):120–138, 2002.

[43] Birgit Frauscher, Alex Iranzo, Birgit Högl, Jordi Casanova-Molla, Manel Salamero, Viola Gschliesser, Eduardo Tolosa, Werner Poewe, and Joan Santamaria. Quantific-

ation of electromyographic activity during REM sleep in multiple muscles in REM sleep behavior disorder. *Sleep*, 31(5):724, 2008.

[44] Diego Garcia-Borreguero, Ana B Caminero, Yolanda De La Llave, Oscar Larrosa, Soledad Barrio, Juan J Granizo, and Juan A Pareja. 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 disorders*, 17(5):934–941, 2002.

[45] Carlos H Schenck and Mark W Mahowald. Rapid eye movement sleep parasomnias. *Neurologic clinics*, 23(4):1107–1126, 2005.

[46] Bradley F Boeve, Michael H Silber, and Tanis J Ferman. REM sleep behavior disorder in parkinson disease and dementia with lewy bodies. *Journal of geriatric psychiatry and neurology*, 17(3):146–157, 2004.

[47] Allan Rechtschaffen and Anthony Kales. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* US Government Printing Office, US Public Health Service, 1968.

[48] Christian Guilleminault and William C Dement. *Sleep apnea syndromes*, volume 11. AR Liss, 1978.

[49] Carlos H Schenck and Mark W Mahowald. Polysomnographic, neurologic, psychiatric, and clinical outcome report on 70 consecutive cases with REM sleep behavior disorder (RBD): sustained clonazepam efficacy in 89.5% of 57 treated patients. *Cleveland Clinic journal of medicine*, 57(Supplement):S–9, 1990.

[50] José Haba-Rubio and Jean Krieger. Evaluation instruments for sleep disorders: A brief history of polysomnography and sleep medicine. In *Introduction to Modern Sleep Technology*, pages 19–31. Springer, 2012.

[51] Henri Piéron. *Le problème physiologique du sommeil*. Masson et Cie, 1913.

[52] Ian M Colrain. The k-complex: a 7-decade history. *Sleep*, 28(2):255–273, 2005.

[53] H Blake and RW Gerard. Brain potentials during sleep. *Am J Physiol*, 119(692-703):796, 1937.

[54] Helen Blake, Ralph W Gerard, and Nathaniel Kleitman. *Factors influencing brain potentials during sleep*. PhD thesis, Am Physiological Soc, 1939.

[55] M Jouvet, F Michel, and J Courjon. On a stage of rapid cerebral electrical activity in the course of physiological sleep. *Comptes rendus des seances de la Societe de biologie et de ses filiales*, 153:1024, 1959.

[56] Mark I Boulos, Karthikeyan Umapathy, Peyman Shokrollahi, Kristiina McConville, Tess Sudenis, Dana R Jewell, Sridhar Krishnan, and Brian J Murray. Automated detection of nocturnal slow eye movements modulated by selective serotonin reuptake inhibitors. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 35(1):126–130, 2011.

[57] Madeleine M Grigg-Damberger. The aasm scoring manual: a critical appraisal. *Current opinion in pulmonary medicine*, 15(6):540–549, 2009.

[58] LO Wahlund, F Barkhof, F Fazekas, L Bronge, M Augustin, M Sjögren, A Wallin, H Ader, Didier Leys, L Pantoni, et al. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke*, 32(6):1318–1322, 2001.

[59] Behnaz Ghoraani. *Time-frequency Feature Analysis*. PhD thesis, Ryerson University, 2010.

[60] Douglas K Lindner. *Introduction to signals and systems*. McGraw-Hill, 1999.

[61] Alan V Oppenheim, Ronald W Schafer, John R Buck, et al. *Discrete-time signal processing*, volume 2. Prentice-hall Englewood Cliffs, 1989.

[62] Charles L Philips, John M Parr, and E Riskin. *Signals, systems, and transforms*. Prentice Hall, 1995.

[63] Anil K Jain. Fundamentals of digital signal processing. *Fundamentals of Digital Signal Processing*, 1989.

[64] John G Proakis. *Digital signal processing: principles algorithms and applications*. Pearson Education India, 2001.

[65] Dimitris G Manolakis, Vinay K Ingle, and Stephen M Kogon. *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing*, volume 46. Artech House Norwood, 2005.

[66] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. comput*, 19(90):297–301, 1965.

[67] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.

[68] Hirotugu Akaike. Fitting autoregressive models for prediction. *Annals of the institute of statistical mathematics*, 21(1):243–247, 1969.

[69] Sanjeev Tavathia, Rangaraj Mandayam Rangayyan, Cyril Basil Frank, GD Bell, KO Ladly, and Y-T Zhang. Analysis of knee vibration signals using linear prediction. *Biomedical Engineering, IEEE Transactions on*, 39(9):959–970, 1992.

[70] Rangaraj M Rangayyan, Sridhar Krishnan, G Douglas Bell, Cyril B Frank, and Katherine O Ladly. Parametric representation and screening of knee joint vibroarthrographic signals. *Biomedical Engineering, IEEE Transactions on*, 44(11):1068–1074, 1997.

[71] Zahra MK Maussavi, Rangaraj M Rangayyan, G Douglas Bell, Cyril B Frank, and KO Ladly. Screening of vibroarthrographic signals via adaptive segmentation and linear prediction modeling. *Biomedical Engineering, IEEE Transactions on*, 43(1):15, 1996.

[72] M. Shokrollahi, S. Krishnan, D. Jewell, and B. Murray. Autoregressive and cepstral analysis of electromyogram in rapid movement sleep. In *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pages 1580–1583. Springer, 2009.

[73] Daniel Griffin and Jae S Lim. Signal estimation from modified short-time fourier

transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(2):236–243, 1984.

[74] Yunfeng Wu, Sridhar Krishnan, and Rangaraj M Rangayyan. Computer-aided diagnosis of knee-joint disorders via vibroarthrographic signal analysis: a review. *Critical Reviews in Biomedical Engineering*, 38(2), 2010.

[75] Jean Morlet, G Arens, E Fourgeau, and D Glard. Wave propagation and sampling theory-part i: Complex signal and scattering in multilayered media. *Geophysics*, 47(2):203–221, 1982.

[76] Stéphane Mallat. *A wavelet tour of signal processing*. Academic press, 1999.

[77] Ingrid Daubechies, Alex Grossmann, and Y Meyer. Painless nonorthogonal expansions. *Journal of Mathematical Physics*, 27(5):1271–1283, 1986.

[78] Ingrid Daubechies. The wavelet transform, time-frequency localization and signal analysis. *Information Theory, IEEE Transactions on*, 36(5):961–1005, 1990.

[79] Stephane Mallat and Sifen Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on pattern analysis and machine intelligence*, 14(7):710–732, 1992.

[80] Richardson N Leao and John A Burne. Continuous wavelet transform in the evaluation of stretch reflex responses from surface emg. *Journal of neuroscience methods*, 133(1):115–125, 2004.

[81] Marco Gazzoni, Dario Farina, and Roberto Merletti. A new method for the extraction

and classification of single motor unit action potentials from surface emg signals. *Journal of neuroscience methods*, 136(2):165–177, 2004.

[82] Matteo Arvetti, Giuseppina Gini, and Michele Folgheraiter. Classification of EMG signals through wavelet analysis and neural networks for controlling an active hand prosthesis. In *Rehabilitation Robotics, 2007. ICORR 2007. IEEE 10th International Conference on*, pages 531–536. IEEE, 2007.

[83] Pietro Liò, Anna T Lawniczak, Shengkun Xie, and Jiaying Xu. Wavelet-domain statistics of packet switching networks near traffic congestion. In *Bio-Inspired Computing and Communication*, pages 268–279. Springer, 2008.

[84] Tianhorng Chang and C-CJ Kuo. Texture analysis and classification with tree-structured wavelet transform. *Image Processing, IEEE Transactions on*, 2(4):429–441, 1993.

[85] S Arivazhagan and L Ganesan. Texture classification using wavelet transform. *Pattern recognition letters*, 24(9):1513–1521, 2003.

[86] Tryphon Lambrou, PSRSM Kudumakis, R Speller, M Sandler, and A Linney. Classification of audio signals using statistical features on time and wavelet transform domains. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 6, pages 3621–3624. IEEE, 1998.

[87] Mehrnaz Shokrollahi, Sridhar Krishnan, Dana Jewell, and Brian Murray. Analysis of the electromyogram of rapid eye movement sleep using wavelet techniques. In

Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, pages 2659–2662. IEEE, 2009.

[88] Ronald R Coifman, Yves Meyer, and Victor Wickerhauser. Wavelet analysis and signal processing. In *In Wavelets and their Applications*. Citeseer, 1992.

[89] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.

[90] Jeffrey R Cram and Jeffrey C Steger. EMG scanning in the diagnosis of chronic pain. *Biofeedback and self-regulation*, 8(2):229–241, 1983.

[91] JM Lopes, E Tabachnik, NL Muller, H Levison, and AC Bryan. Total airway resistance and respiratory muscle activity during sleep. *J Appl Physiol*, 54(3):773–777, 1983.

[92] Coleman Phd, M Richard, Charles P Pollak, and Elliot D Weitzman. Periodic movements in sleep (nocturnal myoclonus): relation to sleep disorders. *Annals of neurology*, 8(4):416–421, 1980.

[93] Carlos H Schenck and Mark W Mahowald. Motor dyscontrol in narcolepsy: Rapid-eye-movement (rem) sleep without atonia and REM sleep behavior disorder. *Annals of neurology*, 32(1):3–10, 1992.

[94] TL Lee-Chiong Jr. Parasomnias and other sleep-related movement disorders. *Primary care*, 32(2):415–434, 2005.

[95] Jun Huang, Stephen E Levinson, and Mark Hasegawa-Johnson. Signal approximation

in hilbert space and its application on articulatory speech synthesis. In *INTER-SPEECH*, pages 775–778, 2000.

[96] Ke Huang and Selin Aviyente. Sparse representation for signal classification. In *Advances in neural information processing systems*, pages 609–616, 2006.

[97] Yuanqing Li, Andrzej Cichocki, and Shun-ichi Amari. Analysis of sparse representation and blind source separation. *Neural computation*, 16(6):1193–1234, 2004.

[98] Yuanqing Li, Praneeth Namburi, Zhuliang Yu, Cuntai Guan, Jianfeng Feng, and Zhenghui Gu. Voxel selection in fMRI data analysis based on sparse representation. *Biomedical Engineering, IEEE Transactions on*, 56(10):2439–2451, 2009.

[99] Albert Nigrin. *Neural networks for pattern recognition*. MIT press, 1993.

[100] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[101] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[102] M.H. Hansen and B. Yu. Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774, 2001.

[103] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[104] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition

via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 210–227, 2008.

[105] H. Xiyi and W. Fang-Xiang. Sparse representation for classification of tumors using gene expression data. *Journal of Biomedicine and Biotechnology*, 2009.

[106] Mehrnaz Shokrollahi and Sridhar Krishnan. Sleep EMG analysis using sparse signal representation and classification. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 3480–3483. IEEE, 2012.

[107] K. Huang and S. Aviyente. Sparse representation for signal classification. *Advances in neural information processing systems*, 19:609, 2007.

[108] C. Lan, X. Jing, S. Li, L. Bian, and Y. Yao. Exploring the natural discriminative information of sparse representation for feature extraction. In *Image and Signal Processing (CISP), 2010 3rd International Congress on*, volume 2, pages 916–920. IEEE, 2010.

[109] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International joint Conference on artificial intelligence*, volume 14, pages 1137–1145, 1995.

[110] D.L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197, 2003.

[111] Min Xu, Chunhou Zheng, D Zhang, and NT-Y Vincent. Tumor classification via sparse representation based on metasample. In *Knowledge Acquisition and Modeling, 2009. KAM'09. Second International Symposium on*, volume 1, pages 31–34. IEEE, 2009.

[112] A. Banerjee, R. Juang, J. Broadwater, and P. Burlina. Sparse feature extraction for support vector data description applications. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 4236–4239, 2010.

[113] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1999.

[114] D.L. Donoho. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, 59(6):797–829, 2006.

[115] Richard G Baraniuk. Compressive sensing. *IEEE signal processing magazine*, 24(4), 2007.

[116] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, 2009.

[117] John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S Huang, and Shuicheng Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, 2010.

[118] Irina F Gorodnitsky and Bhaskar D Rao. Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm. *Signal Processing, IEEE Transactions on*, 45(3):600–616, 1997.

[119] Robert J Adams, Yuan Xu, and Francis X Canning. Sparse pseudo inverse of the discrete plane wave transform. *Antennas and Propagation, IEEE Transactions on*, 56(2):475–484, 2008.

[120] Ehsan Elhamifar. *Sparse Modeling for High-dimensional Multi-manifold Data Analysis*. PhD thesis, Citeseer, 2012.

[121] Michal Aharon and Michael Elad. Sparse and redundant modeling of image content using an image-signature-dictionary. *SIAM Journal on Imaging Sciences*, 1(3):228–247, 2008.

[122] S.J. Raudys and A.K. Jain. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on pattern analysis and machine intelligence*, 13(3):252–264, 1991.

[123] Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. In *Encyclopedia of database systems*, pages 532–538. Springer, 2009.

[124] Anil K Jain, Robert P. W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37, 2000.

[125] S.P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ Pr, 2004.

[126] E. Candes and J. Romberg. l1-magic: Recovery of sparse signals via convex programming. *URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf*, 2005.

[127] S Nightingale, JC Orgill, IO Ebrahim, SF De Lacy, S Agrawal, and AJ Williams. The association between narcolepsy and *rem* behaviour disorder (RBD). *Sleep medicine*, 6(3):253–258, 2005.

[128] Eamonn J Keogh and Michael J Pazzani. A simple dimensionality reduction technique for fast similarity search in large time series databases. In *Knowledge Discovery and Data Mining. Current Issues and New Applications*, pages 122–133. Springer, 2000.

[129] Maria G Jafari and Mark D Plumbley. Fast dictionary learning for sparse representations of speech signals. *Selected Topics in Signal Processing, IEEE Journal of*, 5(5):1025–1031, 2011.

[130] L Rebollo-Neira. Dictionary redundancy elimination. *IEE Proceedings-Vision, Image and Signal Processing*, 151(1):31–34, 2004.

[131] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.

[132] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *Image Processing, IEEE Transactions on*, 17(1):53–69, 2008.

[133] Behnaz Ghoraani and Sridhar Krishnan. A joint time-frequency and matrix decomposition feature extraction methodology for pathological voice classification. *EURASIP Journal on Advances in Signal Processing*, 2009:10, 2009.

[134] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[135] Joel A Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *Information Theory, IEEE Transactions on*, 52(3):1030–1051, 2006.

[136] Moshe Mishali and Yonina C Eldar. Reduce and boost: Recovering arbitrary sets of jointly sparse vectors. *Signal Processing, IEEE Transactions on*, 56(10):4692–4702, 2008.

[137] Mehmet Akçakaya and Vahid Tarokh. A frame construction and a universal distortion bound for sparse representations. *Signal Processing, IEEE Transactions on*, 56(6):2443–2450, 2008.

[138] N. Hurley and S. Rickard. Comparing measures of sparsity. *Information Theory, IEEE Transactions on*, 55(10):4723–4741, 2009.

[139] Hugh Dalton. The measurement of the inequality of incomes. *The Economic Journal*, pages 348–361, 1920.

[140] Scott Rickard and Maurice Fallon. The Gini index of speech. In *Proceedings of the 38th Conference on Information Science and Systems (CISS 04)*, 2004.

[141] Weidong Zhu, Jingyu Feng, and Yongmin Lin. Using gini-index for feature selection in text categorization. In *2014 International Conference on Information, Business and Education Technology (ICIBET 2014)*. Atlantis Press, 2014.

[142] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[143] J. Wright and Y. Ma. Dense error correction via l1-minimization. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 3033–3036. IEEE, 2009.

[144] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst. Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms. *Journal of Fourier analysis and Applications*, 14(5):655–687, 2008.

[145] D.L. Donoho and J. Tanner. Precise undersampling theorems. *Proceedings of the IEEE*, 98(6):913–924, 2010.

[146] Anshul Gupta, George Karypis, and Vipin Kumar. Highly scalable parallel algorithms for sparse matrix factorization. *Parallel and Distributed Systems, IEEE Transactions on*, 8(5):502–520, 1997.

[147] Jont B Allen and LAWRENCE Rabiner. A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11):1558–1564, 1977.

[148] Dennis Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[149] Martin J Bastiaans. Gabor's expansion of a signal into gaussian elementary signals. *Proceedings of the IEEE*, 68(4):538–539, 1980.

[150] John G Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(7):1169–1179, 1988.

[151] Ron Rubinstein, Michael Zibulevsky, and Michael Elad. Double sparsity: Learning sparse dictionaries for sparse signal approximation. *Signal Processing, IEEE Transactions on*, 58(3):1553–1564, 2010.

[152] Qiang Zhang and Baoxin Li. Discriminative K-SVD for dictionary learning in face recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2691–2698. IEEE, 2010.

[153] Zhuolin Jiang, Zhe Lin, and Larry S Davis. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1697–1704. IEEE, 2011.

[154] Ignacio Ramirez, Pablo Sprechmann, and Guillermo Sapiro. Classification and clustering via dictionary learning with structured incoherence and shared features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3501–3508. IEEE, 2010.

[155] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Discriminative learned dictionaries for local image analysis. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[156] P.O. Hoyer. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research*, 5:1457–1469, 2004.

[157] Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

[158] M.J. Lee, A.S. Lee, D.K. Lee, and S.Y. Lee. Video representation with dynamic features from multi-frame frame-difference images. In *Motion and Video Computing, 2007. WMVC'07. IEEE Workshop on*, pages 28–28. IEEE, 2007.

[159] I.B. Ciocoiu. Occluded face recognition using parts-based representation methods. In *Circuit Theory and Design, 2005. Proceedings of the 2005 European Conference on*, volume 1, pages I–315. IEEE, 2005.

[160] N.F. Chikhi, B. Rothenburger, and N. Aussenac-Gilles. A comparison of dimensionality reduction techniques for web structure mining. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 116–119. IEEE Computer Society, 2007.

[161] Kjersti Engan, Sven Ole Aase, and J Hakon Husoy. Method of optimal directions for frame design. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 5, pages 2443–2446. IEEE, 1999.

[162] Michal Aharon, Michael Elad, and Alfred M Bruckstein. K-SVD and its non-negative variant for dictionary design. In *Optics & Photonics 2005*, pages 591411–591411. International Society for Optics and Photonics, 2005.

[163] Kjersti Engan, Bhaskar D Rao, and Kenneth Kreutz-Delgado. Frame design using focuss with method of optimal directions (*mod*). In *Proc. NORSIG*, volume 99, 1999.

[164] Li Zhang, Wei-Da Zhou, Pei-Chann Chang, Jing Liu, Zhe Yan, Ting Wang, and Fan-Zhang Li. Kernel sparse representation-based classifier. *Signal Processing, IEEE Transactions on*, 60(4):1684–1695, 2012.

[165] Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods.* Cambridge University press, 2000.

[166] John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis.* Cambridge University press, 2004.

[167] Tommi Jaakkola, David Haussler, et al. Exploiting generative models in discriminative classifiers. *Advances in neural information processing systems*, pages 487–493, 1999.

[168] Koji Tsuda, Shotaro Akaho, Motoaki Kawanabe, and Klaus-Robert Müller. Asymptotic properties of the fisher kernel. *Neural Computation*, 16(1):115–137, 2004.

[169] Tommi Jaakkola, Mark Diekhans, and David Haussler. Using the fisher kernel method to detect remote protein homologies. In *ISMB*, volume 99, pages 149–158, 1999.

[170] Craig Saunders, Alexei Vinokourov, and John S Shawe-taylor. String kernels, fisher kernels and finite state automata. In *Advances in Neural Information Processing Systems*, pages 633–640, 2002.

[171] L. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 471–478. IEEE, 2011.

[172] Mark Iskander Boulos. Exploring the relationship of sleep-related movement disorders with cerebrovascular disease. Master's thesis, University of Toronto, 2014.

[173] Véronique L Roger, Alan S Go, Donald M Lloyd-Jones, Emelia J Benjamin, Jarett D Berry, William B Borden, Dawn M Bravata, Shifan Dai, Earl S Ford, Caroline S Fox, et al. Heart disease and stroke statistics-2012 update a report from the american heart association. *Circulation*, 125(1):e2–e220, 2012.

[174] A. M Adami, Misha Pavel, Tamara L Hayes, André G Adami, and C Singer. A method for classification of movements in bed. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 7881–7884. IEEE, 2011.

[175] Raffaele Ferri, Marco Zucconi, Mauro Manconi, Oliviero Bruni, Silvia Miano, Giuseppe Plazzi, and Luigi Ferini-Strambi. Computer-assisted detection of nocturnal leg motor activity in patients with restless legs syndrome and periodic leg movements during sleep. *Sleep*, 28(8):998–1004, 2005.

[176] Stephen D Pittman, Mary M MacDonald, Robert B Fogel, Atul Malhotra, Koby Todros, Baruch Levy, Amir B Geva, and David P White. Assessment of automated scoring of polysomnographic recordings in a population with suspected sleep-disordered breathing. *Sleep*, 27(7):1394–1403, 2004.

[177] Thomas C Wetter, Gerhard Dirlich, Jonathan Streit, Claudia Trenkwalder, Andreas Schuld, and Thomas Pollmächer. An automatic method for scoring leg movements in

polygraphic sleep recordings and its validity in comparison to visual scoring. *Sleep*, 27(2):324–328, 2004.