SIGNAL PROCESSING AND PATTERN RECOGNITION FOR NOCTURNAL
POLYSOMNOGRAPHY SLEEP STUDIES

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF

ELECTRICAL ENGINEERING AND

THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Hyatt Moore IV

November, 2013

This dissertation is online at: http://purl.stanford.edu/my525cr4403

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Emmanuel Mignot, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Krishna Shenoy

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Bernard Widrow

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Steven Woodward

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

Emmanuel Mignot, Principal Advisor

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation education for the degree of Doctor of Philosophy.

_____

Bernard Widrow, Co-Advisor

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

Krishna Shenoy

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

Steve Woodward

Approved for the University Committee on Graduate Studies

_____

Patricia J, Gumport, Vice Provost Graduate Education

ABSTRACT

Polysomnography (PSG)-based sleep studies are information-rich datasets, yet are only currently visually interpreted by technicians and medical doctors for retrieval of very basic parameters such as sleep staging of the electroencephalogram (EEG), EMG and EOG, scoring of respiratory events and periodic leg movements. Human sleep is a complex physiological process that plays a significant role in health and wellbeing and millions of PSG studies are now conducted every year in the US alone. There is a growing interest in extracting and validating new PSG biomarkers of disease or health in these datasets. Further, studies having shown many of these parameters to be strongly genetic, genomic studies of large PSG datasets may elucidate and uncover the extent and degree to which quantifiable alterations of sleep reach into our very lives.

Unfortunately however, the growth in algorithms and analytic frameworks necessary to examine sleep has not kept pace with the collection of this data, which is now reaching sufficient statistical power for meaningful genetic, clinical and epidemiological studies. The majority of algorithms and findings presented in the sleep literature are based and validated using only a small number of curated studies with very clean PSG data that are of limited use to others. Flexible, robust solutions are necessary to analyze sleep in large, diverse, and noisy PSG data sets coming from diverse laboratories and containing various sleep pathologies. This requires the design and validation of specialized algorithms that can be versatile enough to use in heterogeneous PSG recordings, and then, and as importantly, can be used to present findings in clear, meaningful ways to sleep researchers and clinicians. There is no silver bullet here, one single right method or algorithm for every problem, however there are simple strategies by which the work flow can be streamlined and remain flexible enough for a variety of different analyses.

In my work, I show how signal estimation and classification techniques, combined with visual interaction and receiver operating characteristics (ROC) studies, a commonly used statistical analysis method, can be used to investigate PSG based sleep studies (and measures) from large, diverse populations for genetic, medical, and clinically relevant purposes. I do this by considering four problems currently faced by the sleep research community and developing the signal processing, classification, optimization, and visualization measures needed for each. These problems include: (1) improving diagnostic criteria for narcolepsy using clinical and PSG measures; (2) selecting EEG power spectral density phenotypes for genome wide association (GWAS) (3) dependably detecting and classifying periodic leg movements (PLM) in sleep; (4) measuring rapid eye movements in patients with post traumatic stress disorder and major depressive disorder. ROC theory was extended and a combinatorial, iteratively bounded search method presented and used to optimize diagnostic testing (both parameter cutpoints and configuration) in a tool we called softROC. The Stanford EEG Viewer (SEV), a MATLAB toolbox, is developed to graphically analyze individual sleep studies and automate analysis of collections of sleep studies. The SEV provided the framework necessary to develop and optimize a new PLM classification algorithm, which implements a novel two pass, variable threshold calculation base on the current noise floor calculation. This algorithm was validated using human scored PLM data from a healthy cohort and a cohort with known sleep disorders. Time locked analysis of PLM with respiratory events illuminated the prevalence of PLM apart from respiratory events. I provide empirical evidence for improving PLM measures that are currently being put into clinical practice. Lastly, several eye movement algorithms are evaluated and a new tracking approach is developed which uses a wavelet denoised signal estimate of movement using two ocular channels (horizontal and vertical). Eye movements are characterized by activity and position and examined progressively by sleep cycle and elapsed sleep in combat veterans across four consecutive sleep studies.

Although these four examples represent only a fraction of what can and should be done with PSG datasets, they constitute a foundation for future work that will have to incorporate multiple algorithms in the analysis of sleep for clinical and research purposes. Today, when so much is still unknown about our health and sleep, and so much is known about processing data, it is imperative to engage medical experts and provide useful tools, methods, and theory to move forward together. The integration of signal processing methods with visual analytics has the potential of accelerating this partnership, as illustrated in this work.

and processing nocturnal polysomnography based sleep studies, particularly those from Stanford's Sleep Clinic. Oscar also introduced me to Steve Woodward.

Steve Woodward was tremendously helpful in working through many of the questions related to posttraumatic stress disorder and specific, technical issues related to the Combat Veteran Cohort. I enjoyed our many conversations on the subject and also the introduction to cardiac activity in relation to periodic leg movements. Thank you. I hope we will be able to continue working together in the future.

Krishna Shenoy remains an inspiration to. I appreciate his willingness to help me as an advisor after hearing just a short description of my proposed research, and for continuing with me in the face of some incredible challenges. Thank you.

And to my primary advisor Emmanuel Mignot, I give my heart felt thanks and appreciation for mentoring me throughout this journey. He challenged, corrected, encouraged, prodded, and continually motivated me to engage in my work, to think clearer, and mature as a researcher and person. He kept the same high standards for me as he kept for himself, and while it was not always easy, I sincerely appreciate and respect what he has done to guide and help me. I am proud to be counted as one of his students. Thank you.

Now I want to recognize specific chapter contributions made to this thesis by others.

Chapter 2: I would like to thank Olivier Andlauer and Emmanuel for their encouragement, support, and interest in improving the diagnosis of narcolepsy using PSG measures and introducing me to this research question. Their patience with explaining the problem and feedback were what led me to investigate new approaches to presenting receiver-operating characteristics discussed in this chapter. I would like to thank Noah Simon for his assistance with development of the convex optimization and bootstrapping approach to reduce bias. Finally, I would like to thank Ruth O'Hara and Simon Warby

movement and Restless Legs Syndrome (RLS) symptoms phenotypes in this chapter. Juliane Winkelmann's expertise in RLS and genetics was invaluable; thank you. And of course, Emmanuel Mignot who had the vision for this project and caught its "scent" well in advance, thank you.

Chapter 7: I would like to thank Steven Woodward for his mentorship, help, and the opportunity to work with him on research of sleep in post-traumatic stress disorder. I would also like to thank all of the combat veterans and participants of the sleep studies investigated in this chapter.

DEDICATION

To Geneva, we miss you. May the sleep of others be so sweet.

TABLE OF CONTENTS

17

LIST OF TABLES

22

26

LIST OF FIGURES

34

CHAPTER 1   INTRODUCTION

In this thesis I investigate several sleep related questions that require processing big data sets of polysomnogram (PSG)-based sleep studies, and develop interactive tools which process and prepare this data for meaningful biomarkers and present it to researchers in novel ways that engage their expertise and brings about helpful solutions. Instead of developing additional hardware, I develop interactive software solutions that rely on signal processing, pattern recognition and classification methods to prepare PSG data so it can be mined for specific biomarkers relevant to the investigation. These tools have been used to find better diagnostic criteria for narcolepsy, produce and then explore electroencephalography (EEG) power spectral density (PSD)-based phenotypes in sleep, automatically recognize periodic leg movements (PLM) in sleep and allow me to identify genotypes susceptible to PLM, and discover cross hemispheric EEG breakdown during sleep of combat veterans diagnosed with post traumatic stress disorder.

There are significant opportunities right now for electrical engineering to partner with sleep science, genetics, psychiatry, and epidemiology in evaluating PSG data collected in large diverse cohorts, for lesser known, studied, or well understood health-related issues that will lead to collective improvement of our lives. This thesis provides examples and applications of how this can be done successfully.

Two types of sleep study are examined in this work: nocturnal PSG (NPSG) and the multiple sleep latency test (MSLT). Evaluation of NPSG studies is the primary focus of this thesis and discussed first in Section 1.1, followed by the MSLT in Section 1.2. Section 1.3 introduces specific sleep disturbances that are examined in the later chapters, and Section 1.4 introduces the three sleep study cohorts that are investigated in the course of this dissertation. Motivating remarks for this dissertation are given in Section 1.5, and Section 1.6 concludes the chapter with an outline of the dissertation and its contributions.

1.1 NOCTURNAL POLYSOMNOGRAPHY BASED SLEEP STUDIES

NPSG based sleep studies are used mostly to determine if a patient's sleep is disturbed by sleep apnea, leg movements, insomnia, seizures or other abnormalities. How much sleep is disturbed could explain how tired a person is. During NPSG, the PSG monitors various physiological parameters and rhythms that can be used to differentiate healthy from abnormal or disordered sleeping (how someone manages to sleep wearing so much equipment remains a mystery to me). Electroencephalography or EEG is used for measuring brain function, electromyography (EMG) for muscle tone, electrooculagraphy (EOG) for ocular movement, and electrocardiography (ECG) for cardiac activity. These measurements are often taken from set locations and recorded to a specific *channel*. Nasal pressure, oximetry (ear lobe or finger), esophageal pressure, body position, airflow (from sensors at nose and mouth), snoring sounds, rib cage and abdominal movement, and intercostal EMG may also be measured during PSG.

A PSG *montage* refers to specific selection and configuration of channels to monitor these signals and often reveals a thoughtful balance between resources (e.g. channel inputs available for a recording system) and the line of investigation. For example, a high-density EEG array with hundreds of electrodes may be used for specific investigation and localization of brain activity to an evoked response, or multiple leads may be placed about the eyes to more precisely measure the position of each to determine the degree of synchrony.

Routine PSG sleep studies, however uses a set number of channels and are evaluated according to standardized, published criteria (e.g. 2007 American Academy of Sleep Medicine (AASM) scoring manual). I will discuss the specific montage configurations and criteria used for evaluating the different collection of NPSG sleep studies, or *cohorts* used in this dissertation in Section 1.4. For now, it is important to know that these channels are primarily used to (1) score sleep stages, (2) detect and quantify abnormal

breathing, (3) identify periodic leg movements (PLM), and (4) evaluate seizure or any unusual pattern.

1.1.1 SLEEP STAGES

Staging sleep is important as general basic knowledge, and for recognizing narcolepsy and insomnia. Technicians examine the EEG, EOG, and chin EMG following an NPSG study for specific changes, interactions, and patterns of brain, eye, and muscle tone activity in order to score sleep stages on 30 second intervals, called *epochs*, according to standard Rechtschaffen and Kales (R&K) criteria [1], slightly updated by the AASM[2].

Sleep is primarily categorized in two parts, non-rapid eye movement (NREM) sleep and rapid eye movement sleep (REM) sleep, which typically cycle back and forth between each other across a typical night's sleep with NREM sleep occurring first. Complete skeletal muscle loss or *atonia* is a distinguishing feature of REM sleep (REMS) that is typically determined by a flaccid chin EMG. The EEG gives a rough estimate of brain activity as measured through surface electrodes placed at various positions on the scalp. A typical EEG montage paces electrodes at the left central position along the top of the scalp (C-3 position), and at the left occipital position along the back of the scalp (O-1 position). Voltage from these electrodes is measured in reference to the right mastoid. A *wake-like* EEG and REM are also necessary for classifying or *scoring* REMS. REM is observed with EOG, which places surfaces electrodes about the outer canthi.

The EEG's PSD profile is helpful in staging sleep. Frequency ranges or *bands* of interest include delta (< 4 Hz), theta (4 to < 8 Hz), alpha (8 to < 13 Hz), and beta (13 to 40 Hz) [3]. Quiet wakefulness (Stage 0) refers to the stage prior to sleep onset when the patient's eyes are closed and there is a predominance of alpha band activity in the occipital (e.g. O-1) and central EEGs (e.g. C-3). Alpha activity diminishes or disappears with the transition to Stage 1 sleep which is further characterized by theta activity and slow rolling eye movements seen in the EOG. Characteristic EEG features of Stage 2 sleep include K-complexes (sharp negative deflection followed by slower positive rise) and spindles

(0.5 to 2.0 sec burst of 12-14Hz activity with a waxing and waning shape). Stage 2 indicates stable NREM sleep, which is deeper than Stage 1, but lighter than Stages 3 and 4. Stages 3 and 4, the deepest sleep, are often referred to as slow wave sleep because of the characteristic 0.5 to 2 Hz high amplitude ($> 75 \mu$V) EEG waveforms produced. In REM sleep (REMS, Stage 5 by convention here), the chin EMG drops to the lowest levels, rapid eye movements appear, and EEG activity looks similar to wakefulness. Portions of a study that cannot be determined or staged (i.e. 0-5) due to technical problems or artifacts are annotated (Stage 7 by convention here) and removed from analysis. In Chapter 4 I develop a visualization-based approach to explore these bands by sleep stage for markers to genotype.

The AASM recommends that several parameters, derived from sleep staging, be reported as part of the study. These include lights out clock time, lights on clock time, total recording time (TRT), total sleep time (TST), sleep latency (SLAT), REMS latency, wake after sleep onset (WASO), sleep efficiency (SEI), time in each stage, and sleep stage percentages. Lights-out clock time is the exact time that the subject is instructed to allow him or herself to fall asleep. Lights-on clock time is the exact clock time the subject is awakened. TRT is the temporal duration between lights out and lights on, while TST is the duration the subject spent sleeping (i.e. in Stage 1-5). SLAT is the elapsed time from lights out until first sleep onset and helps identify difficulty falling asleep. WASO is the duration spent awake between sleep onset and lights on. SEI is the ratio of TST to TRT, which can help characterize how well sleep is maintained. REMS latency is the time elapsed between sleep onset and the first epoch of REMS. Short REMS latency is a characterizing feature of narcolepsy.

1.1.2 SLEEP DISORDERED BREATHING

Sleep disordered breathing (SDB) has gained increasing notoriety because of its harmful effects and widespread prevalence. Abnormal breathing is seen by the presence of sleep apneas or hypopneas; respiratory events measured by oxygen saturation, breathing airflow, and respiratory effort of the rib cage/abdominal movements and/or intercostal EMG. The overlapping effects observed from SDB events and PLM are investigated in Chapter 5. In the cohorts used in this chapter apneas were defined as a cessation of airflow lasting ≥10 sec. Hypopneas were defined as a ≥ 30% reduction in nasal pressure signal excursions and associated ≥ 4% reduction in oxyhemoglobin saturation or arousal. This is close to 2007 AASM recommended (Medicare) or Chicago criteria for scoring hypopneas [4]. The apnea-hypopnea index (AHI) is a useful metric for characterizing the severity of SDB and is defined as the average number of apneas plus hypopneas per hour of objectively measured sleep. An AHI≥15 is used to identify subjects with SDB in Chapter 5, which investigates PLM in patients with and without SDB.

1.1.3 PERIODIC LEG MOVEMENTS

Electrodes placed on the left and right anterior tibialis (LAT/RAT) are used to determine significant leg movements (LM) and the presence of PLMs. These channels may also be combined into a single channel for LM and PLM classification. PLM are a phenomena or movement disorder defined by a minimum sequence of four consecutive leg movements (LMs) with intervals between 5 and 90 sec. The leg EMG, measured from surface electrodes placed on the left and or right anterior tibialis, are necessary to classifying the LMs, while respiratory effort must be observed to differentiate between idiopathic PLM and secondary PLM due to sleep disordered breathing (SDB). Figure 1.1 shows a picture of a man prepared for NPSG sleep study. The strain gauges around his chest and waist are designed to measure respiratory effort and expand and contract with the diaphragm and rib cage during normal breathing. Electromyography (EMG) measures muscle activity, such as teeth grinding or leg movements, or inactivity.

Monitoring periodic leg movements

Polysomnography

Leg movements are measured through electromyography, placing electrodes on the left and or right anterior tibialis muscles (outer part of the shin, close to the knee)

Human scoring

Observe polysomnogram, traditionally viewed as a secondary symptom to sleep disordered breathing and often not scored as diligently as respiratory events which have direct treatment.
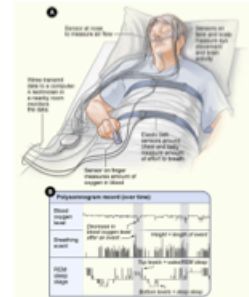
Figure 1.1: Polysomnography is useful for evaluating sleep in human subjects and abnormalities such as periodic leg movements. Technicians (lower right) evaluate the recorded PSG signals for sleep patterns and behaviors according to specific criteria. Monitoring periodic leg movements (PLM) requires electromyography via surface electrodes placed on the anterior tibialis to detect movement as well as bands and respiratory sensors (e.g. nasal cannula) to measure respiratory effort and oxygen levels in order to differentiate idiopathic PLM from PLM secondary to sleep disordered breathing.

The most current criteria for identifying PLM is the 2007 AASM Manual for Scoring Sleep. Here a significant leg movement (LM) as a period of 0.5-10 sec where LAT/RAT EMG activity exceeds 8 μV above baseline and then falls below 2 μV from baseline for 0.5 sec or longer[2]. A PLM is defined by the consecutive sequence of four or more LMs whose inter-movement intervals are between 5 and 90 sec. The 2007 classification excludes LMs 0.5 sec before, during or 0.5 sec after a respiratory event[2]. The PLM classifier in Chapter 5 modifies these rules' respiratory exclusion window to only exclude

LM detected at the edges of respiratory events: 5 seconds prior to onset until 0.5 sec after and then again 0.5 sec prior to offset 5.0 sec after. This adjustment was made based on empirical analysis of time locked respiratory events and leg EMG activity in patients stratified by an AHI of 15. Prior to the AASM 2007, the World Association of Sleep Medicine published their standards for scoring PLM in sleep (PLMS) and wakefulness (PLMW). The WASM 2006 gives considerably more attention to identifying and characterizing PLM particularly in regard to Restless Legs Syndrome or RLS. More will be said about PLM and RLS in Section 1.3.2.

1.1.4 AMBULATORY STUDY

NPSG may be performed at a sleep clinic or in the patient's home as an *ambulatory* study. Ambulatory studies have gained popularity because of their low cost and ease of use by the patient who can sleep in a familiar setting. These studies often serve as a good screening tool for SDB, but are often not adequate for more sophisticated research studies as recommended by current guidelines (e.g. AASM 2007 scoring manual). Only data from clinical NPSG are used in this work. These datasets and their specific NPSG configurations are explained in Section 1.4. Still, the continual growth in technology combined with the impetus to reduce medical costs makes the future of ambulatory studies promising.

1.2 MULTIPLE SLEEP LATENCY TEST

The MSLT is a sleep EEG test often used as a screening test for narcolepsy. The test presents five daytime opportunities for a patient to fall asleep within 20 minutes and measures the average time it takes to fall asleep (based on EEG changes) and the number of nap periods in which they entered REMS (SOREMP). A NPSG sleep study is also conducted the night prior to the MSLT to ensure the patient is well rested and not tired for other reasons (e.g. sleep apnea, sleep deprivation). In addition to the MSLT, REML is obtained during the NPSG and can assist the diagnosis. I show how to optimize the diagnosis of narcolepsy using a combination of these measures in Chapter 2.

1.3 SLEEP DISTURBANCES

In this thesis I give particular attention to sleep disturbances related to narcolepsy, PLM disorder and RLS, and PTSD. Medication is another sleep altering component that must be considered and accounted for. The following sections give more detailed description of these phenomenon and criteria used.

1.3.1 NARCOLEPSY

Narcolepsy is a REM sleep disorder classically characterized by excessive daytime sleepiness, sleep paralysis, cataplexy, and hypnagogic hallucinations. More recently, a subset of narcolepsy patients has been defined biochemically by measuring hypocretin (orexin), a wake-promoting peptide in the brain. In cases where the disorder is caused by the lack of hypocretin, hypocretin levels are greatly diminished in the cerebral spinal fluid (CSF), providing a gold standard [5]. These cases commonly have the symptom of cataplexy. Recent classifications of sleep disorders are now calling this subset of narcoleptic patients, "narcolepsy Type-I" or "narcolepsy with hypocretin deficiency".

As CSF can only be collected using lumbar punctures, a traumatic procedure, older tests such as the MSLT are still most often used to diagnose narcolepsy, but have not been optimized in relation to the new biochemical standard. Traditionally, the patient is diagnosed with narcolepsy if the mean sleep latency (MSL) during the five naps is less-than-or-equal eight minutes and two or more SOREMPs are observed [6].

1.3.2 PERIOD LEG MOVEMENTS AND RESTLESS LEGS SYNDROME

As mentioned earlier, PLMs are episodic, involuntary muscle contractions (myoclonus) that may occur during sleep or wake. RLS is often associated with PLMs, with four out of five patients diagnosed with RLS exhibiting PLMs[7]. However, PLMs can also occur without RLS symptoms. Because the movements may be accompanied by an arousal or sleep fragmentation, a PLM Index (PLMI) over 15 is likely to have an effect on an individual's overall health and wellbeing[8]. Little is known about the cause of PLMs or

their impact on daytime sleepiness or insomnia symptoms. A study review of cardiac risk for RLS and PLMS found associations between PLMS and congestive heart failure[9]. Additionally, patients with RLS were at higher risk for heart disease and hypertension[9]. Another, recent prospective study found increased mortality in men with RLS[10]. PLM are known to be associated with several other disorders and pathologies such as depression, cardiovascular disease, REM behavior disorder, narcolepsy, Parkinson's disease and multiple system atrophy[11-14].

Recent investigations have revealed genetic components of RLS. Most notably, 13 single nucleotide polymorphisms (SNP) were identified as RLS susceptibility markers in a genome-wide association study (GWAS) described in 2011[15]. I discuss this study further and investigate these SNPs for association to PLM in Chapter 6.

1.3.3 POSTTRAUMATIC STRESS DISORDER

PTSD is a disorder that develops in a subset of individuals who have experienced a traumatic event [16]. PTSD develops when there is a failure to recover from stress induced by such trauma [17]. Estimates have shown that 68% of children [18] and 55-70% of adults have experienced or witnessed a traumatic event [17, 19, 20]. That is, two-thirds of the general population is estimated to experience a traumatic event during their lifetime [21]. While experiencing a traumatic event does not imply PTSD, it is a necessary criterion for its development and it is striking to see the high percentage of people who are at risk for developing PTSD. Recent studies have shown that 5-20% of military members serving in combat zones suffer from PTSD [22]. Furthermore the prevalence of PTSD in returning combat veterans is likely to increase based on VA reports of 33.4% of Operations Iraqi Freedom/Enduring Freedom (OIF/OEF) veterans being diagnosed with PTSD [23]. According to epidemiological studies, 7.8-9.2% of the population suffers from PTSD with women being twice as likely to be effected as men [21, 24, 25]. The cost of this disorder to society and health care as whole is enormous

[26-28]. Notable symptoms of PTSD include anger, irritability, isolation, hypervigilance, context avoidance, flashbacks, suicidal ideation, and intrusive thoughts [29].

The correlation between PTSD and troubled sleeping is well established. In the past, sleep problems were viewed as a secondary symptom to PTSD [30], but recent literature has greatly elevated the significance of this relationship. Sleep problems are now recognized as a primary component of PTSD [31], with nightmares and insomnia included in diagnostic criteria for PTSD as established by the DSM-IV [32]. The correlation between sleep problems and exposure to a traumatic event and PTSD has been shown numerous times [33, 34]. Roughly 70% of those diagnosed with PTSD report co-occurring sleep problems [35]. Reported sleep disturbances associated with PTSD include sleep terrors, avoidance of sleep, periodic leg movement disorders, vocalizations and acting out behavior, and apnea [36, 37].

Poor sleep and anxiety are mutually aggravating with each negatively impacting the other. Increased anxiety during the day, for example, can lead to poor sleep at night, which in turn can lead to increased agitation and anxiety the following day. The spiraling relationship between PTSD and sleep disturbances is a dangerous threat to increased depressive severity, drug and alcohol abuse, poor quality of life, and suicidality [38-42]. While the relationship between sleep and PTSD has been repeatedly investigated using PSG-based sleep studies, these studies have provided both conflicting results and limited insight to the physiological underpinnings that drive this order. One reason for this may be the historical focus on global power spectral analysis of the EEG or broad sleep architecture measures like REM latency in only a single study. In Chapter 7 I examine discrete *phasic* eye movements in REMS and cross-hemispheric EEG coherence during sleep of combat veterans diagnosed with and without PTSD.

1.4 SLEEP STUDY COHORTS

I examine three sleep cohorts over the course of this dissertation. These are the Wisconsin Sleep Cohort (WSC), the Stanford Sleep Cohort (SSC), and the Combat Veteran Cohort (CVC).

1.4.1 WISCONSIN SLEEP COHORT

The WSC is a longitudinal study of sleep habits and disorders in the general population [43]. It was established in 1988 from a sample of employees of four state agencies in south central Wisconsin, USA, aged 30-60 years. NPSGs were initially stored on paper until digital collection began in 2000 and the PSG sleep studies were stored to disk at 100 Hz sampling in European Data Format (EDF), with file sizes ranging from 80 to 100 megabytes, and paired with scoring files of stages and events. Sleep was characterized using a 16-channel PSG recording system (16-channel Grass-Telefactor Heritage digital sleep system Model 15). Arterial oxyhemoglobin saturation was measured by pulse oximetry using a 3 sec averaging rate. Oral and nasal airflow were measured using thermocouples (ProTech). Nasal air pressure was measured with a pressure transducer (Validyne, Northridge). Thoracic and abdominal respiratory motions were measured with inductance plethysmography (Respitrace, Ambulatory Monitoring). These signals were used to identify SDB events. Electrodes placed on the left and right anterior tibialis (LAT/RAT) were combined into a single EMG channel and used to determine PLM.

The first PSGs for each subject that were performed between 2000 and 2004 were used in Chapter 4 and Chapter 5. Table 1.1 describes this WSC sample together, and stratified by an AHI of 15. The reason for stratifying by AHI becomes clear in Chapter 5, which describes the design and validation of an automatic PLM detection algorithm. Chapter 4 does not utilize the stratified groupings here, but focuses on the visualization of the EEG's power spectrum for the entire set. All PSGs available between 2000 and 2012 were used in Chapter 6 to show the genetic link to PLM apart from RLS symptoms.

Table 1.1 Wisconsin Sleep Cohort (WSC). Participants selected from 2000-2004 and stratified by apnea-hypopnea index (AHI) of 15.

	All (1,073)	AHI≤15 (738)	AHI>15 (264)
Demographics			
Age	56 ±0.24	55.3 ±0.28	57.8 ±0.49
Sex, Male (%)	53.2%	48.9%	63.3%
Clinical Data			
BMI (kg/m²)	31.7 ±0.22	29.8 ±0.22	35.4 ±0.45
AHI	12 ±0.50 (1004)	4.71 ±0.15	32.2 ±1.11
AHI > 15 (%)	26.5% (1004)	0.0%	100.0%
Polysomnogram			
TST (hour)	6.15 ±0.03	6.28 ±0.04	6.08 ±0.06
Stage 1 (%)	10.1 ±0.21	8.88 ±0.18	12 ±0.44
Stage 2 (%)	69.2 ±0.26	68.8 ±0.30	70.5 ±0.52
SWS (%)	4.06 ±0.18	4.49 ±0.22	2.96 ±0.27
REM (%)	16.5 ±0.19	17.6 ±0.21	14.4 ±0.39

1.4.2 STANFORD SLEEP COHORT

The Stanford Sleep Cohort (SSC) is a naturalistic sample of 760 successive patients (Table 1.2), including a wide range of sleep disorders, recruited to the Stanford Sleep Disorders Clinic and who had a nocturnal PSG from 1999-2007[44]. The SSC is used as a validation population sample for the automatic PLM detector introduced in Chapter 5. This cohort was particularly useful in the development of a robust detection algorithm because of the diversity of pathophysiologic conditions and poor signal conditions that had to be addressed and overcome.

Table 1.2 reports on summary statistics for the SSC, broken down by diagnostic category. The only exclusion criterion was the use of continuous positive airway pressure (CPAP) device for previously documented sleep apnea. PSGs were collected using Sandman Elite digital sleep software and Sandman SD32+ amplifiers. The Stanford Sleep Disorders Clinic protocol exceeds the AASM's clinical guidelines for the assessment of SDB by recording extra respiratory signals and using additional precision and processing.

Eighteen channels of information are recorded including: EEG, EOG, EMG of the submentalis muscle as well as the anterior tibialis muscles of each leg, electrocardiogram (ECG), snore using neck vibration, breathing effort using Braebon respiratory inductance plethysmography (RIP) system, airflow from Braebon PureFlo Duo Cannula and Nasal Pressure sensors and oxygen saturation (SpO2) through Finger PhotoPlethysmography Pulse Rate. EEG was recorded from conventional 10-20 system electrode sites using a 256 Hz sampling rate. EMG, ECG and snore signals used a sampling rate of 512 Hz while the sampling rates were 64 Hz for breathing effort and airflow and 4 Hz for SpO2, pulse rate. All AC channels used were hardware filtered between 0.1 Hz and 0.45 times the sampling rate and leg EMG channels were additionally high pass filtered at 10 Hz. In this cohort, PLM were not consistently scored thus not used.

Table 1.2: Stanford Sleep Cohort (SSC)

	All (760)	Delayed Phase Syndrome (14)	Insomnia (141)	Narcolepsy (19)
Demographics				
Age	45.9 ±0.52	36 ±4.67	46.2 ±1.29	40.2 ±5.09
Sex, Male (%)	58.8%	64.3%	45.4%	42.1%
Clinical Data				
BMI (kg/m²)	27.1 ±0.24 (741)	25.7 ±2.76	25.7 ±0.39 (140)	26.7 ±1.08
AHI	13.7 ±0.70	9.77 ±2.82	10.9 ±1.37	10.3 ±2.80
AHI > 15 (%)	33.0%	28.6%	27.0%	21.1%
PSG				
TST (hour)	6.12 ±0.04	6.18 ±0.34	6.13 ±0.10	6.69 ±0.24
Stage 1 (%)	11.7 ±0.35 (754)	7.57 ±1.37	9.74 ±0.61 (140)	12.4 ±2.31
Stage 2 (%)	62.7 ±0.45	60.6 ±2.75	63.4 ±0.90	57.5 ±3.25
SWS (%)	11.2 ±0.39 (622)	17.5 ±2.52 (13)	12 ±0.76 (116)	13.1 ±2.83 (18)
REM (%)	16.5 ±0.25 (753)	15.4 ±1.50	17.2 ±0.50 (139)	17.4 ±1.96

	REM Behavior Disorder (4)	Restless Legs Syndrome (23)	Sleep Disordered Breathing (607)	Other (39)
Demographics				
Age	59.3 ±3.84	49.2 ±3.10	45.5 ±0.59	42.2 ±2.62
Sex, Male (%)	100.0%	52.2%	58.5%	43.6%
Clinical Data				
BMI (kg/m²)	28.9 ±0.75	24.9 ±1.38	27.3 ±0.27 (599)	27.5 ±1.37
AHI	43 ±16.83	11.7 ±3.89	15.5 ±0.81	10.1 ±3.01
AHI > 15 (%)	75.0%	26.1%	36.7%	23.1%
Polysomnogram				
TST (hour)	6.01 ±0.66	5.96 ±0.26	6.14 ±0.05	5.91 ±0.22
Stage 1 (%)	20.8 ±5.74	13.6 ±2.51	11.7 ±0.38 (604)	11.7 ±1.53
Stage 2 (%)	56.2 ±6.47	60.6 ±2.81	62.6 ±0.50	61.4 ±2.65
SWS (%)	6 ±3.00 (3)	11.3 ±2.95 (17)	11.2 ±0.45 (498)	13 ±2.93 (33)
REM (%)	18 ±1.87	17.4 ±1.55	16.5 ±0.27 (600)	16.1 ±1.25 (38)

Data are mean ± Standard Error Mean, or percentage. The number of subject used for calculations are shown in parentheses. AHI: Apnea hypopnea index; REM: Rapid Eye Movement; TST: Total sleep time; SWS: Slow wave sleep (stages 3 and 4 combined); BMI: Body mass index

1.4.3 COMBAT VETERAN SLEEP STUDIES

Combat-related PTSD inpatient subjects were recruited from the Specialized Inpatient PTSD Unit at the Veterans Administration Medical Center, Palo Alto, California. Potential subjects with evidence of medical disease and/or chronic pain that could influence sleep were excluded, as were individuals with risk factors for obstructive sleep apnea (frequent snoring, obesity, or partner reports of interrupted breathing during sleep). Subjects were further excluded from analysis if, on any laboratory night, there was evidence of sleep apnea as indicated by an apnea hypopnea index (AHI) greater than 10 events per hour or a periodic limb movement arousal index of greater than 10 events per hour. Subjects were excluded if there was a recent history of heavy alcohol use (intake greater than 5 oz/day for any 30-consecutive day period during the preceding 6 months). All included subjects reported being abstinent from alcohol or illicit drugs for at least 30 days before hospitalization. These subjects were also free of prescription psychotropic medication on admission to the inpatient program and remained off such medications before testing. Mean duration of hospitalization before testing was 35 days. Finally, no subject was medicated with a selective serotonin reuptake inhibitor within 80 days of testing. Table 1.3 presents summary information for this *Combat Veteran Cohort* (CVC).

The final sample consisted of 148 Vietnam era combat veterans: 32 healthy controls, and 116 PTSD inpatients meeting DSM-IV criteria for current PTSD as the primary diagnosis determined through administration of the Structured Clinical Interview for the DSM-III-R (SCID; American Psychiatric Association 1994) or the Clinician-Administered PTSD Scale (CAPS; Blake et al 1995). Subjects ranged in age from 42 to 48 years (mean = 45.2, SD = 3.1). Of the PTSD sample, 72 subjects were currently diagnosed with major depressive disorder. All subjects gave informed consent.

Table 1.3: Combat Veteran Cohort by psychiatric diagnosis: combat with posttraumatic stress disorder (PTSD) and major depressive disorder (MDD), combat with PTSD, combat only, and all.

Demographics/Groups	Combat with PTSD and MDD	Combat with PTSD
Male	97.18% (69 of 71)	89.47% (34 of 38)
Female	2.82% (2 of 71)	10.53% (4 of 38)
AHI less than 15	100.00% (71 of 71)	100.00% (38 of 38)
Combat only	100.00% (13 of 13)	100.00% (13 of 13)
PTSD only	100.00% (38 of 38)	100.00% (38 of 38)
PTSD and MDD	100.00% (71 of 71)	100.00% (71 of 71)
Total Sleep Time (TST)	5.23 ±0.13 (71)	5.46 ±0.15 (38)
TST 1	0.14 ±0.01 (71)	0.15 ±0.01 (38)
TST 2	0.54 ±0.01 (71)	0.51 ±0.02 (38)
TST SWS	0.08 ±0.01 (66)	0.10 ±0.02 (32)
TST NREM	0.77 ±0.01 (71)	0.74 ±0.01 (38)
TST 5	0.23 ±0.01 (71)	0.26 ±0.01 (37)
Antidepressants	16.90% (12 of 71)	10.53% (4 of 38)
REM suppression	18.31% (13 of 71)	10.53% (4 of 38)
Fluoxetin	7.04% (5 of 71)	2.63% (1 of 38)

Demographics/Groups	Combat only	All
Male	89.47% (34 of 38)	94.26% (115 of 122)
Female	10.53% (4 of 38)	5.74% (7 of 122)
AHI less than 15	100.00% (38 of 38)	100.00% (122 of 122)
Combat only	100.00% (13 of 13)	100.00% (13 of 13)
PTSD only	100.00% (38 of 38)	100.00% (38 of 38)
PTSD and MDD	100.00% (71 of 71)	100.00% (71 of 71)
TotalSleepTime	5.46 ±0.15 (38)	5.26 ±0.09 (122)
TST 1	0.15 ±0.01 (38)	0.14 ±0.01 (122)
TST 2	0.51 ±0.02 (38)	0.53 ±0.01 (122)
TST SWS	0.10 ±0.02 (32)	0.09 ±0.01 (110)
TST NREM	0.74 ±0.01 (38)	0.76 ±0.01 (122)
TST 5	0.26 ±0.01 (37)	0.24 ±0.01 (121)
Antidepressants	10.53% (4 of 38)	13.11% (16 of 122)
REM suppression	10.53% (4 of 38)	13.93% (17 of 122)
Fluoxetine	2.63% (1 of 38)	4.92% (6 of 122)

Data are mean ± Standard Error Mean, or percentage. The number of subject used for calculations are shown in parentheses.

Subjects slept three or four nights in the sleep laboratory located immediately adjacent to the inpatient unit. Scheduling was arranged to accommodate subjects' typical bedtimes. Inpatient subjects terminated their sleep at will, but not later than 6:00 AM, the standard wake-up time for the inpatient program.

The recording montage included two channels of bipolar EOG and four channels of scalp EEG (F-3, F-4, Cz, and Pz referenced to linked mastoids). Other channels, not pertinent to this study, were also scored. EOGs and EEGs were filtered to a 0.3-Hz to 30-Hz bandwidth. Electromyograms were filtered to a 30-Hz to 300-Hz bandwidth, then rectified and integrated over a 20-msec time constant. After conditioning, all physiologic data were digitized at 125 Hz and streamed to disk. A trained sleep technician manually scored sleep stages in 30-s epochs according to standard R&K criteria.

1.5 MOTIVATION

Sleep researchers today are faced with the enormous and continually growing problem of investigating the plethora of information available in NPSG sleep studies for clinically significant biomarkers; markers that will reveal trends and behaviors that can be used to understand our own sleep and how it, and by consequence our very lives, may be improved. A recent report estimated 50-70 million adults in the United States suffer from a sleep or wakefulness disorder and found those with poor sleep more likely to suffer from chronic diseases like hypertension, depression, diabetes and obesity[45]. The number of sleep studies performed each year has soared with the growing awareness of the risks of chronic insomnia. Over a million studies are performed yearly in the U.S. with NPSG being used as the gold standard for most diagnoses. However, while PSG gives a physiologically rich snapshot into each person studied and so much of it now exists, clinicians often evaluate it at broad levels only, quickly searching for indications of sleep apnea or other disorder in question and taking a few cursory notes to render diagnosis and/or treatment. I do not mean to say that nothing is being done here. Indeed, people are often receiving the help and care they need, but there is an aspect of sleep

research that remains in its infancy and which bears attention. If separate, large collections of NPSG studies exist in sleep cohorts – cohorts that include survey responses, medication records, genetic information, psychiatric diagnoses, and various demographic information for each participant – then why is there a current dearth of literature explaining automatically derived PSG characterizing features of such data?

I believe the answer to this question lies in the host of challenges that must be overcome to develop robust solutions that can interpret and explore large collections of diverse PSG data. First, a large number of people must be found who are willing to participate in a large sleep study or *cohort*, which requires various medical exams and surveys of their history, habits, and health. The sleep study must be conducted and then normalized in some fashion to account for the variety of different channel identifiers, PSG configurations, and storage formats, which often have to be reverse engineered before they can be accessed. The criteria used to manually evaluate and *score* sleep studies changes with time and takes creativity and insight to salvage in light of the most recent guidelines. Furthermore, logistical and technical challenges of de-identifiying the PSGs exist to protect patient privacy while simultaneously organizing them in a centralized, accessible manner.

Once the data is collated the work of developing signal processing methods, pattern recognition algorithms, and automated computational framework design can begin. Methods to reliably extract sleep biomarkers can be developed and in turn used for direct analysis, statistical modeling in conjunction with genetic/genomic data, or visually mined with other exploratory tools.

1.6 CONTRIBUTIONS AND THESIS OUTLINE

There is currently a lack of tools in the sleep field to robustly process, parse, and present PSG data in flexible, visual ways that engage sleep scientists and brings their expertise to the forefront of sleep related problems facing us today. The contributions of my thesis

consist primarily in the development of such tools and the results obtained by applying them to different problems currently faced in the sleep research community. Apart from the introduction and conclusion, each chapter in this thesis represents a contribution to research.

Chapter 2 introduces receiver-operating characteristics and a user directed, iteratively bounded, combinatorial approach for optimizing cut-points of diagnostic tests. These methods are implemented in the software program, *softROC*, and presented in the setting of determining optimal diagnostic criteria of narcolepsy with and without cataplexy using PSG measures and hypocretin levels.

Chapter 3 transitions to micro or *quantitative* PSG analysis, and discusses the *SEV* (Stanford EEG Viewer). The SEV is a multipurpose MATLAB toolbox that I developed during this thesis to prototype pattern recognition and classification algorithms of PSG data and automate their use across our cohorts. The SEV provides the groundwork necessary for the research presented in Chapters 4 – 7.

Chapter 4 considers the problem of investigating the sleep EEG's power spectrum in patients of the Wisconsin Sleep Cohort for phenotypes to evaluate by genome-wide association study. Power spectral output of two EEG channels in 1,836 NPSG studies is processed by the SEV and then consolidated and visually presented based on design decisions from a small user design study with several sleep researchers.

Chapter 5 begins analysis of time series biomarkers and introduces the problem of automatically identifying PLMs in sleep. Motivation for this topic is provided, as are the challenges to automatically identifying PLMs in noisy and variable leg EMG recordings. Cardiac interference, noise floor changes, transient or spurious activity, low signal strength, and respiratory-related events which interfere with PLM classification are presented, resolved, and validated using gold standards from two separate sleep cohorts.

In Chapter 6, I utilize the automatic PLM detector to investigate 13 single nucleotide polymorphisms (SNP), associated with RLS, for susceptibility to PLM apart from RLS symptoms in the Wisconsin Sleep Cohort.

A mix of cross spectral EEG analysis and time series phasic EOG events is presented in Chapter 7 with application to sleep of combat veterans diagnosed with and without post traumatic stress disorder as measured during three to four night sleep studies. In this chapter, I first examine the utility of automated detection algorithms for analyzing phasic REMS in combat veterans diagnosed with PTSD and matched controls. Four previously published and validated eye movement detection algorithms, and two unpublished eye movement detection algorithms, are optimized with reference to our dataset using adaptive filtering and wavelet transform techniques which remove noise and increase rapid eye movement detections during REMS. Horizontal and vertical phasic-REM activities are compared longitudinally between patient groups by visit sequence and NREM/REM cycle. A new approach to quantifying ocular activity, *EOG quadrant analysis*, is presented which takes advantage of the EOG montage used in this data set. Lastly, I test an ethologically-driven hypothesis that PTSD is associated with reduced inter-hemispheric EEG coherence during sleep.

Concluding remarks are given in Chapter 8.

CHAPTER 2   EXPLORING MEDICAL DIAGNOSTIC PERFORMANCE USING INTERACTIVE, MULTI-PARAMETER SOURCED RECEIVER OPERATING CHARACTERISTIC SCATTER PLOTS

This chapter introduces a method I developed for conducting a bounded search of receiver operating characteristic space for optimal diagnostic test parameter configurations. It is presented in the context of diagnosing narcolepsy from a selection of PSG measures using a program called *softROC* (see the supplementary section of this for further details on obtaining and using softROC).

I would like to thank Olivier Andlauer and Emmanuel for their encouragement, support, and interest in improving the diagnosis of narcolepsy using PSG measures and introducing me to this research question. Their patience with explaining the problem and feedback were what led me to investigate new approaches to presenting receiver-operating characteristics discussed in this chapter. I would like to thank Noah Simon for his assistance with development of the convex optimization and bootstrapping approach to reducing bias. Finally, I would like to thank Ruth O'Hara and Simon Warby for their helpful reviews of the initial manuscript, which is currently in review.

Two other manuscripts were generated during this time that used softROC to investigate narcolepsy. I was second author to Olivier Andlaurer on these manuscripts, and discuss portions of our findings in this chapter. The first manuscript, "Predictors of Hypocretin (Orexin) Deficiency in Narcolepsy Without Cataplexy" is published in the journal *SLEEP*. The second manuscript, "Nocturnal Rapid Eye Movement Sleep Latency for Identifying Patients with Narcolepsy/Hypocretin Deficiency" is published in the *Journal of American Medical Association Neurology*.

2.1 INTRODUCTION

Diagnoses are made based on the presence of symptoms or the results of biological or physiological tests. No test is perfect, and for continuous variables, it is essential to select an optimal cut-off versus a gold standard. Depending on the application, one may select cut-offs with equal specificity and sensitivity, or instead favoring one characteristic at the cost of the other.

ROC curves are typically used to visualize sensitivity and specificity tradeoffs for various diagnostic cut-offs. Points located closer to the ideal test point (100% sensitivity, 100% specificity) are often seen as good candidates for cut-off values. However, sensitivity and specificity by themselves provide little meaning, as it is possible to make a test that achieves 100% sensitivity or 100% specificity by simply always giving a positive test result or always giving a negative test result. In an important variation, the quality Receiving Operating Characteristic curves (qROC), sensitivity and specificity values are remapped to Kappa values, or quality indices, which provide a measure of the quality of the original test ROC (tROC) values. These weighted Kappa coefficients, the original ROC values adjusted to be 0% for a random test and 100% for a perfect test, make it easy to identify and lay claim to the test with optimal sensitivity and specificity [46].

While several ROC software packages exist, few are dedicated to medical diagnostic discovery or incorporate the ability to group and combine multiple variables, a critical feature optimizing medical diagnostic when multiple parameters are involved. One review of eight ROC programs, commercially available and freeware, showed mixed results [47]. While statistically sound, the programs covered (e.g. MedCalc and Chicago University's Metz ROC Software) were described as unfriendly or overly complicated to use because of the interface or statistical background required. Little validation of ROC results was provided from these programs in terms of generalization and bias, and none plotted qROCs.

Another ROC program, not included in this review, which does incorporate quality indices, is ROC5 [48]. This program produces decision trees aimed at providing clinical practitioners with a plan of sequential tests to follow based on the outcome at each stage of the tree. This program was the closest to helping us achieve our goals. Decision trees can be helpful in determining the order of tests to provide to a patient and are particularly appropriate when tests must be performed in a sequential manner due to increasing costs. However, the nature of decision trees to follow one branch while rejecting all others, limits evaluation and exploration of simultaneous possibilities when multiple tests are simultaneously available. The limitations we saw in ROC5 were its inability to draw multiple ROC plots and its complex interface. Also, it was not possible to explore the entire space of simultaneous and sequential possibilities, relying solely on a hierarchy of cut-offs.

Considering this need, we developed softROC, a MATLAB-based software package for exploratory ROC analysis. Specifically, softROC provides a GUI that allows users to quickly configure candidate diagnostic criteria combinations and evaluate them for optimal performance using test and quality ROC metrics. The software uses either bootstrapping or training-with-validation techniques to provides generalizability of selected diagnostic criteria to other populations. The developer chose MATLAB because of its relatively straightforward, stable, development environment suitable for both statistical and graphically-interactive based software programs. softROC is available in the supplementary material section which also includes the instruction manual and tutorial dataset. We provide background on the statistical methods used and an overview of narcolepsy diagnosis in Section 2.2. Section 2.3 covers the design and implementation of softROC for investigating diagnostic tests. Discussion of softROC's application, limitations, and extension are covered in Section 2.6, followed by concluding remarks in Section 2.7.

2.2 BACKGROUND

Statistical background of receiver operating characteristics and Boolean algebra are presented in this section. Information on narcolepsy and its diagnosis is presented in Section 2.3.

2.2.1 RECEIVER OPERATING CHARACTERISTICS

The contingency table, or confusion matrix, shown in Table 2.1 contains the collection of possible outcomes, as a percentage, for a predicted medical diagnosis of a patient and the true diagnosis, or pathology, as revealed through an accepted gold standard. The *confusion* of Table 2.1 lies on the diagonal where the evaluation is different from the "ground truth" - a patient with true disease is missed (i.e. false negative) or one without disease is wrongly diagnosed with it (i.e. false positive).

Table 2.1: The contingency table captures the four possible outcomes when comparing a diagnosis based on medical testing to its gold standard (*truth*): true positive (TP), false negative (FN), false positive (FP), and true negative (TN). The values, given as fractions of the whole, sum to produce quality (Q) and prevalence (P) which are necessary to calibrate ROC values. Medical testing is frequently defined as positive or negative depending on a threshold biological, symptomatic or physiological value. For example, abnormally high fasting blood sugars used to diagnose diabetes are typically defined as above 126 mg/dl (7.0 mmol/L).

		Medical Test		
		+	-	
Ground Truth	+	TP	FN	P
(Gold Standard)	-	FP	TN	$P' = 1\text{-}P$
		Q	$Q' = 1 - Q$	1

Frequently, medical tests give a continuous value that needs to be dichotomized as positive and negative for practical reasons based on a threshold or cut off value. Altering this threshold value modifies specificity and sensitivity which can be optimized for a given application. ROCs are frequently used to evaluate these trade-offs. These and other measures, which are derived from Table 2.1 and implemented in softROC, are

given in Table 2.2. The derivations require contents of Table 2.1 to be given as fractions of the total count ranging in value from 0.0 to 1.0.

Table 2.2: List and Definition of softROC statistics obtained from Table 2.1.

Term	Notation	Definition
Sample Size	N_0	
True Positive	TP	$1/N_0 \cdot \sum_{i=1}^{N_0}(TrueDiagnosis_i^+ \cap Test_i^+)$
False Negative	FN	$1/N_0 \cdot \sum_{i=1}^{N_0}(TrueDiagnosis_i^+ \cap Test_i^-)$
False Positive	FP	$1/N_0 \cdot \sum_{i=1}^{N_0}(TrueDiagnosis_i^- \cap Test_i^+)$
True Negative	TN	$1/N_0 \cdot \sum_{i=1}^{N_0}(TrueDiagnosis_i^- \cap Test_i^-)$
Prevalence	P	TP + FN
Quality	Q	TP + FP
Sensitivity	SE	TP/P
Specificity	SP	TN/P$'$
Positive Predictive Value	PPV	TP/Q
Negative Predictive Value	NPV	TN/Q$'$
Efficiency	EFF	TP + FN
Quality Index 1,0	$\kappa(1,0)$	(SE-Q)/Q$'$
Quality Index 0,0	$\kappa(0,0)$	(SP-Q$'$)/Q
Cohen's Kappa	$\kappa(0.5,0)$	$(PQ' \cdot \kappa(1,0) + P'Q \cdot \kappa(0,0))/(PQ' + P'Q)$
Chi-square	χ^2	$N_0 \cdot \kappa(1,0) \cdot \kappa(0,0)$

ROC curves plot sensitivity versus specificity for different thresholds used to classify a patient positively or negatively for diagnosis. When evaluating multiple curves, points along the outermost curve -- the ROC convex hull -- are superior to any along the other curves. The area under the ROC curve (AUC) offers insight into a diagnostic test's overall ability to discriminate between positive and negative cases and is equivalent to the Wilcoxon test of ranks [49]. A random ROC curve, which places a straight line from (0,0) to (1,1), has an AUC of 0.5. A useful diagnostic test will have an AUC between 0.5 and 1.0.

Alternative quality indices have also been proposed to optimize the selection of cut-offs. In this case, a qROC is created. Kraemer points out that a sensitivity of 0.99 is meaningless on its own. The measures need to be placed in context. A random test, with

sensitivity and specificity of 0.5 has $\kappa(1,0)$ and $\kappa(0,0)$ values of 0.0. A test which identifies the gold standard diagnosis perfectly has $\kappa(1,0)$ and $\kappa(0,0)$ values of 1.0. Efficiency, when similarly recalibrated, produces $\kappa(0,0.5)$ – Cohen's Kappa [46]. The use of quality indices rather than specificity and sensitivity produces qROC plots, which, although less frequently used, have the advantage of showing the best cut-off overall independent of the application.

2.2.2 BOOLEAN ALGEBRA

Boolean algebra provides a simple framework to combine binary data (Table 2.3). Continuous data can easily be broken into two groups by applying a comparison operation (e.g. $<, \leq, >, \geq, =, \neq$) at the desired split. Dichotomizing data and combining with Boolean operations this way produces a useful framework for flattening an otherwise hyper-dimensional diagnostic decision space into the visible 2D ROC realm.

Table 2.3: Boolean operations on binary data. Binary value pairs combine to form a single binary result based on the logic operator used.

		AND $x \wedge y$	OR $x \vee y$	NOT $\neg x$
x	y			
0	0	0	0	1
0	1	0	1	1
1	0	0	1	0
1	1	1	1	0

The choice of Boolean operator can have significant effect on the relative weight that each subtest carries toward forming the final result. Evaluations combined with AND carry less strength together than they do apart, while the opposite is true for evaluations combined with OR. These rules' impact on the combined test's sensitivity and specificity must be assessed by performing each test separately and then combining and comparing the predicted result to the gold standard. Kraemer points out that there is no easy mathematical shortcut to reach this result [46]; it is a computational problem and must be

calculated directly. Recognizing the utility of Boolean logic and ROCs, and their familiarity with medical practitioners, we developed softROC to provide the flexibility in diagnostic test design while off-loading the computational burden.

2.3 IMPLEMENTATION AND THE EXAMPLE OF NARCOLEPSY

Medical diagnosis, positive or negative, is often obtained after using not one test, but a combination of tests. Furthermore, these *subtests* may take on a range of values that must first be dichotomized at known cut-points (e.g. low blood sugar, fever, high white blood cell count, etc.) before being combined into the final diagnosis. In this section, we used the example of narcolepsy, a diagnosis that can be made based on a gold standards, (low cerebral spinal fluid (CSF) hypocretin-1) and physiological measures of sleep abnormalities (with each potential diagnostic variable being used as a subtest). *softROC* could as easily be used to determine cut-offs for a combination of biological measures (dementia psychological test thresholds, CSF amyloid levels) versus a pathologically-based defined gold standard for another disease (Alzheimer disease based on postmortem examination).

2.3.1 NARCOLEPSY: DEFINITION AND DIAGNOSTIC PROCEDURES

Narcolepsy was introduced and described in Section 1.3.1. To reiterate briefly, it is a rapid eye movement (REM) sleep disorder classically characterized by excessive daytime sleepiness, sleep paralysis, cataplexy, and hypnagogic hallucinations. Furthermore, hypocretin levels measured from cerebral spinal fluid (CSF) are greatly diminished and provide a gold standard [5]. These cases commonly have the symptom cataplexy. The MSLT (described in Section 1.2) is often used to screen for narcolepsy and to avoid the traumatic lumbar puncture that is necessary to obtain CSF. The patient is diagnosed with narcolepsy if the mean sleep latency (MSL) during the five naps is less-than-or-equal eight minutes and two or more onset into rapid eye movement periods (SOREMPs) are observed [6]. In addition to the MSLT, REM latency (REML), the number of minutes

elapsed from sleep onset to the first REM period is obtained during the prior night's NPSG and can assist the diagnosis.

To investigate how sleep tests can be best used to predict narcolepsy/hypocretin deficiency, we used softROC to optimize diagnostic cut-offs for PSG REM latency and MSLT MSL and number of SOREMPs. The full analysis, which uses multiple datasets, is detailed in [50] and [44]. As a result of this analysis, short REM latency during NSPG is now likely to be incorporated in newly revised AASM diagnostic criteria for narcolepsy used by clinicians worldwide. In this paper, we are using and providing one exemplar dataset from this publication as a training dataset for users interested in using softROC for other applications.

2.3.2 SOFTWARE CONFIGURATION FOR SEARCHING COMBINED BIOMARKERS IN MEDICAL DIAGNOSIS

Figure 2.1: softROC configured to analyze the diagnostic performance of narcolepsy using multiple subtests. Menu choices for subtests are populated using the input data file's header column.

Figure 2.1 shows softROC in one of the preliminary configurations we used to examine diagnostic tests of narcolepsy and which will be referred to throughout this section. The accompanying Microsoft Excel input file includes polysomnogram measures Mean Sleep Latency and number of SOREMPs on the MSLT and NSPG REM latency as shown in Figure 2.2. The provided tutorial gives specific instruction on configuring softROC in this manner and utilizing its various features.

2.3.2.1 Input

	A	B	C	D	E	F	G	H	I
1	Diagnosis	PSG.REML	SOREMPs	MSL	Test4	Age	Gender	ID	
2	1.00	53.30	3.00	1.61	4.83	24.00	M	1.00	
3	1.00	3.50	2.00	1.5	3.00	32.00	F	2.00	
4	1.00	48.00	2.00	0	0.00	31.00	F	3.00	
5	1.00	6.00	3.00	3.4	10.20	43.00	M	4.00	
6	1.00	1.00	4.00	0.8	3.20	45.00	M	5.00	
7	1.00	121.00	-1.00	-1	1.00	29.00	M	6.00	
8	1.00	8.50	4.00	1.4	5.60	46.00	M	7.00	
9	1.00	0.00	4.00	3.5	14.00	51.00	F	8.00	
10	1.00	98.10	5.00	1.8	9.00	29.00	M	9.00	
11	1.00	5.00	-1.00	-1	1.00	32.00	M	10.00	
12	1.00	3.50	4.00	2.3	9.20	33.00	M	11.00	
13	1.00	1.00	3.00	1	3.00	37.00	F	12.00	
14	1.00	4.00	4.00	1.6	6.40	22.00	M	13.00	
15	1.00	3.50	4.00	4.4	17.60	34.00	F	14.00	
16	1.00	223.50	1.00	3.1	3.10	28.00	F	15.00	
17	1.00	21.00	5.00	0.4	2.00	38.00	F	16.00	
18	1.00	4.00	3.00	0.12	0.36	39.00	F	17.00	
19	1.00	0.50	3.00	1	3.00	31.00	M	18.00	
20	1.00	84.00	1.00	7	7.00	42.00	M	19.00	
21	1.00	2.00	4.00	3.1	12.40	38.00	F	20.00	
22	1.00	1.50	2.00	3.75	7.50	25.00	M	21.00	
23	1.00	128.00	4.00	2.7	10.80	42.00	M	22.00	

Figure 2.2: Contents of input Excel file with de-identified patient information, subtests, and ground truth diagnosis. The file is laid out similarly to the ones we used in our

investigation of narcolepsy. softROC uses the header of each column to identify the subtest or gold standard grouping. Each subsequent row represents the patient's measured result. Data may be removed using exclusion rules in the settings. Here, patient data in row seven is excluded from analysis because it contains negative values.

The input file, once selected, provides all of the information necessary to configure a diagnostic test. It presents the file header names for selection as subtest or gold standard (when data is dichotomous) and resizes as necessary. In the exemplar excel file (Figure 2.2), narcolepsy is defined biochemically by low CSF hypocretin (gold standard diagnosis: 0,1) in each patient and age matched control.

2.3.2.2 Cut-point ranges

Candidate cut-points are generated by establishing a range and step-size to iterate through from *START* to *STOP*. In the example provided in Figure 2.1, PSG REM latency cut-offs are calculated for each REM latency from 1 to 200 minutes, in 1-minute increments.

2.3.2.3 Dichotomize

Each cut-point in the established range is split using a comparison operator. Figure 2.1 shows the ≤ and ≥ comparison operators selected for use. For the PSG REM latency subtest, measured patient values less-than-or-equal to the cut-points are classified negative for narcolepsy while measured values above these cut points are classified positively for diagnosis.

2.3.2.4 Combine

Subtests are combined using the logic operators AND, OR, AND NOT, and OR NOT before they can be collectively compared to the gold standard. Nesting of subtests is also possible in softROC. For example, the MSL and SOREMPs results in Figure 2.1 are first combined using AND before, subsequently, being combined with PSG.REML subtest result using the OR operation.

2.3.3 COMPUTATION

The softROC configuration shown in Figure 2.1 can be viewed algorithmically as

$$\text{Medical Evaluation} = \begin{cases} & \text{PSG.REML} \leq [1.0:200.0] \\ & \text{OR} \\ & (\text{MSL} \leq [1.0:20.0] \text{ AND SOREMPS} \geq [1.0:5.0]) \end{cases}$$

and interpreted in words as "For each combination of values in the established ranges, make a positive diagnosis if *either* the measured PSG.REML is less than the current diagnostic test value *or* if *both* the measured MSL is less-than-or-equal to its current diagnostic value *and* the measured SOREMPs is greater-than-or-equal to its current diagnostic test value." The number of cut-points evaluated grows combinatorially as a Cartesian-product with each subtest added. There are 200,000 cut point combinations in the current diagnostic configuration. Each combination is evaluated and compared against the selected gold standard's positivity marker (i.e. a Boolean value) to generate the contingency table and produce the statistics outlined in Table 2.2.

Let $D_n(\bullet)$ represent the diagnostic classification function, evaluated for some patient n where $n \in [1, N]$ (N is the sample size), whose output is in $\{0,1\}$ with 0 meaning a negative diagnosis and 1 meaning a positive diagnosis. The diagnostic classification function is defined in softROC, and varies in definition and arguments with each unique configuration. Using the current example, let $PSG.REML(n)$, $MSL(n)$, and $SOREMPs(n)$ be the corresponding measurements for n, the current patient, as obtained from the input file (Figure 2.2). The diagnostic classification function can be defined as

$$D_n(i,j,k) = v(n,i) \wedge \{\phi(n,j) \vee \psi(n,k)\}$$
$$v(n,i) = PSG.RML(n) \leq i$$
$$\phi(n,j) = MSL(n) \leq j$$
$$\psi(n,k) = SOREMPs(n) \geq k$$

Defining G_n as the gold standard diagnosis for patient n (with similar $\{0,1\}$ output and interpretation of $D_n(\bullet)$) we generate the contingency table values as follows

$$TP = \frac{1}{N} \sum_{n=1}^{N} (D_n \vee G_n)$$

$$FP = \frac{1}{N} \sum_{n=1}^{N} (D_n > G_n)$$

$$TN = \frac{1}{N} \sum_{n=1}^{N} (\neg D_n \vee \neg G_n)$$

$$FN = \frac{1}{N} \sum_{n=1}^{N} (D_n < G_n)$$

In order to optimize our implementation, we reshape the standard contingency matrix into a single contingency row vector C with elements in the order of [TN, FP, FN, TP]. A 0-base indexing scheme is then to generate the C with a simple, iterative update

$$C(i_n) = C(i_n) + 1$$

where i_n is the contingency vector index for patient n defined as

$$i_n = D_n + 2 \cdot G_n \quad n \in [1, N]$$

2.3.4 PLOTTING
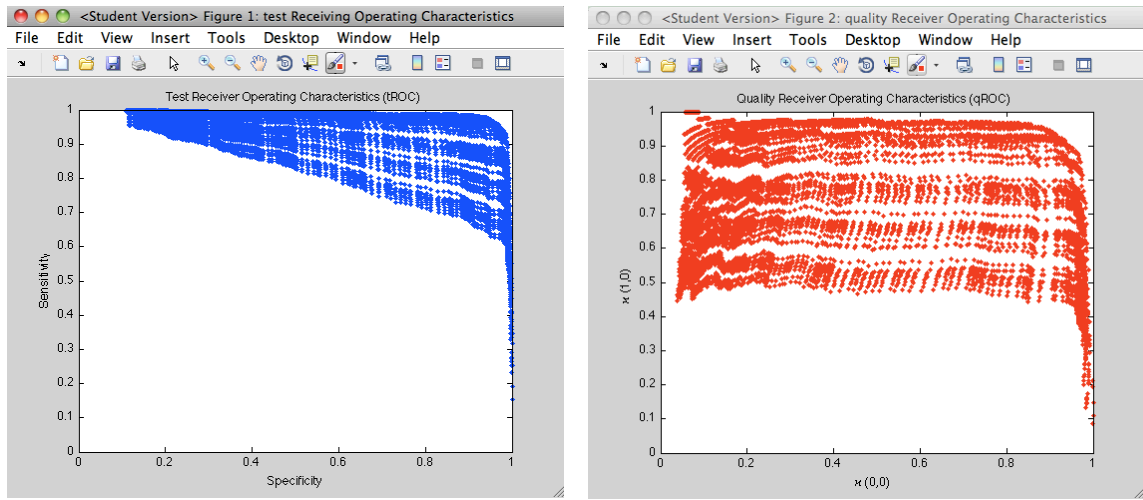
a. Test ROC (left) and quality ROC (right).
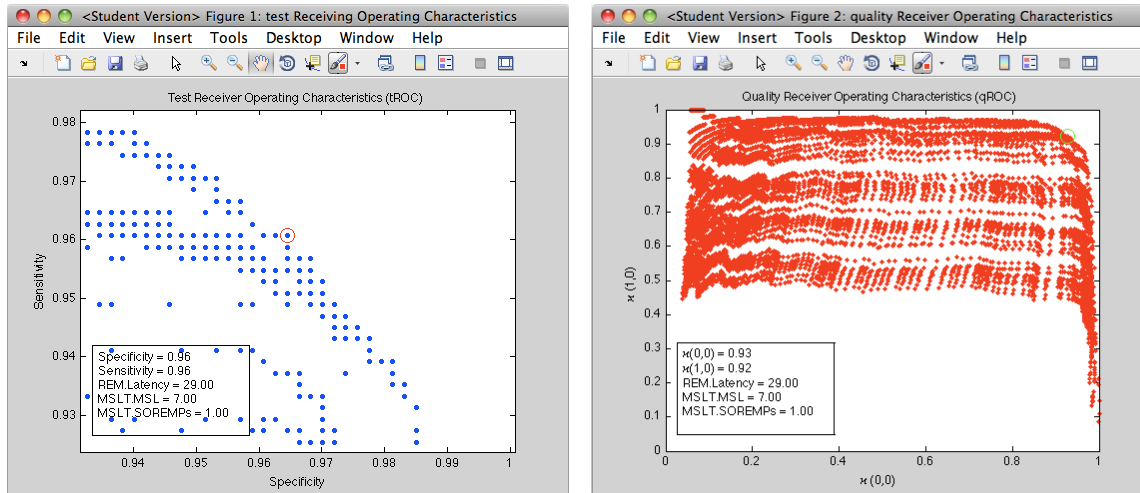
b. Zoomed in view of test ROC with the selected point shown with a circle and the corresponding qROC (right) values shown with the matching subtest configuration cutoffs.

Figure 2.3: Interactive plots. ROC and qROC plots produced using *All ROC results* selection (a), and zoomed in view of the results (b).

The tROC and qROC plots of every diagnostic test examined in this configuration using the sample data set are shown in
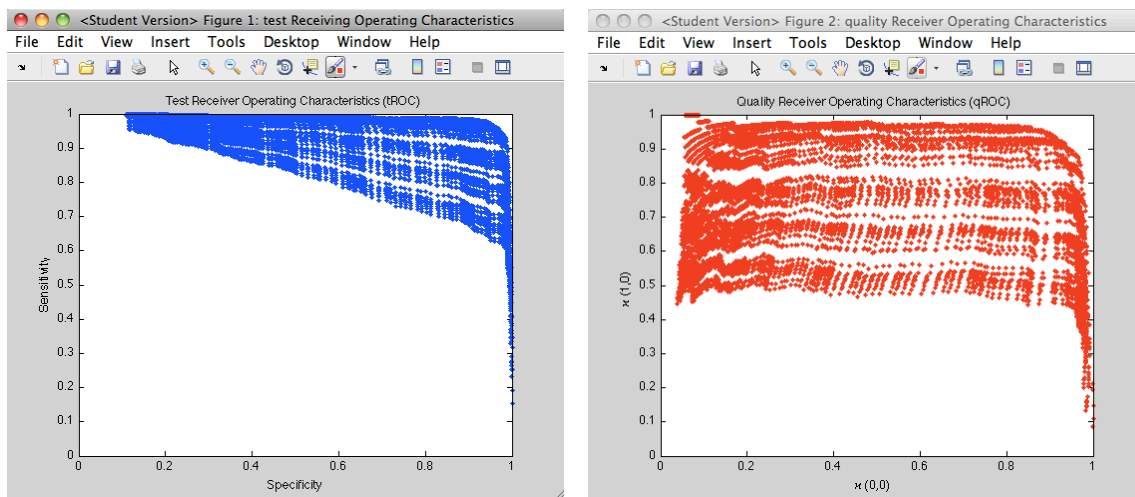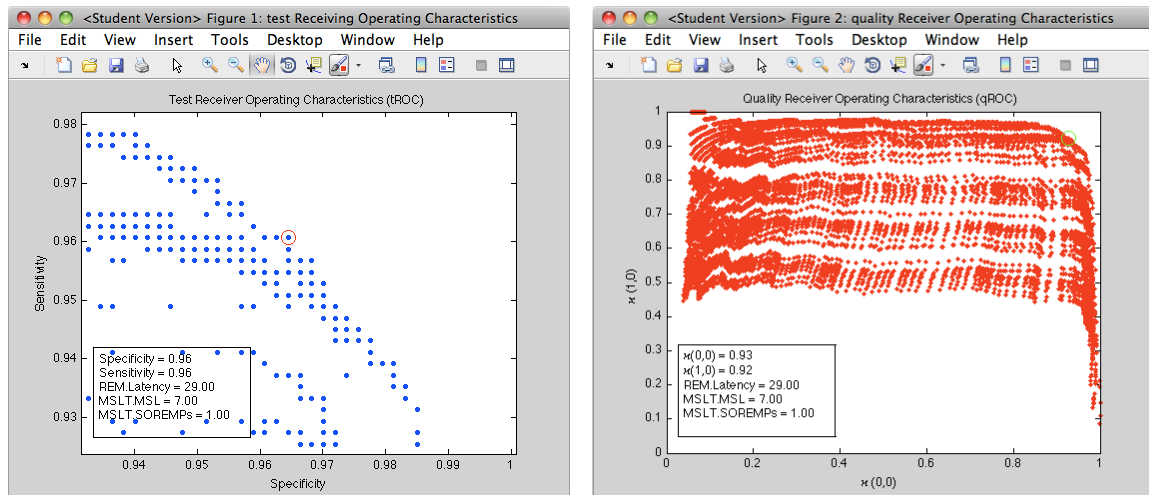
c. Test ROC (left) and quality ROC (right).

70

d. Zoomed in view of test ROC with the selected point shown with a circle and the corresponding qROC (right) values shown with the matching subtest configuration cutoffs.

Figure 2.3(a). tROC results are shown as blue dots and qROC results as red. Figure 2.3(b) shows the results of zooming in and selecting a point in the ROC space. The quality indices for the selected points matching cut points are shown simultaneously (and vice versa) to avoid possible misperception that tROC and qROC values lay in similar locations.

2.3.5 OPTIMAL DIAGNOSTIC TESTS
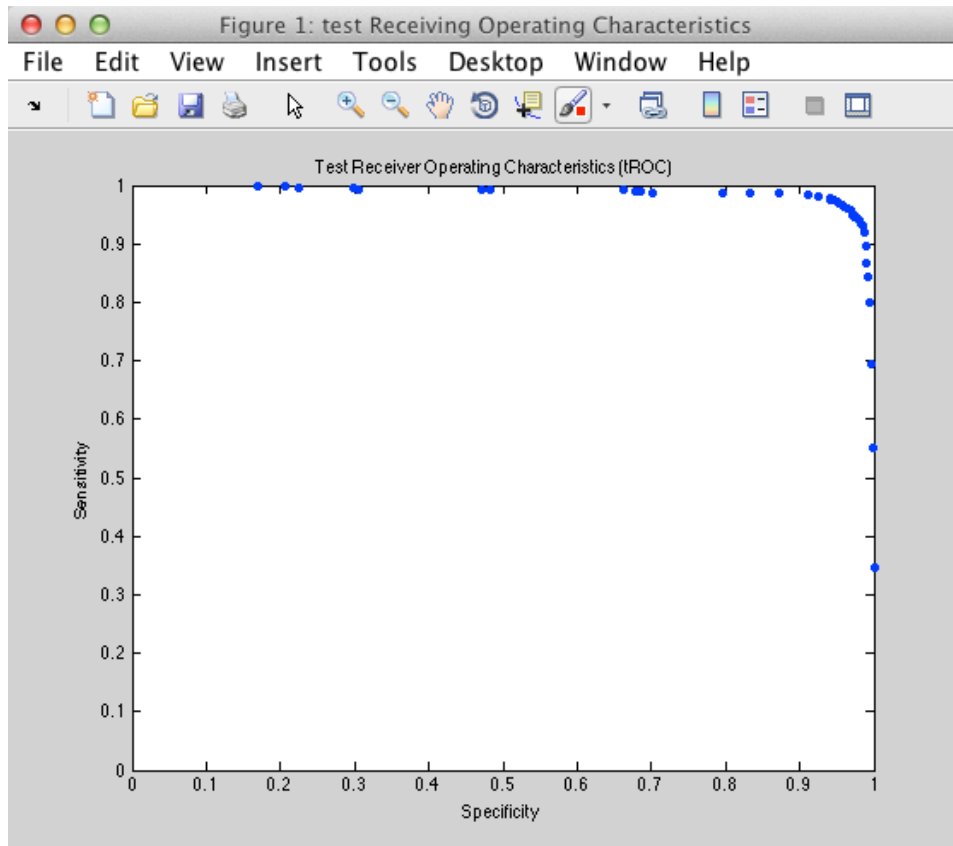
a. tROC convex hull
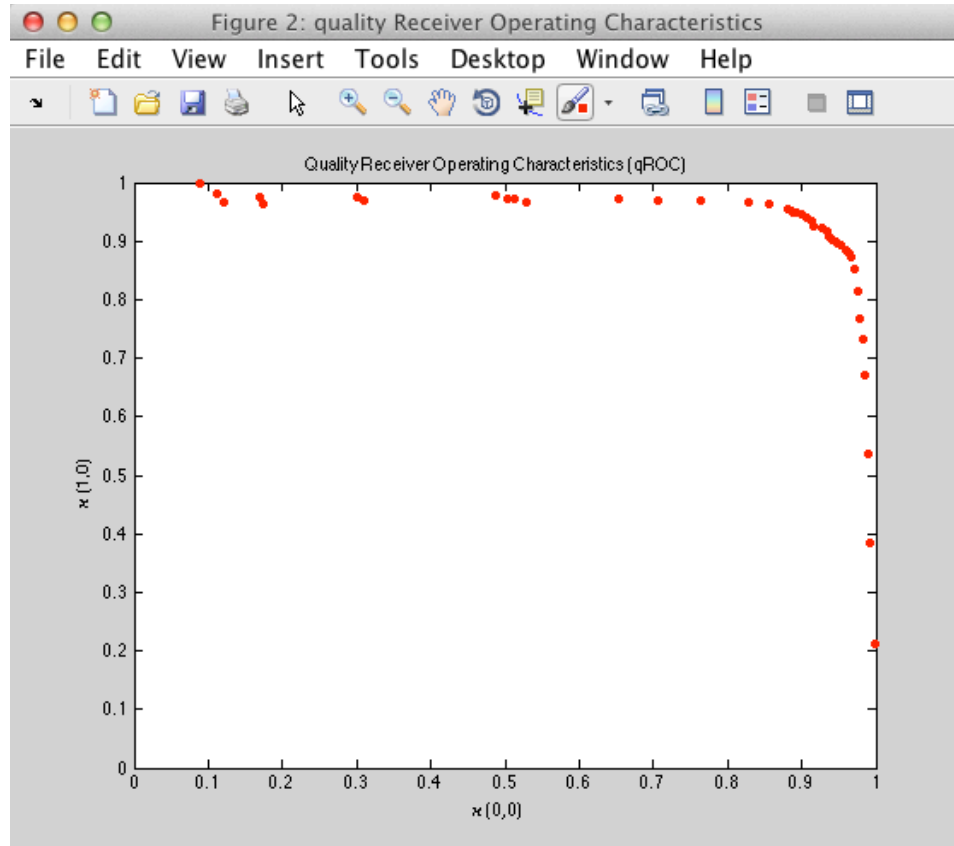
b. qROC convex hull

Figure 2.4: ROC convex hull. Plots of ROC convex hulls are produced using the *Optimal ROC results* selection.

Clinicians are not generally interested in examining *all* diagnostic tests, just the *best*. But what is the best diagnostic test, or, rather, what makes one test better than another? Points along the ROC convex hull are superior to others when evaluating curves from multiple tests. These points can be filtered for display as seen in Figure 2.4. However, because there are two criteria of interest, sensitivity and specificity, even on this outermost curve there is no uniformly best model. The answer depends on the desired trade-off between these criteria, which must be aggregated to select and validate a single model. The software does this by using a convex combination of the two

$$(1 - \alpha) \cdot \text{sensitivity} + \alpha \cdot \text{specificity}$$

73

for a user selected $\alpha \in [0,1]$. One way a user specifies α is explicitly in softROC's settings dialog (Figure 2.5). Here the *Sensitivity-Specificity balance* is at its default, central position with an α of 0.5, but it can easily be adjusted to place proportionally more weight on either sensitivity or specificity.
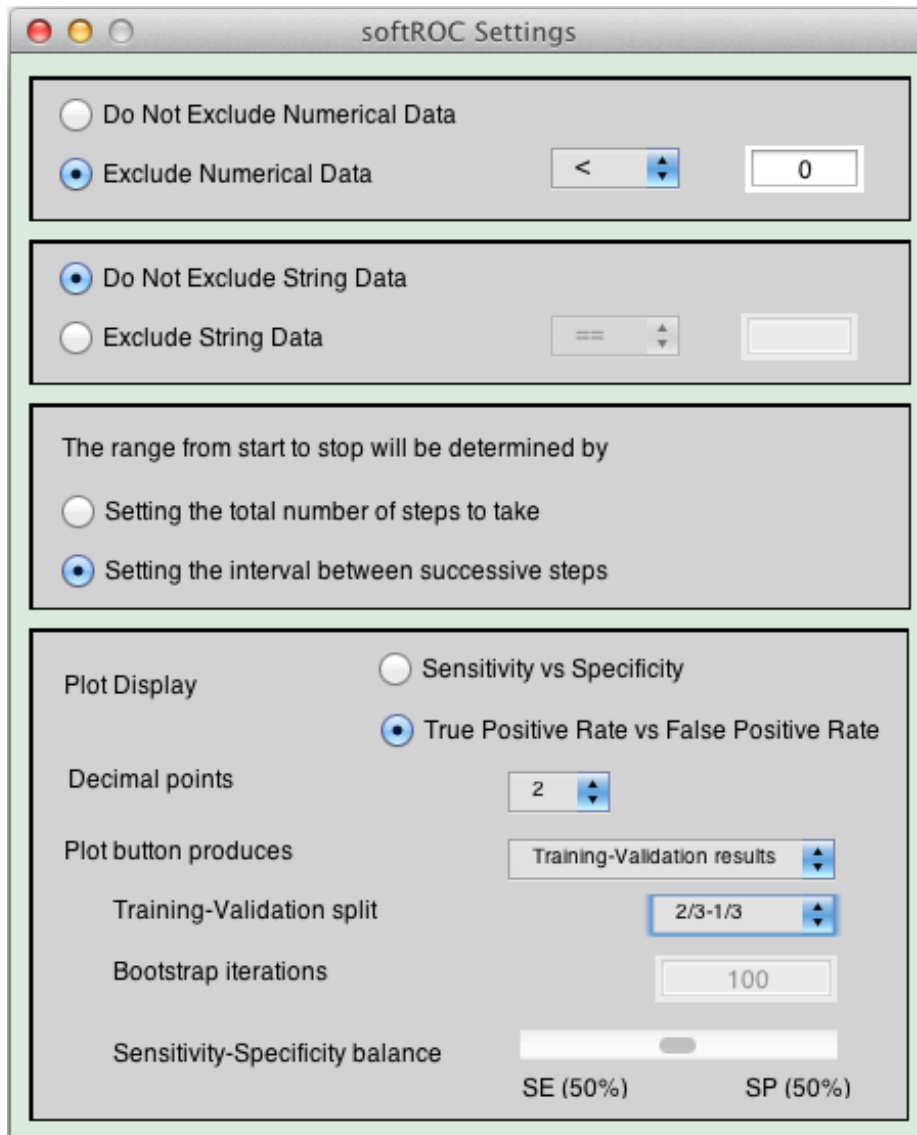
Figure 2.5: softROC's plot settings dialog. The settings GUI allows users to set the desired tradeoff between sensitivity and specificity, the number of bootstrap iterations to perform, and various plot and output formatting options.

The user may decide to define α implicitly by selecting a point from the plot directly and letting the software find the α for which this point is *optimal* (generally there is an interval of such α, and the center value of that interval is chosen). If no α exists, it is interpolated from the nearest points that are optimal for some α. Regardless of the approach, once α is obtained, the user can select a *best* model and assess its variability and performance.

2.3.6 GENERALIZATION

One drawback to a multiple-parameter exhaustive search is the increased potential to over-fit the data. A clinician is generally interested in how well their diagnostic criteria will perform on future patients – using the same data to train and evaluate a model can lead to an overly optimistic estimate of future performance. We used split-set validation to generate unbiased estimates and bootstrapping to obtain confidence intervals.

2.3.6.1 Split set validation

In evaluating and constructing the diagnostic tool, many different combinations of candidate cut-points are considered. Because the *best* set of cut-points is selected, the ROC characteristics estimated on the training data are no longer unbiased estimates of the ROC characteristics for future patients --- they are potentially overly optimistic. A training-set/test-set split is typically used to address this issue and correctly assess ROC characteristics of the selected best point. The data are divided in two. The first set (training set) contains a random 2/3 of the observations (stratified to include 2/3 of the cases and 2/3 of the controls). The second set (test set) contains the remaining 1/3 of the observations. The first sub-dataset is used to *train the model*: the ROC characteristics of all cut-point sets are assessed with these observations, and the best set of cut-points is selected. The second dataset is used to assess the ROC characteristics of this optimal model for future data. All of the test patients are then classified with the optimal model to

obtain the unbiased estimate. The 2/3--1/3 split is based on the recommendation in [51], though a 1/2--1/2 split may be used instead.

Because some degree of data exploration was presumably performed on the complete dataset before selecting the Boolean settings and ranges of the cutoffs, the estimates above are not entirely unbiased. This bias will increase if the full dataset is more fully explored. This issue is discussed in depth in [52]. In our case however, we are not working in the very high-dimensional framework. The issue is further mitigated by variable selection during data exploration being neither automated nor exhaustive on its own but driven by the user.

softROC generates four plots in the split-set validation mode: a tROC-qROC pair generated from the training-set, and a tROC-qROC pair generated from the testing-set. Clicking on any point in the one of the four produced plots will identify that points diagnostic cutoffs and display the matching ROC (test and quality) points in each of the remaining three plots. Estimates may also be taken from selected points as described in Section 2.3.7.

2.3.6.2 Bootstrapping

In addition to sensitivity and specificity, one may be interested in the diagnostic's stability – i.e. if this model were built on a different set of patients from the same population, how different might the "optimal" cut-points look? While one generally cannot obtain more datasets from the populations, this can be approximated with the bootstrap. For each of many bootstrap replicates, patients are resampled with replacement, and an optimal cut-point replicate is found based on the α value used. To assess the variability of the optimal cut-point estimates, one looks at the histogram of the optimal cut-point replicates. The empirical inner 95 percentile of these replicates is used as a confidence interval. The user may set the number of iterations to use when bootstrapping in softROC's settings dialog (Figure 2.5).

2.3.7 IMPLICITLY DEFINING ALPHA

Test Receiver Operating Characteristics (tROC)

Bootstrap this point
Get unbiased ROC for this point

Specificity = 0.30
Sensitivity = 0.99
PSG.REML = 173.00
MSL = 2.00
SOREMPs = 1.00
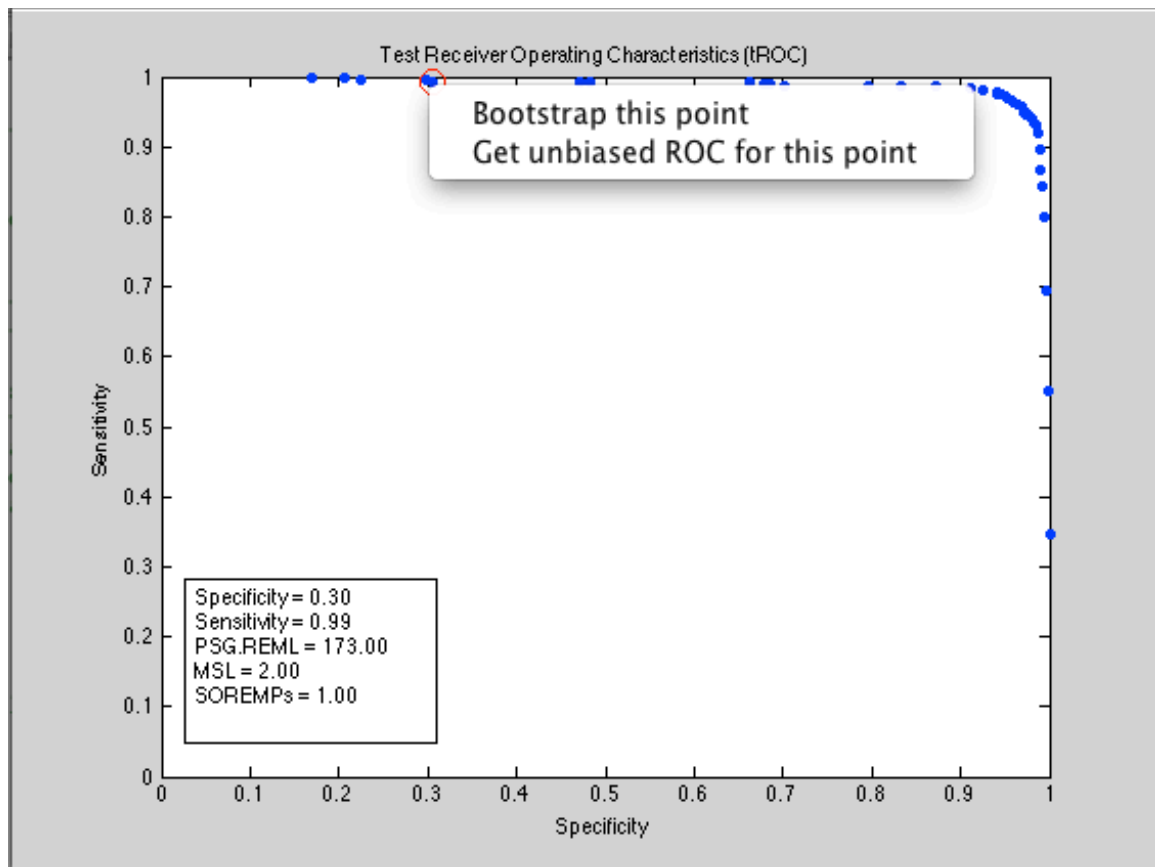
a. Context menu from selected value.

Unbiased ROC estimate using training-validation split based on
Training sample size = 722 samples
Validation sample size = 362
SE = 0.98
SP = 0.19

OK
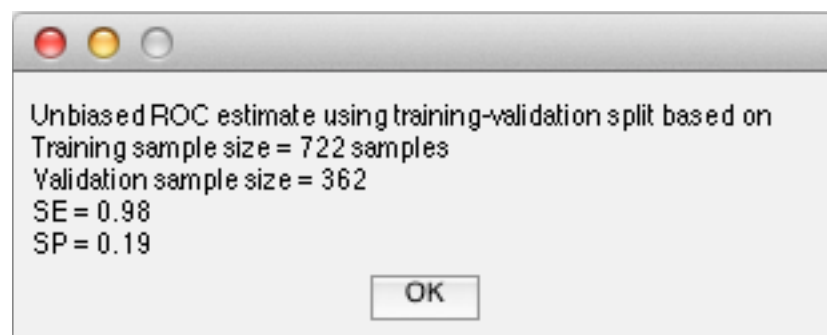
b. 95% confidence interval message box.
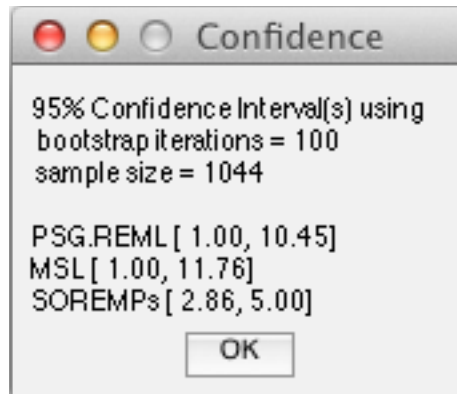
c.   Unbiased estimate message box.

Figure 2.6: Context menu produced when an ROC value is selected and corresponding results.

Users may obtain unbiased ROC estimates and confidence intervals directly from selected points. This approach implicitly defines α by using an iterative slope calculation method to produce a range of optimal values and associated subtest configuration cut-points on the outermost ROC curve. The final α value used is taken from the linear interpolation of the selected point's position within the generated α-range to ROC mapping. The α value obtained in this way is optimal for points along the outermost curve, and cannot locate suboptimal points. Thus a point selected inside the outer curve is replaced with the nearest optimal one in the ensuing analysis. Figure 2.6(b) and Figure 2.6(c) respectively show the results of choosing to obtain an unbiased estimate and bootstrapping a selection (Figure 2.6(a)). The user can similarly obtain confidence intervals for the AUC of the convex hull. The AUC is calculated using trapezoidal integration.

2.3.1 Narcolepsy studies

Two retrospective studies were conducted to evaluate PSG related diagnostic measures for narcolepsy.

2.3.1.1 Cut-offs for narcolepsy without cataplexy

The goal of the first study was to compare clinical, electrophysiologic, and biologic data in narcolepsy without cataplexy with low ($\leq$ 100pg/ml), intermediate (110-200 pg/ml), and normal (> 200 pg/ml) concentrations of CSF hypocretin-1. This study used 171 subjects with narcolepsy, who did not have cataplexy, and 170 subjects without narcolepsy as controls. Patients with narcolepsy without cataplexy, and having CSF hypocretin-1 measures, were found from databases at Stanford's Center for Narcolepsy and similar ones in collaborative labs located in China. Narcolepsy without cataplexy was determined using the International Classification of Sleep Disorders (ICSD)-2, which implies that all cases had excessive daytime sleepiness and a positive MSLT (i.e. MSL $\leq$ 8 min, $\geq$ 2 SOREMPs).

softROC served two purposes in the study of narcolepsy without cataplexy. First, it was used to determine optimal cut-off of CSF hypocretin-1 concentration for narcolepsy without cataplexy as defined by ICSD-2 (n=171), compared to unrelated controls (n=170). Second, it was used to determine best MSLT cutoffs for predicting hypocretin deficiency in narcolepsy without cataplexy, that is, CSF hypocretin-1 $\leq$ 200 pg/ml.

2.3.1.2 Short REML during NPSG for narcolepsy screening

The goal of the second narcolepsy study was to determine if short REML during the NPSG (i.e. prior to MSLT) could be used as a screening test to predict narcolepsy cases associated with hypocretin deficiency. Four comparisons were made using a total of 1,749 different participants.

Comparison 1 examined 516 narcoleptics with either low CSF hypocretin-1 levels or clear cataplexy with HLA-DQB1*06:02 positivity to 516 adults age- and gender-matched adults randomly selected from the general population. Comparison 2 used a clinical sample of patients referred to the sleep clinic who underwent NPSG (n=749). Similarly characterized cases of narcolepsy were selected from this group as cases (n=14) and all others with sleep disorders were used as controls (n=735). Comparison 3 used a sample

of 254 successive patients with hypersomnia and a high pretest probability of narcolepsy. Similarly characterized cases of narcolepsy were selected as cases (n=122), while remaining patients diagnosed with narcolepsy without hypocretin deficiency, idiopathic hypersomnia, or Kleine-Levin syndrome were used as controls (n=118). This third comparison allowed us to evaluate our REML criterion and compare it to those of the MSLT in a clinical setting of patients with hypersomnia. Lastly, comparison 4 examined 118 narcoleptic patients with low CSF hypocretin-1 levels to 118 age- and gender-matched narcoleptic patients with normal concentrations CSF hypocretin-1. In this last sample, MSLT positivity (i.e. MSL $\leq$ 8 min, > 2 SOREMP) was the primary diagnostic criterion for narcolepsy, as 62% of the controls with normal CSF hypcretin-1 levels did not have cataplexy.

2.4   RESULTS

The softROC program was helpful in exploring and determining optimal diagnostic measures in two narcolepsy studies. The first study investigated narcolepsy without cataplexy, while the second investigated narcolepsy with cataplexy.

80

Figure 2.7 ROC curve for CSF hypocretin-1 levels as a predictor of narcolepsy without cataplexy (96 cases versus 96 controls). The circle highlights the best CSF-hypocretin-1 cut-off (201 pg/ml). Each dot identifies the sensitivity and specificity resulting from a different test configuration applied to the cases and controls in conjunction with the gold standard.
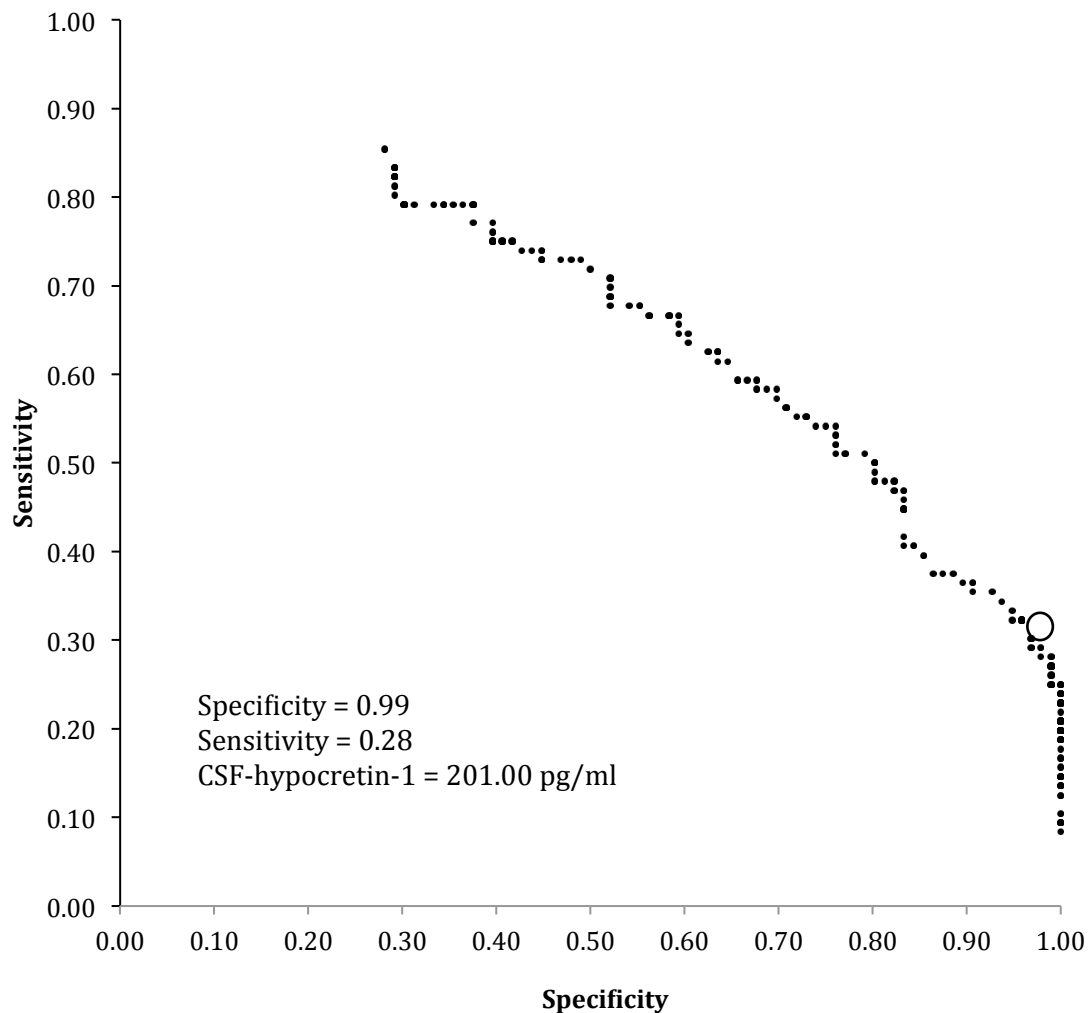
Using softROC, we found that a CSF hypocretin-1 concentration of approximately 200 pg/ml was the best cut-off of narcolepsy without cataplexy (Figure 2.7). This had a high specificity of 99%, but a low sensitivity of 33% and reflects the fact that only a minority

of patients without cataplexy are hypocretin-deficient. This cut-off matches that of previously used groupings of CSF-hypocretin-1 concentrations mentioned earlier (i.e. low, intermediate, and normal). Using this cut-off of 200 pg/ml, we next found the best MSLT cut-offs for predicting hypocretin deficiency in narcolepsy without cataplexy was an MSL ≤ 2 min with ≥3 SOREMPs (Figure 2.8). This had a high specificity of 95%, but again a low sensitivity (39%).

Figure 2.8 Multiple ROC curve for MSLT parameters as a predictor of narcolepsy without cataplexy with CSF hypocretin-1 ≤ 200 pg/ml (96 cases versus 96 controls). The circle highlights the best MSLT MSL (≤ 2 min) and MSLT SOREMPs (≥ 3) cut-off point.

Only 24% of our sample had low or undetectable concentrations, which was in line with previous findings showing only minority of narcolepsy without cataplexy cases having low concentrations of CSF hypocretin-1. Reports of subjective symptoms like hypnogogic hallucinations, triggers, or severity of atypical cataplexy, when present, or number of naps did not differ between the hypocretin-1 subgroups. However, PSG abnormalities were generally more pronounced in patients with low concentrations of CSF hypocretin-1 than those with higher levels (e.g. shorter MSL and a higher number of SOREMPs by MSLT).

2.4.2 SHORT REML DURING NPSG

The goal of the second narcolepsy study was to determine if short REML during the NPSG (i.e. prior to MSLT) could be used as a screening test to predict narcolepsy cases associated with hypocretin deficiency. Four comparisons were made using a total of 1,749 different participants. The case control sample pairings described in Chapter 1.

The first comparison examined 516 narcoleptics with either low CSF hypocretin-1 levels or clear cataplexy with HLA-DQB1*06:02 positivity to 516 adults age- and gender-matched adults randomly selected from the general population. The second comparison used a clinical sample of patients referred to the sleep clinic who underwent NPSG (n=749). Similarly characterized cases of narcolepsy were selected from this group as cases (n=14) and all others with sleep disorders were used as controls (n=735). The third comparison used a sample of 254 successive patients with hypersomnia and a high pretest probability for narcolepsy. Similarly characterized cases of narcolepsy were selected as cases (n=122), while remaining patients diagnosed with narcolepsy without hypocretin deficiency, idiopathic hypersomnia, or Kleine-Levin syndrome were used as controls (n=118). This third comparison allowed us to evaluate our REML criterion and compare it to those of the MSLT in a clinical setting of patients with hypersomnia. The fourth and final comparison in this particular study examined 118 narcoleptic patients with low CSF hypocretin-1 levels to 118 age- and gender-matched narcoleptic patients with normal

concentrations CSF hypocretin-1. In this last sample, MSLT positivity (i.e. MSL ≤ 8 min, > 2 SOREMP) was the primary diagnostic criterion for narcolepsy, as 62% of the controls with normal CSF hypcretin-1 levels did not have cataplexy.

2.5 RESULTS

The softROC program was again helpful in exploring and determining optimal diagnostic measures in two narcolepsy studies using separate populations of narcoleptics and controls. The first study investigated narcolepsy without cataplexy, while the second investigated narcolepsy with cataplexy.

2.5.1 NARCOLEPSY WITHOUT CATAPLEXY

Figure 2.9 ROC curve for CSF hypocretin-1 levels as a predictor of narcolepsy without cataplexy (96 cases versus 96 controls). The circle highlights the best CSF-hypocretin-1 cut-off point (201 pg/ml).
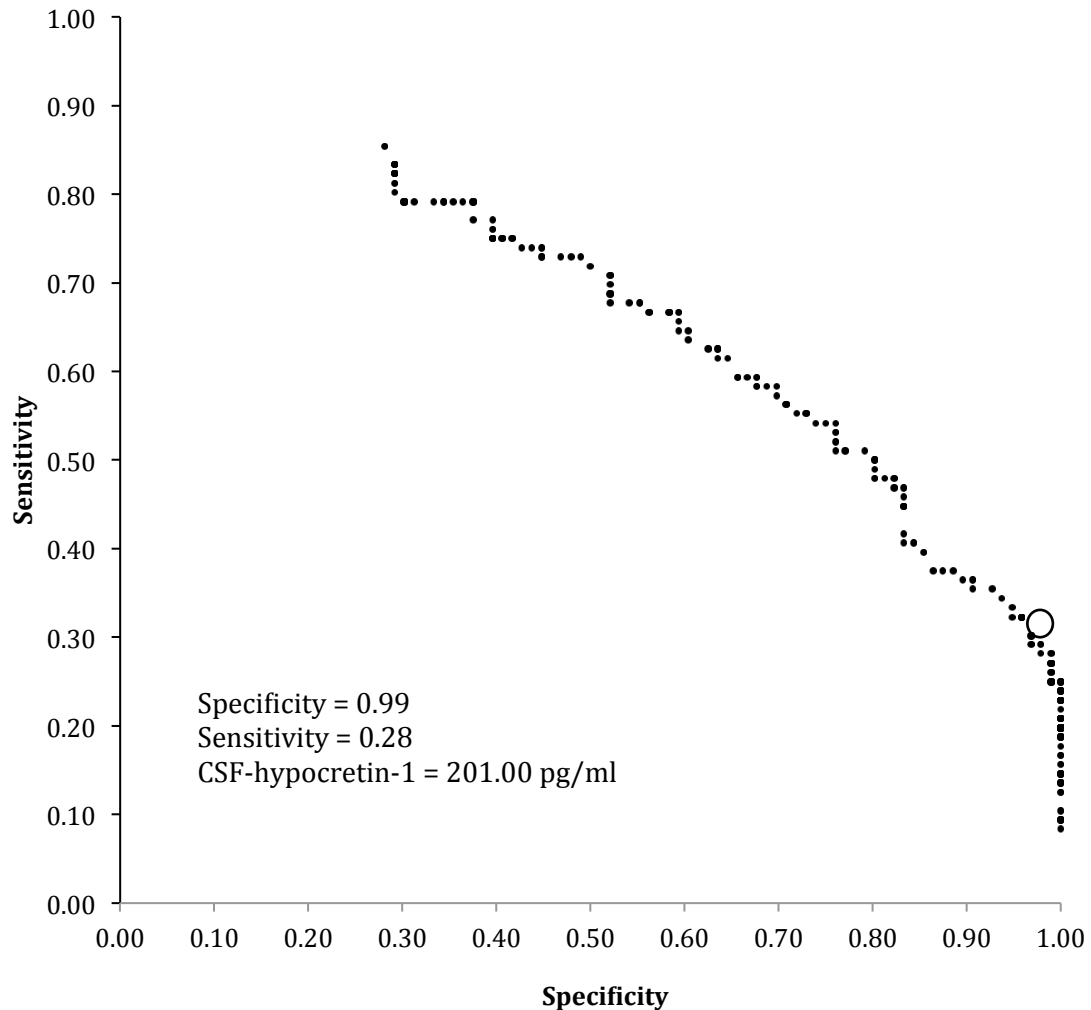
Using softROC, we found that a CSF hypocretin-1 concentration of approximately 200 pg/ml was the best cutoff of narcolepsy without cataplexy (Figure 2.7). This had a high specificity of 99%, but a low sensitivity of 33% and reflects the fact that only a minority of patients without cataplexy are hypocretin deficient. This cut-off matches that of previously used groupings of CSF-hypocretin-1 concentrations mentioned earlier (i.e.

low, intermediate, and normal). Using this cut-off of 200 pg/ml, we next found that the best MSLT cut-offs for predicting hypocretin deficiency in narcolepsy without cataplexy was an MSL ≤ 2 min with ≥3 SOREMPs (Figure 2.8). This criterion had a high specificity of 95%, but again a low sensitivity (39%).

Figure 2.10 Multiple ROC curve for MSLT parameters as a predictor of narcolepsy without cataplexy with CSF hypocretin-1 ≤ 200 pg/ml (96 cases versus 96 controls). The circle highlights the best MSLT MSL (≤ 2 min) and MSLT SOREMPs (≥ 3) cut-off point.

Only 24% of our sample had low or undetectable concentrations, which was in line with previous findings showing only minority of narcolepsy without cataplexy cases having

low concentrations of CSF hypocretin-1. Reports of subjective symptoms like hypnogogic hallucinations, triggers, or severity of atypical cataplexy, when present, or number of naps did not differ between the hypocretin-1 subgroups. However, PSG abnormalities were generally more pronounced in patients with low concentrations of CSF hypocretin-1 than those with higher levels (e.g. shorter MSL and a higher number of SOREMPs by MSLT).

2.5.2 SHORT REML DURING NPSG

ROC curves of comparison 1 are shown in Figure 2.11 which consists of 516 patients with narcolepsy with either low CSF hypocretin-1 levels or cataplexy and HLA-DQB1*06:02 positivity vs 516 age- and gender-matched controls. An REML of 18 minutes or less gave best specificity cut-off. In this case, specificity was 99.2% (95% CI: 98.5-100.0), sensitivity was 51.6% (95% CI: 47.2-55.9]), and the AUC was 0.799 (95% CI: 0.771-0.826). Further analysis was performed on subsets of this sample. Patients with low CSF hypocretin-1 levels (n=89) had an optimal REML cut-off of 17 minutes or less with 97.8% specificity (95% CI: 92.1-99.7), 44.9% sensitivity (95% CI: 34.4-55.9), and an AUC of 0.704 (95% CI: 0.625-0.786). For patients with cataplexy and HLA-DQB1*06:02 positivity (n=427), the optimal cut-off was an REML of 21 minutes or less with 99.5% specificity (95% CI: 98.3-99.9), 53.9% sensitivity (95% CI: 49.0-58.7), and an AUC of 0.820 (95% CI, 0.789-0.850). The results were all very similar.

Figure 2.11 ROC curves for REML comparison 1.
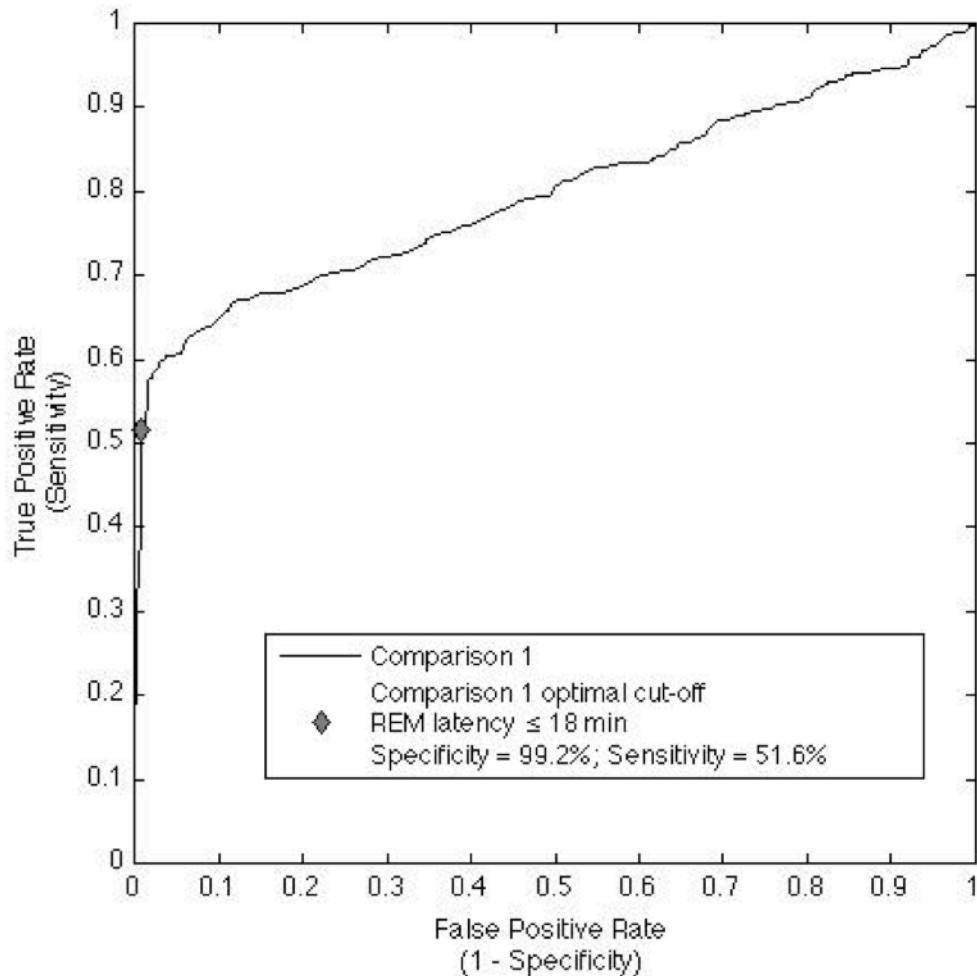
ROC curves for comparisons 2, 3, and 4 are shown in Figure 2.12. In comparison 2, an REML of 16 minutes or less is the optimal cut-off (specificity, 99.6% [95% CI, 99.1-100.0]; sensitivity, 42.9% [95% CI, 16.9-68.8]) for the diagnosis of narcolepsy/hypocretin deficiency compared to other sleep disorders (AUC, 0.704 [95% CI, 0.524-0.907]). In comparison 3, an REML of 17 minutes or less (specificity, 95.5% [95% CI, 90.4-98.3]; sensitivity, 58.2% [95% CI, 48.9-67.1]) was the best cut-off for the diagnosis of narcolepsy with low CSF hypocretin-1 levels versus narcolepsy with normal hypocretin-1 levels, idiopathic hypersomnia, and Kleine-Levin syndrome (AUC, 0.765

88

[95% CI, 0.707-0.831]). In comparison 4, an REML of 8 minutes or less was the optimal (specificity, 87.3% [95% CI, 81.3-93.3]; sensitivity, 44.1% [95% CI, 35.1-53.0]) cutoff for diagnosing narcolepsy with low CSF hypocretin-1 levels against narcolepsy with normal CSF hypocretin-1 levels (AUC, 0.628 [95% CI, 0.558-0.702]).

Figure 2.12 ROC curves for REML comparisons 2, 3, and 4.
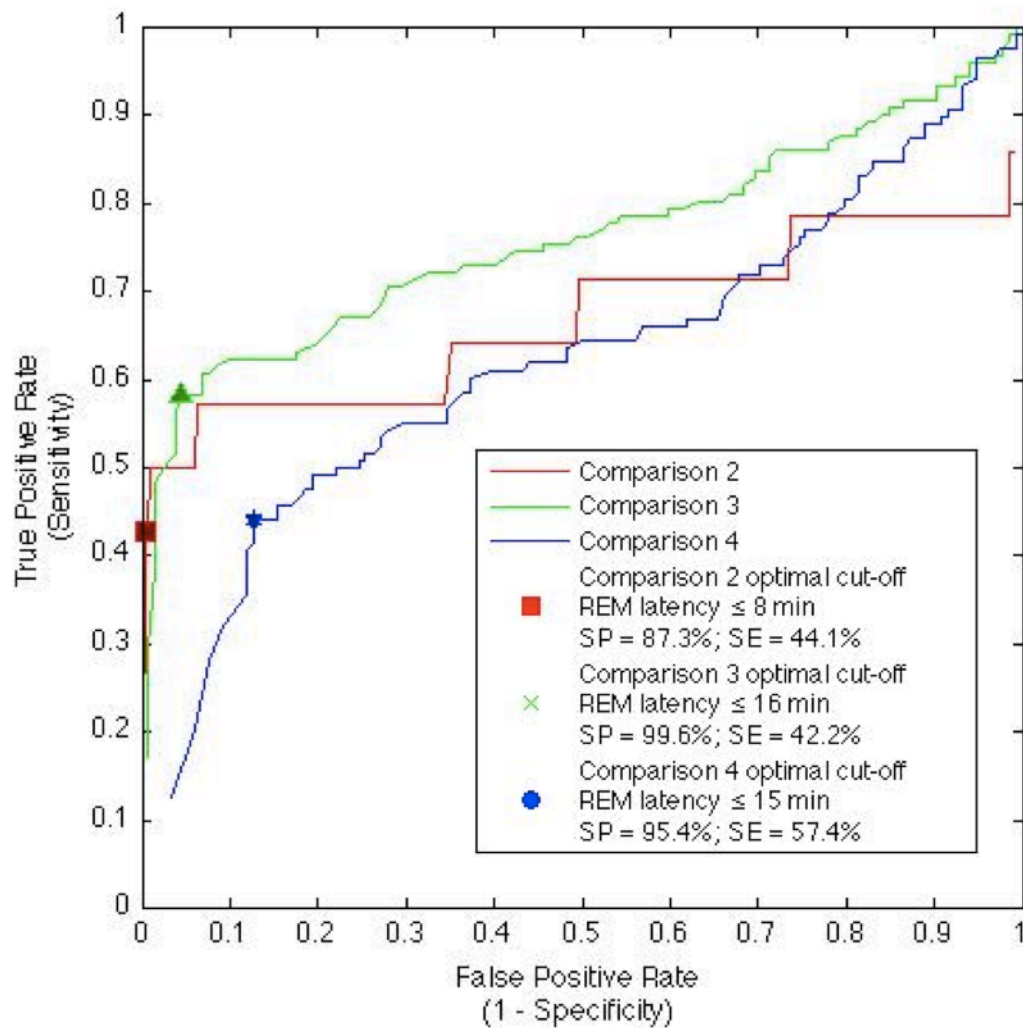
2.6 DISCUSSION

Machine learning continues to play an important part in medical practice and decision making; however, it is clear that human expertise is not readily taken out of the equation when it comes to diagnosing ailments. Cooper *et al* [53] evaluated the performance of several machine-learning methods to predict mortality of patients hospitalized with pneumonia with the goal of reducing treatment costs (both personal and financial). In this thorough study, a total of 63 variables (continuous and categorical) from each patient were used to training (n=9,487) and evaluation (n=4,352) [53]. The authors point out the need for feature reduction as the otherwise large feature sets present difficulty to implement in practice compared to a smaller set of features, which may be converted into a paper-based model that a clinician, who ultimately acts on the diagnosis, can readily apply. Solutions that engage medical expertise and serve as part of the process may gain the trust necessary to put into practice as we found in our case using interactive combinatorial ROC scatter plots.

2.6.1 APPLICATION AND EXTENSION

softROC is a general software tool that can be used for diagnostic discovery for any disease or pathology given an input file containing symptom data and "ground truth" diagnoses. Our group has successfully used the software to identify and support new diagnostic measures for narcolepsy.

For example, a CSF cut-off value of hypocretin-1 of 200 pg/ml was found to be optimal for identifying narcolepsy without cataplexy, and MSLT measures were found (i.e. MSL $\leq$ 8 min, $\geq$ 2 SOREMPs) for predicting hypocretin deficiency above and below this cut-off in narcolepsy without cataplexy. These and other findings from this study may be found in the manuscript we published in the journal *SLEEP* [44]. One drawback to this study is that lumbar punctures (used to obtain CSF) are only ordered for severely affected patients who have typically shown presence of the disease for an average of ten years. The inclusion requirement for patients to have CSF hypocretin-1 measures likely biases

patients with normal concentrations in our sample to show greater increase in severity than would be found in a randomly selected pool of narcoleptics without cataplexy [44].

Furthermore, we used softROC to find optimal NPSG REML cutoffs for diagnosing narcolepsy and evaluating its performance to the MSLT in several comparisons of different samples for narcolepsy and controls [50]. A tutorial Excel file based on one of the datasets from this study is included for reference with softROC's instruction manual. The strength of our program versus those available in statistical packages is its user friendliness and the fact it can easily combine multiple parameters to optimize diagnostic criteria. softROC is filling a niche for medical diagnosis users and we hope to extend its use by answering the call for open source machine learning software [54].

2.6.2 DIAGNOSTIC DISCOVERY

Performance evaluation of candidate diagnostic configurations can be done graphically within softROC, or offline by outputting the results to a tab-delimited text file. Graphical output produces two separate, interactive plots. One plot displays test ROC results showing sensitivity and specificity of each configuration in the search space, while the other plot displays the quality ROC values. Clicking on any ROC value in either plot causes the matching configuration result to be circled in both plots and produces a text box showing its ROC and configuration values. This allows easy selection of ROC values and corresponding diagnostic configuration cut-offs that are closest to the ideal test point within the range of specificity or sensitivity desired as shown in Figure 2.1. Text output includes additional statistical information for each diagnostic configuration test against the selected gold standard, including tROC sensitivity and specificity, qROC Kappas, Cohen's Kappa, contingency table values, positive and negative predictive values, and corresponding diagnostic cut-off parameter values used to produce the results in comparison to the gold standard. These values are defined in Table 2.2.

2.6.3 PIGEONHOLING

softROC's output is shown as a scatterplots and not ROC curves. ROC curves are not displayed because the points plotted are truly multidimensional. For example, the vertical clustering or layering of ROC points shown in Figure 2.3 reflects choices of SOREMPs and MSL, while horizontal changes are driven at each layer primarily by changes in PSG.REML cut points. The ROC convex hull thus consists primarily of using set cut-points for SOREMPs and MSL and varying the cut-point of PSG.REML. Applying thresholds in this manner produces a step-function that can theoretically be made smoother and more curve-like with a smaller step-size. In practice, this is not the case. Aside from computational ceilings, the data simply has a limit to which it can be split. The subtest itself may only offer a limited, discrete number of results (such as the case of MSLT SOREMPs in our example whose only outcomes are in the set {0,1,2,3,4,5}). Establishing finer cut points beyond a subtest's range futilely increases computation without ROC resolution. This *pigeonholing* effect occurs in softROC whenever different cut-points produce the same ROC result and prevents generating individual curves by simply "connecting-the-dots". When duplicate entries occur inside the ROC limits of 0 and 1, a many-to-one-to-many relationship exists which cannot be visually resolved in ROC space. These case may be addressed offline by outputting the results to a text file from softROC and analyzing them with another program such as R, MATLAB, Excel or any text editor.

2.6.4 LIMITATIONS

Generating ROC curves by simply varying a threshold across a range of values may be viewed as a computationally poor approach [55], however we found it to be a necessary burden when evaluating the performance of combined subtests as pointed out by Kraemer [46]. Understanding the granularity of one's data makes computational processing a non-issue in most cases. A basic laptop computer (Intel Core 2 Duo 2.33GHz) evaluated ~4,000 candidate configurations per second. Furthermore, we found computation time to be a non-issue in most cases for our research using relatively current issues. The

computation time scales linearly by the number of iterations performed when bootstrapping. In these cases, it may be necessary either to let the simulation run overnight or reduce the number of iterations desired.

Other approaches to evaluating the performance of combined tests exist and include bagging, voting, and regression schemes. Furthermore cross-validation is another useful approach that can be taken over the split-set validation method we used. These methods have their place. The open source community can add them to softROC if they desire. Ultimately, we sided with Kraemer's argument for quality indices and chose an approach fitting to our research – one we hope others in the medical profession can easily adopt.

2.7 CONCLUSION

softROC is a novel, graphical program designed to assist clinical researchers in their search for optimal diagnostic configurations and cut-off values according to test and quality ROC. Interactive plots make it easy to quickly select and compare diagnostic configurations and verify the quality of each selection. Bootstrapping and training-test split validation methods are available to measure 95% confidence intervals of test cut-offs and obtain unbiased ROC values.

softROC's core features include input handling, dynamic search configuration, diagnostic exploration and future performance evaluation. A diagnostic candidate search configuration is established by first selecting the gold standard field and positivity value used to separate cases from controls. Next, test or symptom entries are added to the search configuration and selected from the drop-down menu. The user creates a range of potential cutoff values for each test by selecting a start and stop range to compare across. The range can be broken up by setting the step-size (delta) or by setting the total number of divisions to break the range into. An inequality or equality is selected for comparison against the generated test values. Similar entries can be added or removed and are compared with one another using Boolean logic (e.g. AND, OR). A series of entries can

be grouped using checkboxes, which explicitly sets the order in which Boolean comparisons are made (i.e. how candidate symptoms are combined). ROC curves generated across the configuration space may be explored interactively. Confidence intervals of selected diagnostic configurations are determined by bootstrap while bias is dealt with using split-set validation. softROC provides a flexible, robust setting for exploring diagnostic tests and measuring their usefulness. All these features play a critical role in improving the search for ideal diagnostic criteria using ROCs scatterplots, derived from user-configured medical subtest combinations and variable thresholds of measured results and known outcomes, which would not have been possible using other, existing software packages.

2.8   SUPPLEMENTARY MATERIALS

Supplementary materials include the following items that can be found online at http://www.stanford.edu/~hyatt4/software/softroc/.

- softROC toolbox for MATLAB (.zip): This package contains the MATLAB files necessary to run the softROC software described in this article. A tab-delimited text file containing de-identified narcolepsy patient diagnosis and biomarker data for user testing are also included.

- softROC instruction manual (.pdf): This document provides instructions for installing and using softROC in conjunction with the included data file.

CHAPTER 3   THE STANFORD EEG VIEWER – A TOOLBOX FOR LARGE SCALE ANALYSIS AND VISUALIZATION OF POLYSOMNOGRAPHY DATA

In this chapter, I present the Stanford EEG Viewer. This is a software package I developed to assist in the initial investigation and automation of the different collection of sleep studies, or cohorts, mentioned in Chapter 1. The Stanford EEG Viewer has two distinct modes of operation, single study and batch, which provide the framework for the investigation of periodic leg movements (and their genetic basis), eye movements, and EEG coherency in the following chapters.

There are several people who helped make this software possible who I want to thank. First, Justin Talbot for his instructive insights and data visualization feedback, and Pat Hanrahan, Bernard Widrow, and Emmanuel Mignot for the opportunity to start this work and their guidance. Oscar Carrillo and Simon Warby (in particular) provided motivational problem statements and hours of critical feedback leading to multiple design decisions in context of sleep research needs and practice. Eileen Leery provided feedback for the user interface as a registered polysomnography technician, as did Robin Stubbs as an information technologist.

3.1 INTRODUCTION

Polysomnography (PSG) based sleep studies are data-rich yet are typically visually interpreted by technicians and medical doctors for very basic parameters such as sleep staging, scoring of respiratory events and periodic leg movements. Our laboratory is interested in extracting and validating new PSG biomarkers for disease or genetic studies. Analysis of this scale requires thousands of human subjects (patients and controls) to reach statistical power necessary for genetic analysis or epidemiological relevance. It also requires flexibility in generating new phenotypes that requires specialized algorithms. Open source software packages such as FieldTrip[56], EEGLAB[57], and PSGLab[58] are available but slow, do not accommodate large datasets, and/or cater to very specific needs (e.g. PSGLab was designed for neonatal analysis).   Commercial software packages from companies including Embla (e.g. SandMan), PhiTools (e.g. PRANA), Grass Technologies, and Huneo are more user friendly and broadly-purposed, but the algorithms used are not reviewable and cannot be customized due to intellectual property concerns. Output format in these is specific for the clinical end user and there is no flexibility for use by researchers.

To fill this need, we developed the Stanford EEG Viewer (SEV), a MATLAB [59] toolbox to graphically analyze individual sleep studies and automate analysis of collections of sleep studies through a specially designed graphical user interface (GUI). As use cases, we have been using sleep studies from three sources, the Wisconsin Sleep Cohort (WSC) [43], the Stanford Sleep Clinic (SSC) and the Veteran Affairs (VA) Medical Center, Menlo Park Division. The WSC cohort, designed for longitudinal analysis of sleep in a clinical population, comprises approximately 2,400 sleep studies in 1,090 subjects. The SSC sample includes 762 sleep studies in adult patients with various sleep disorders [44]**.** The Menlo Park VA sample consists of 422 combat veteran sleep studies in controls and Post Traumatic Stress Disorder (PTSD) patients[60].   Sleep studies of these cohorts are exported as European Data Format (EDF) [61] and analyzed using the SEV for various projects including development of a periodic leg movement

detector [62], studies of REM sleep in PTSD and narcolepsy, and studies of spindle activity in correlation to genome wide association studies.

In all these use cases, biomarker classifiers, filters, and power spectral methods are often explored and adjusted on individual studies before application to the entire data set is possible. This cycle of improvement is typically repeated to evaluate performance of classifiers using receiver-operating characteristics (ROC) against a manually scored data set or to investigate interesting and/or atypical results. Despite progress in the development of sleep biomarkers, our judgment is that current methods too often rely on small, accrual-biased datasets unrepresentative of the general population. The SEV simplifies the process of adjusting these and other algorithms to different datasets by exposing pertinent parameter fields to the user through a GUI interface. Researchers can then optimize preexisting classifiers to their data without having to program.

The potential for automated batch processing of artifact detection, power spectral analysis, and biomarker classification sets the SEV apart from other PSG analysis tools. Automated processing tasks of large cohorts can be accomplished within a few hours using the SEV on a single desktop computer. Short processing times allow for more rigorous artifact detection and investigation of optimal biomarker classifications. Biomarker and artifact classifiers incorporated in the SEV include previously published and validated methods as well as new, unpublished methods. Raw and statistically analyzed output from the SEV is exportable to standard, tab-delimited text files or a MySQL database. Additional tools exist within the SEV for further analysis of results; however, the primary purpose of the SEV is to create a useful, flexible framework to prepare large volumes of sleep data for downstream statistical analysis. This may be done using other analytical tools and environments more familiar to the researcher, such as Microsoft Excel or R. The software is freely available for academic use and may be downloaded at http://www.stanford.edu/~hyatt4/.

This chapter is organized in the following manner. Section 3.2 discusses the SEV's single study mode, which allows viewing and exploration of individual PSG-based sleep studies. Section 3.3 covers the SEV's batch mode, which automates signal processing and pattern recognition algorithms across cohorts of PSG-based sleep studies. Concluding remarks are given in Section 3.4.

3.2 SINGLE STUDY ANALYSIS

The SEV's single study analysis mode begins when an EDF file is loaded from the startup screen. In addition, any independent event file (hypnogram/sleep staging, scored respiratory and leg movement events constituting the scoring mask), can be loaded, proving they are in a pre-specified format file (tab-delimited text files or a MySQL database) (see supplementary tutorial video). The single study mode provides the visual thrust of the SEV. Patient data is presented at several scales for frequency and time domain analysis, and classification, filtering, and power spectral analysis performance is easily explored using various settings. Analysis made in the single study mode can then be automated across a directory of studies in the batch mode.

3.2.1 VIEWING
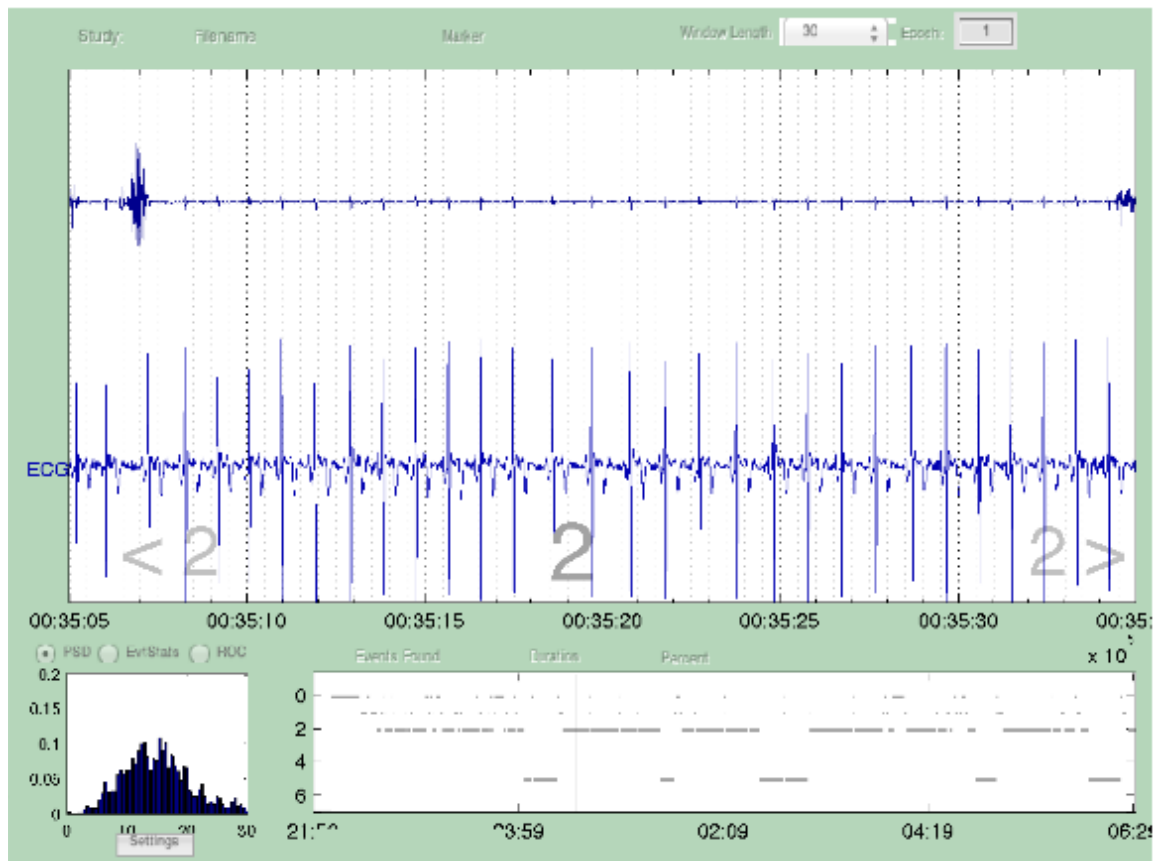
Figure 3.1: The SEV's main application window for single study analysis. The large center window shows the selected polysomnography channels which may be viewed at different scales and manipulated through context menus. The hypnogram window in the bottom-right shows the patients sleep stage and classified biomarker or artifact activity progression for the entire night. The multipurpose window, bottom-left, shows the power spectrum of a selected channel for the current time scale, a classifiers distribution by sleep stage, or a classifiers performance to a ground truth using ROC metrics.

The SEV's home screen is shown in Figure 3.1. Signal channels, corresponding to the polysomnography recordings of a sleep study are loaded from the tool bar or menu bar. Widgets, toolbars, menus, and context menus become available once a study is loaded. Three interactive subwindows provide the single study analysis component of the SEV and are intertwined with one another so that updates in one window are complimentarily reflected in another.

3.2.1.1 Epoch window

The channel data is shown in the primary axes and can be manipulated in several ways using context menus. Right clicking with the mouse on the channel or window of interest will access the associated context menus. Channels can be resized, recolored, relabeled, hidden, duplicated and moved about the display to visualize similarities and differences with other channels. The user can view data at a variety of time resolutions (default is 30-seconds epochs) or as entire sleep stage collections (e.g. Stage 2 or all Rapid Eye Movement (REM) sleep). Manually scored events (imported from event files) can be added using the mouse, and periods of activity can be selected for plotting or power spectral analysis in a pop-out window.

3.2.1.2 Entire night window

Manually staged scoring data can be loaded automatically when available in the correct pre-specified format and shown in the hypnogram window at the bottom-center of the viewer. The hypnogram shows the progression of sleep stages across the entire night in a single view as a compressed collection of 30-s epochs. Users can identify epochs of interest from this view and see the signal data at that time by clicking the corresponding point in time in this window. Epoch adjustment is easily accomplished using the arrow keys or the epoch position edit box or slider widget. Sleep events that are loaded or produced from the detection toolbox are displayed in the axes above the hypnogram. Thicker, longer vertical lines above the staging lines naturally draw the eye to epochs with heightened activity while thinner, shorter vertical lines highlight areas of sparse activity. Epochs of interest are quickly identified so that a classifier's performance in relation to sleep stages can easily be visualized.

3.2.1.3 Multipurpose window

A third window, on the lower left, provides three additional views which the user selects using a radio button. Each view can be further configured using its context menu or the settings button.

- Power Spectral Analysis

Power spectral analysis of an epoch, stage, or selected time period is possible using both periodogram averaging, with a variety of Fast-Fourier-Transform (FFT) windows of selectable size and overlap periods [63], and Multiple Signal Classification (MUSIC) [64] methods. Figure 3.2 shows the configuration screen used to adjust power spectral analysis parameters in the SEV's single-study and batch modes. Power spectral analysis results can be configured to exactly match those of commercial programs like PRANA. Of course finer configuration control and the extension of batch processing further separates the SEV its counterparts.
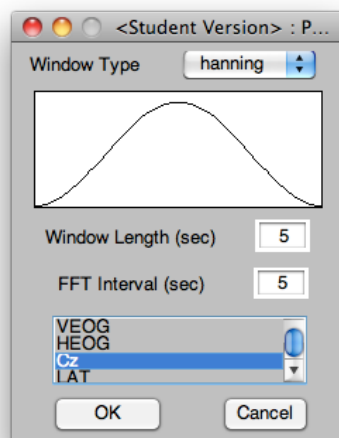
Figure 3.2: The Power Spectral Analysis GUI lets a user change the window type, size, and interval to use for periodogram averaging. The GUI can be used to configure analysis for a single study, or to establish analysis parameters for many studies processed during batch mode.

- Event distribution

A simple bar graph shows the distribution of events detected automatically with one of the classifiers or loaded from a file. In terms of classifier development, this histogram highlights its ability to distinguish between stages. It also provides a quick understanding

of the relationship between the specified events and sleep physiology determined by stage.

- Receiver Operating Characteristics (ROC)

A classifier's performance against another classifier or ground truth dataset is viewed using either test ROC (sensitivity and specificity) or quality ROC (weighted kappa values) [46]. The user selects two classification methods to determine the level of sensitivity and specificity of one classification method versus the other. A third classifier, when selected, serves as an artifact detector. Events detected during periods of artifact are removed from comparison and provide more accurate analysis where artifact rejection is employed. Classifier adjustments made at run time in the epoch window are immediately reflected in the ROC view giving researchers instant feedback on the performance effect of different parameter values for that study.

Direct evaluation of ROC performance on a sample-by-sample based confusion matrix often reveals inflated specificity or incongruous results when evaluating comparatively rare events (e.g. less often than 1/100 of a second), as is the case with many biomarkers of interest. An interaction matrix is first employed to transform classifications from digital time domain to an event space more naturally suited for ROC evaluation. The final evaluation is comparable to precision recall analysis.

3.2.2 EVENT CLASSIFIERS

Several published and unpublished time series biomarker event and artifact classifiers available with the SEV and listed in Table 3.1. Classifiable biomarkers include slow wave activity, spindles, ocular, cardiac activity (e.g. heart rate), and periodic leg movements. Artifact detectors, which classify non-biologic activity such as high frequency activity, flat line, and electrode-pop are also included and help identify unwanted portions of signal. The filter toolbox provides data cleaning solutions as an alternate to data rejection.

Table 3.1: SEV features

FEATURE	DESCRIPTION
Classifiers	<u>Artifact:</u> Flatline, electrode pop, muscle activity[65], high frequency interference <u>Biomarker:</u> Cardiac activity, periodic leg movement[66-68], rapid eye movement[69-72], slow eye movement[73], spindles[74], slow wave activity[75] <u>General:</u> Amplitude threshold, dual amplitude threshold
Power Spectral Analysis	<u>Welch averaging – (FFT based)</u> <u>Multiple signal classification (MUSIC) spectrum</u>
Signal processing tools	<u>Data cleaning and signal estimation</u>: Empirical mode decomposition (EMD), wavelet denoising, adaptive noise cancelling [76] <u>Finite Impulse Response Filters</u>: High pass, low pass, band pass, band stop <u>Moving window methods</u>: Root mean square, mean, median, minimum, maximum, standard deviation, exponentiation. <u>Arithmetic</u>: Absolute value, sum, difference
Single study view windows	<u>Polysomnogram</u>: Multichannel; adjustable time scale; event selection/annotation <u>Entire Study</u>: Sleep staging (hypnogram); biomarker and artifact activity <u>Multipurpose</u>: Power Spectrum; Classification distribution by sleep stage; Receiver Operating Characteristics
Batch processing Mode	Automated processing of multiple EDFs Configurable parameter search space for biomarker classifiers Power spectral analysis with artifact detection
Input formats	European Data Format (.EDF) Hypnogram sleep stage files (.sta) Specialized event files (.sco, apnea or PLM scoring event files) SEV event files (see below output formats)
Output formats	Tab delimited text files MATLAB .MAT files MySQL database JPEG or PNG formatted images

Classifiers are configured to use single or multiple channels. Multichannel algorithms often provide more robust classifications by incorporating physiological significance to further validate or reject initial classifications. For example, inclusion of oxygen saturation measures with detection of channels flat-lining is helpful in artifact detection to determine periods where a patient has likely been wakened and disconnected from the recording system during the middle of a sleep study in order to use a rest room. Similarly, several ocular detection algorithms incorporate two EOG channels to ensure movements are synchronous and thus physiologically relevant.
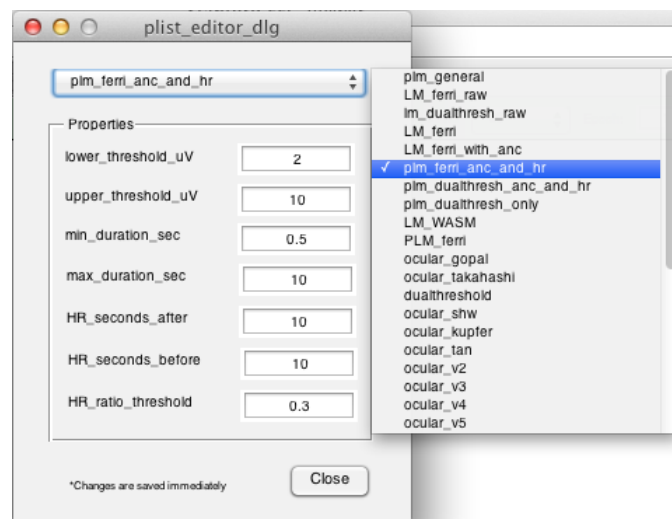
Figure 3.3: Biomarker and artifact classifier selection and parameter adjustment GUI. The classifier of interest is selected from the drop down menu at the top. Parameters associated with the selected classifier are listed with edit boxes for user adjustment. Changes take place immediately and are reflected in the viewer window(s).

A typical detector preprocesses its input signal(s) before applying classification rules. New signals can be synthesized using the intermediate, classifier processed data to give users a better understanding of the method's results. Detection parameters are stored in an extensible markup language (XML) file and may be adjusted using a common dialog window either directly from the detector toolbox, context menus, or the file itself. Figure 3.3 shows the general GUI used for adjusting the parameters of any of the SEV's

classifiers. The viewer instantly updates parameter adjustments made interactively to existing detections.

The event toolbox is easily extended using a simple MATLAB template that provides developers immediate access to the raw signal and metadata. Classifiers developed with this template interface are immediately accessible within the SEV and benefit from the MATLAB's editable environment.

3.2.3 Filter Toolbox

The filter toolbox provides a variety of filtering and decomposition tools for signal analysis and algorithm development that are summarized in Table 3.1.

Filtering methods range from conventional linear Finite-impulse-response (FIR) filters (e.g. band pass, high pass, etc.) to moving-window, non-linear filters (e.g. standard deviation, median, root mean square, and exponentiation). Channels can also be de-noised using wavelet decomposition or adaptive noise cancellation methods. Adaptive noise cancelling using recursive least squares with one or more reference channels is particularly helpful in removing unwanted ECG artifact or undesirable EEG contamination in EOG activity and vice versa. Simple linear combinations, such as summing or subtracting channels, make re-referencing signals easy. Empirical mode decomposition of signals is also possible.

Filters from the toolbox can be cascaded in any order and applied to one or several channels at a time. Swapping views between the original and filtered data is quickly accomplished through a context menu. The user may return to original, unfiltered data using a context menu or the filter toolbox. Duplicating a channel, and subsequently filtering it, provides a way to synthesize new channels for analysis and classification development. Figure 3.4 shows an example of a single horizontal EOG (HEOG) channel duplicated several times with different, selected filtering methods applied to each.
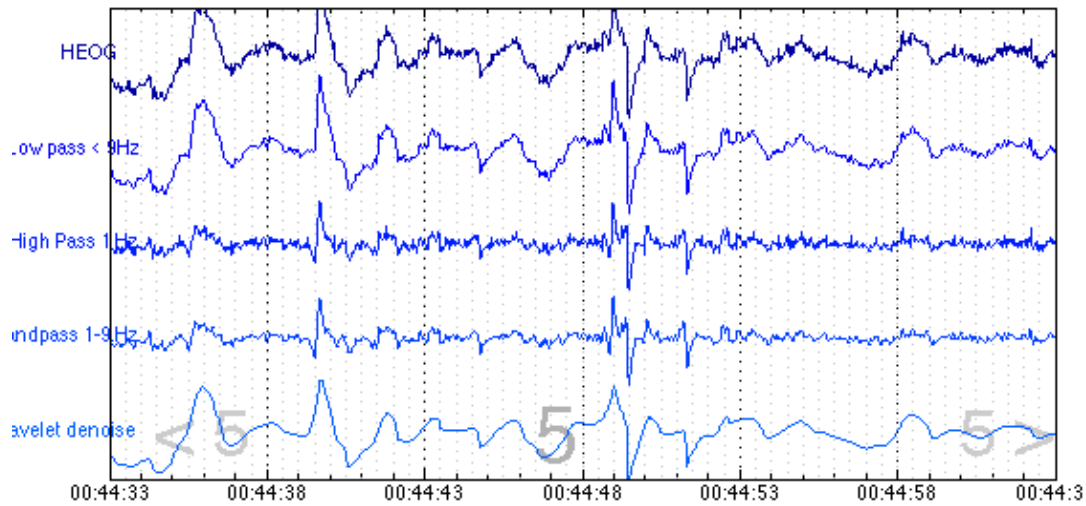
Figure 3.4: Output from filter toolbox. The HEOG channel (top) is duplicated and filtered differently to produce four new channels for visualization and analysis.

3.3   BATCH PROCESSING

The SEV's batch processing mode allows users to automate the core functionality available in the single study mode across a collection of studies. This provides a unique analysis framework not available in other tools. Previous to the SEV, scientists resorted to scripting mouse events in an effort to automate use of other software programs with limited success. The SEV's batch mode, initially developed to provide a clean, transparent interface to process sleep studies for power spectral analysis and identification of artifacts, incorporates biomarker classification for both original and post-processed (i.e. newly synthesized) PSG channels.

The batch mode, accessed from the SEV's main window and shown in Figure 3.5, provides a separate configuration GUI for establishing these automated procedures. This GUI is comprised of five sections which are listed from top to bottom: (1) cohort selection, (2) output preferences, (3) source channel synthesis, (4) classifier selection and configuration, and (5) power spectral analysis configuration. The GUI adjusts in size to

106

accommodate the number of analyses configured. The batch mode allows concurrent importation of EDF files and sleep stage file, but not other event files at this stage.

Figure 3.5: The SEV's Batch Mode provides configuration options for automating filters, classifiers, and spectral analysis of polysomnography data.

3.3.1 CONFIGURATION

3.3.1.1 Cohort selection

Cohort selection is accomplished by selecting the directory that contains the EDF sleep studies and their 30-second staging data (hypnogram file). Automated analysis requires staging files to fill event and PSD output fields, and is particularly useful for clustering results later. Channel names, obtained from the EDF header, are used to populate the various source channel pull-down menus. The user chooses which channels to analyze from these menus.

3.3.1.2 Output preferences

Output preferences, accessed through the settings GUI shown in Figure 3.6, include options for file output (e.g. directories, format, naming convention) and selection of optional statistical summary and log files. Classifier events may be exported as tab delimited text files, MATLAB binary files, or database records of a configured database. Database use requires additional installation steps (i.e. MySQL and MATLAB interface code *mym*). Raw power spectral analysis results are exported as tab delimited text files. Screenshots of classified events can also be exported in either JPEG or PNG format.
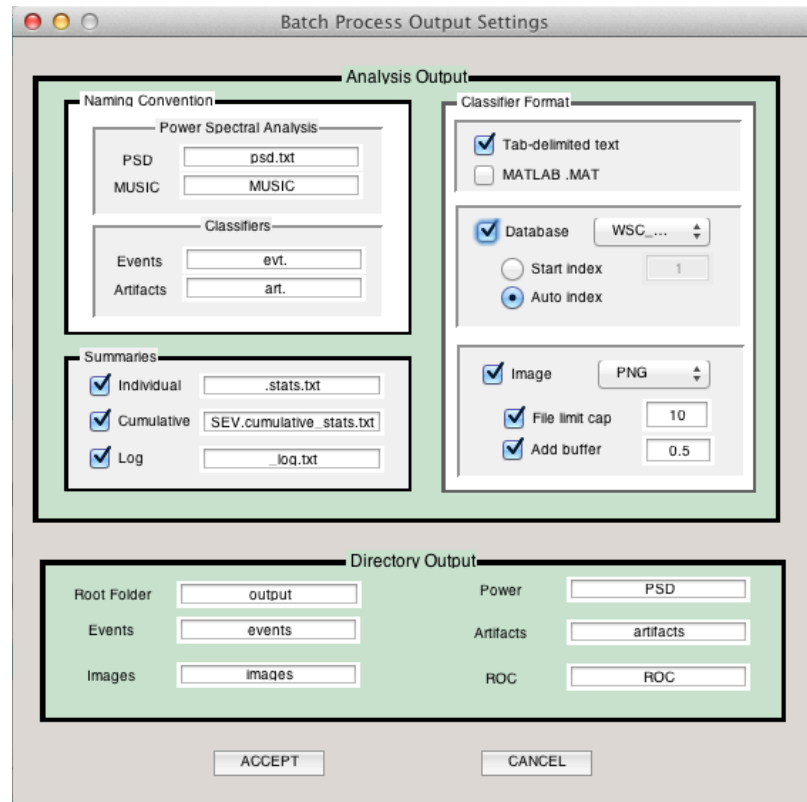
Figure 3.6: Output settings of the Batch Mode are adjusted using a separate GUI. Directory placement, file naming conventions, and selection of statistical and raw output are made using this GUI.

3.3.1.3 Channel synthesis

Additional PSG channels can be synthesized for processing. A source channel, from the EDF or a previously synthesized channel, is first identified and then filtering methods are selected using the Filter Toolbox GUI. Once configured, the synthetic channel is added to the source channel dropdown menus through the batch viewer, making it immediately available for analysis or processing by any of the classifiers or power spectral analysis tools.

3.3.2 BIOMARKER AND ARTIFACT CLASSIFICATION

Batch processing of time domain classifications can proceed in two ways that are best suited for biomarker versus artifact processing respectively. Classifiers can be selected

for use as either biomarker or artifact. The output from each either may saved as text, picture, or database entry, however the methods differ in terms of parameter space configuration (biomarkers) and integration with power spectral analysis (artifact).

Classifiers selected from the biomarker section can have their parameters configured to span a range of values. Automated processing of this parameter search spaces lets users explore a range of classifier specific parameter values for each sleep study in the batch job. The performance of these settings is readily evaluated using ROC curves, as discussed in the ROC section.

Classifiers chosen from the artifact section are integrated with power spectral analysis and do not incorporate configurable parameter search spaces. Events classified as artifact are flagged in the power spectral analysis output. In this mode, not only can SEV generated artifact files be selected, but other previously available synchronized event files, such as files containing respiratory disturbances or scored PLMs can also loaded so that data surrounding a respiratory event can flagged for removal.

3.3.3 POWER SPECTRAL ANALYSIS

The SEV facilitates power spectral analysis by calculating the power spectral density (PSD) of consecutive, fixed-width blocks of selected PSG channel(s) and saving the results to a text file. The settings GUI, shown in Figure 3.2, lets users adjust the window type, size, and interval between successive PSD calculations or periodograms. The method to calculate the PSD (i.e. Fast Fourier Transform (FFT) or Multiple Signal Classification (MUSIC) is selected from the main batch mode view along with the specific channel to analyze (Figure 3.5).

A separate text file is generated for each channel analyzed, of each patient in the batch-processed cohort. Configuration settings included in each file's header along with a row of column labels for the ensuing periodograms that occur in consecutive interval succession. For example, the second output row from an analysis conducted at 1 sec

intervals on widows of 2 sec duration corresponds to the power spectrum from 1-3 sec from the studies start. The third row corresponds to the power spectrum from 2-4 sec since the studies start, and so on. Periodogram frequency resolution is determined by the window length and sampling rate of the channel. The sleep stage, frequency bands of sleep, and flags for artifact – corresponding to each periodogram – are included as well. The files are highly suitable for analysis in other programs, like R or Excel, where custom scripts can be run to minimize variance by averaging overlapping periodograms [63], exclude artifact, and shape the analysis to a specific research thrust. For example, trends in power spectral activity can be analyzed individually, demographically (when known), or longitudinally (successive sleep studies) by time, sleep stage, or even sleep stage cycle using output files from the SEV. Database output is readily leveraged by web-based programs to quickly query and cluster detected biomarker activity, and other programs, like the PhenoFinder, can be developed in MATLAB to further visualize power spectral trends in a selected cohort[77].

3.3.4 RECEIVER OPERATING CHARACTERISTICS OPTIMIZATION VIA EXHAUSTIVE SEARCH

In addition to the single study mode, an ROC viewer is provided within the SEV for analyzing data produced in the batch mode. Figure 3.7 shows an example of running the ROC viewer on a small set of sleep studies using multiple parameter values and a gold standard to compare against. Sensitivity and specificity performance of a classifier's various parameter values are evaluated visually by selecting the values of interest from the drop down menus provided with each adjustable parameter for the ROC input under evaluation. The performance in Figure 3.7 shows the clustered performance of all the studies, but individual studies can also be shown simultaneously to spot outliers and group effects.
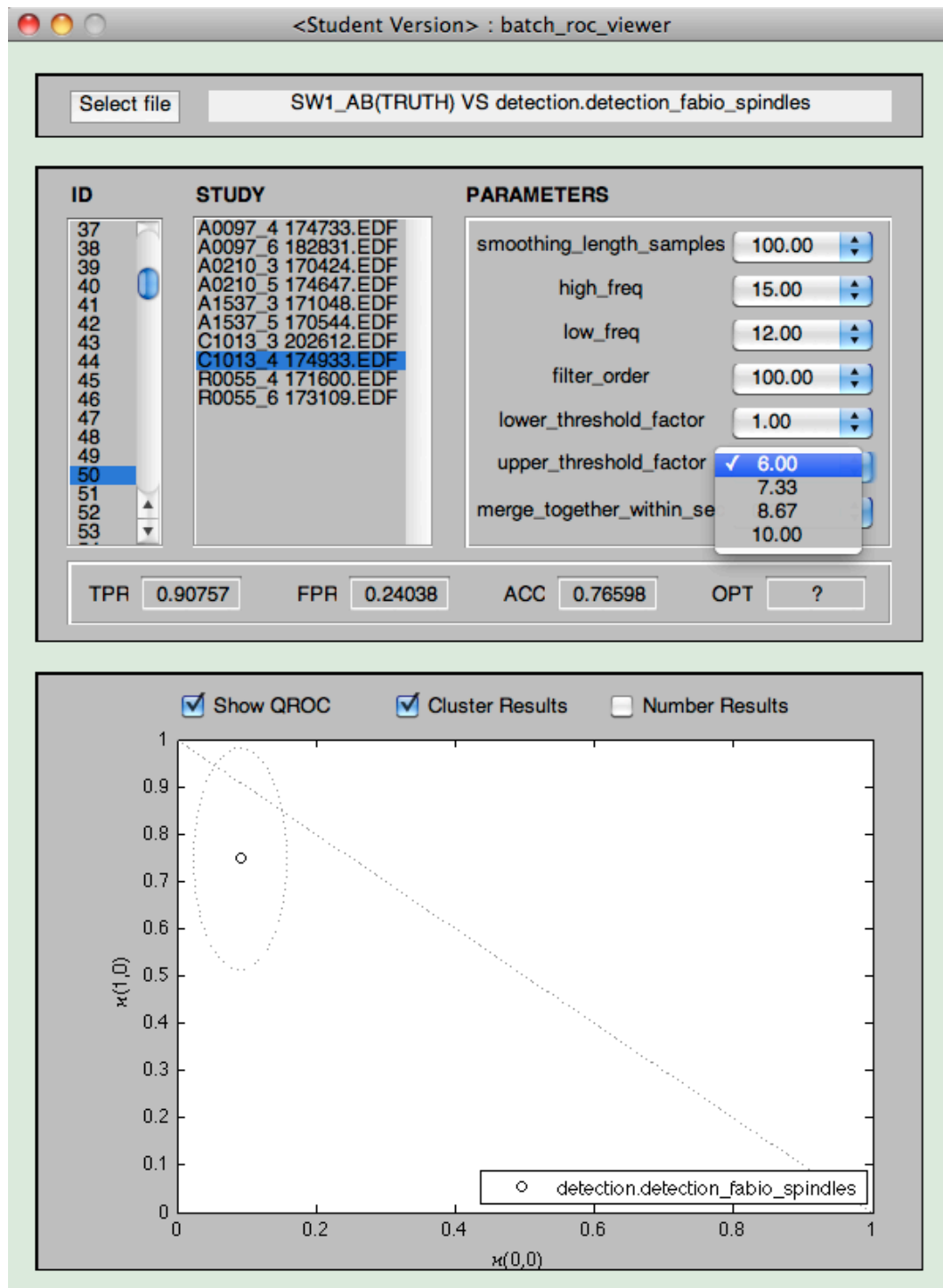
Figure 3.7: The SEV's Batch Mode ROC Viewer allows users to determine the impact of various parameter values on a classifiers performance in terms of quality ROC values (e.g. Kappa values) and test ROC values (e.g. sensitivity and specificity).

3.3.5 AUDITING

Several steps are taken to ensure parameter configurations for batch processing are stored so that results are easily replicated. Text files of classified events or power spectral analysis include header information identifying the channels run and settings used. The output filenames also identify this information, less the parameters. A log file is generated for each batch run with indication of the channel, methods, and settings used as well as the status of each file analyzed and any errors associated with it. For example, studies with missing or corrupt data that fail to load are caught and annotated in the log file with information regarding the error for post-hoc investigation as necessary.

3.4 CONCLUSION

The SEV is a MATLAB toolbox developed for single and multi-study sleep analysis. It is well suited for epidemiologic and genetic studies evaluating polysomnographic biomarkers such as power spectral characteristics and phasic activity in studies with relatively few polysomnography channels (e.g. less than 20). Raw output is provided for further statistical analysis based on the end user's specific goals. The SEV is currently being applied to several, large-scale datasets for the investigation of periodic leg movements, restless legs syndrome, REM activity in posttraumatic stress disorder, sleep duration, and spindle activity [78-82]. It is the cornerstone of the work described in subsequent chapters, and will continue to be used and improved by future students in the sleep program. Future work will be directed toward the clinical integration of the algorithms developed and validated using the SEV.

3.5 SUPPLEMENTARY MATERIALS

Supplementary materials include the following items that can be found online at http://www.stanford.edu/~hyatt4/software/sev/ or http://www.github.com/informaton/sev.

- SEV for MATLAB (.zip): This package contains the MATLAB files necessary to run the SEV software described in this chapter.

113

- SEV tutorials (.mp4): Video instruction tutorials for using the SEV are available online.

- Application programming interface (API): Programming instructions for using SEV's MATLAB classes and functions.

CHAPTER 4   VISUALIZATION OF EEG ACTIVITY FOR STIMULATING SLEEP RESEARCH

This chapter presents both an approach and a related software tool for visualizing EEG activity in large samples of sleep studies. The tool was developed through a user design study I conducted for the purpose of evaluating power spectral based phenotypes in the Wisconsin Sleep Cohort as processed by the SEV's batch mode, which is described in the preceding chapter. This chapter is published in the journal Computer Methods in Biomechanics and Biomedical Engineering [77].

4.1 INTRODUCTION

Brain-wave activity, monitored using electroencephalography (EEG) during nocturnal polysomnography (NPSG) studies, provides an informative view of sleeping individuals. The EEG reflects local changes in electrical activity across the scalp and provides insight into electric potential differences generated by the cerebral cortex. This complex activity often follows predictable patterns, which are used to *stage* an individual's sleep and determine abnormal behavior. Alterations in sleep's EEG power spectral density (PSD) have been observed in individuals diagnosed with depression, schizophrenia, Williams syndrome, autism, and developmental dyslexia. Discovering a genetic basis that links measured sleep activity between these and other disorders is of tremendous medical and societal importance [83-85].

A team of scientists at Stanford's Center for Sleep Sciences and Behavioral Medicine, including expertise in sleep medicine, neurology, biology, immunology, and genetics, has been exploring genetic contributions to EEG activity in human sleep. Such investigations require large samples to obtain to enable reliable statistics tests and a good "phenotype". A phenotype is any observable and reliable characteristic or trait, and may be produced by genetics or the environment. A genotype is a gene or DNA sequence producing a specific phenotype in the host organism. Links between genotypes and phenotypes are tested using a variety of research designs. Genome wide association studies (GWAS) can provide tremendous insight into the genetic bases of phenotypes. To perform such studies, our team also has access to advanced DNA-sequencing technology from Affymetrix and Taqman.

4.1.1 CURRENT TOOLS FOR ANALYZING SLEEP

Commercial sleep analysis programs like PhiTools' PRANA or Embla's Sandman currently provide measures for sleep data visualization. Such tools are typically optimized for rapid viewing of polysomnograms (PSG) while offering pseudo diagnostic and signal processing modules on a per study basis. The opaqueness surrounding the

116

inner workings of the software's algorithms and features is a hindrance to phenotype selection. Access to the data at intermediary stages in these programs is seldom given, and summary results are nicely presented but leave little room to verify or analyze low level results.

Stanford's researchers expressed concern with using such "black-box" commercial tools to select a phenotype. Discovering that an interesting observation is actually an artifact of the software's methodology with no physiological basis is a costly, time-consuming mistake surpassed only by not making the discovery at all.

Alternative, data transparent programs exist for phenotype selection, but come with different problems.

Programs like R or Microsoft's Excel provide means for direct low-level analysis of raw data, but offer limited support for spectral analysis. Data analysis with these programs often develops into a series of unique scripts that are continuously altered to produce new plots or examinations as new questions or ideas arise, and is a tedious, clunky process.

Given these concerns, an alternative solution for investigating the sleep EEG's PSD for phenotypes was sought and developed over a series of informal user discussions with this team of scientists.

4.1.2 CONTRIBUTION AND ORGANIZATION

The PhenoFinder, introduced here, is a tool for visualizing EEG activity in large sleep cohorts for the purpose of exploring and identifying power spectral sleep phenotypes to genotype. This MATLAB toolbox is transparent in its data manipulation and presentation functions, and remains open to adjustment or extension by others. Its graphical interface makes the crucial step of selecting and observing sleep phenotypes across a cohort of patients a quick, easy process. The software, along with a small set of anonymized PSD data (n=30), is available online at http://www.stanford.edu/~hyatt4/software/phenofinder/.

117

It is the first tool we are aware of that assists users in visualizing the sleeping brain's PSD EEG activity for phenotypes in such a large cohort of NPSG studies (n=1,836). The paper's organization continues in the following manner. Section 2 introduces dataset and gives background information on sleep. Section 3 presents the approach and methods taken to visually investigate sleep for PSD phenotypes. Section 4 presents the results with discussion, and Section 5 the conclusion.

4.2 BACKGROUND

This section gives background on the NPSG sleep data set used, its collection and initial sleep stage scoring, and its transformation to the frequency domain (i.e. the PSD).

4.2.1 DATA SET

The EEG data set examined here comes from the Wisconsin Sleep Cohort (WSC), which was introduced in Chapter 1.4.1. Chapter 1.1.1 describes manual process for scoring sleep stages in 30-second intervals or *epochs*. To briefly summarize, the WSC includes a collection of NPSG based sleep studies taken longitudinally (at four-year increments) from a general population of middle-aged adults. It was established in 1988 from a sample of employees of four state agencies in south central Wisconsin, USA. WSC's NPSGs were initially stored on paper until digital collection began in 2000 with a 16-channel PSG recording system (16-channel Grass-Telefactor Heritage digital sleep system Model 15). A total of 1,836 studies, conducted between 2000 and 2008, from the WSC are used in this chapter. The sample represents 1,054 patients: 571 men between 39 and 78 years of age (mean=56.7 std=7.7), and 502 women between 37 and 78 years of age (mean=55.3, std=7.6).

4.2.2 DATA TRANSFORMATION

The EEG's PSD is calculated for each study using the Stanford EEG Viewer (SEV) [86]. The SEV, introduced in Chapter 3, is a PSG processing MATLAB toolbox with a batch that automates power spectral analysis along with classification and detection algorithms

across arbitrarily sized datasets. It was configured for two-second periodogram averaging with one-second overlap to calculate the PSD with 0.5 Hz resolution [63]. These periodograms and corresponding sleep stage scoring are saved to text files with additional flags to indicate the presence of ocular or muscle artifact, which could contaminate the periodograms. The C-3 and O-1 EEG outputs are saved to separately named files for each patient, which the PhenoFinder parses and condenses on initial loading for efficient use of either EEG signal. Power above 30 Hz is less frequently used in clinical practice and is not considered here. (The American Association of Sleep Medicine recommends low pass filtering the EEG at 30 to 35 Hz [2].)

4.2.3 RELATED WORK

Recent approaches to phenotyping EEG include the use of high-density EEG arrays [87], mouse models [88], and neonatal examination [89]. Davis *et al* review of experimental findings re delta EEG in animals (e.g. mice, rats, rabbits), neonates, adolescents, and adults under numerous conditions and settings [90].

Twin studies have shown that across- and within-night variation in NREM frequency band measures is highly heritable [85, 91-93]. In addition, certain genetic disorders are associated with modification of the sleep EEG. Bodizs *et al* found EEG PSD changes (at 8 and 16 Hz) in Williams Syndrome (See also [94] and [95]).

4.3 METHODS

Microsoft's Excel and R were used primarily to analyze and present the PSD data at first. However, investigating phenotypes with these tools was a slow, laborious process. The team wanted to ensure the pros and cons of candidate phenotypes were fully addressed and considerable time and effort was spent between meetings to analyze the dataset in different ways. Efficient, flexible analysis of the dataset was not possible under this workflow and a new solution to streamline the process was needed.

4.3.1 DESIGN DEVELOPMENT

PhenoFinder's design began with a brainstorming session with the end-users that focused on how to best visualize and explore the EEG's power spectrum to identify good phenotypes. Two motivating questions from this session were: (1) How can individuals be shown similar, and (2) How can they be shown different? Specific design requirements to answer these questions began next by identifying specific phenotypes already of interest and then deciding on what core features were necessary to visually evaluate their significance. Stage 2 sigma power, believed to correspond mostly with spindle activity, was one of the initially selected phenotypes, and alpha power during NREM sleep was another. Using these specific examples, the following design requirements emerged:

- Normalize and display data in a way that enhances meaningful comparisons of patients and encourages exploration of phenotypes
- Interactive, intuitive interface
- Multiple phenotype views
- Data integrity confidence measures

The project developed iteratively with working copies of the software provided periodically to the users for feedback to the primary designer (HM). The feedback was often shared with supporting members of Stanford's Department of Computer Science Department who provided input on data visualization driven responses (see acknowledgements).

4.3.2 NORMALIZE AND DISPLAY DATA TO ASSIST COMPARISON OF PATIENT DATA

A person's sleep is as unique as their fingerprint and some measure of smoothing or grouping must occur to observe the physiological trends and similarities between people. Data must be reduced enough to find trends, but not so much that meaningful variability is lost. The dimension of time is collapsed into scored sleep stage. Periodograms are identified and averaged according to sleep stage. Larger categories for NREM sleep and total sleep (NREM + REM sleep) are also provided.
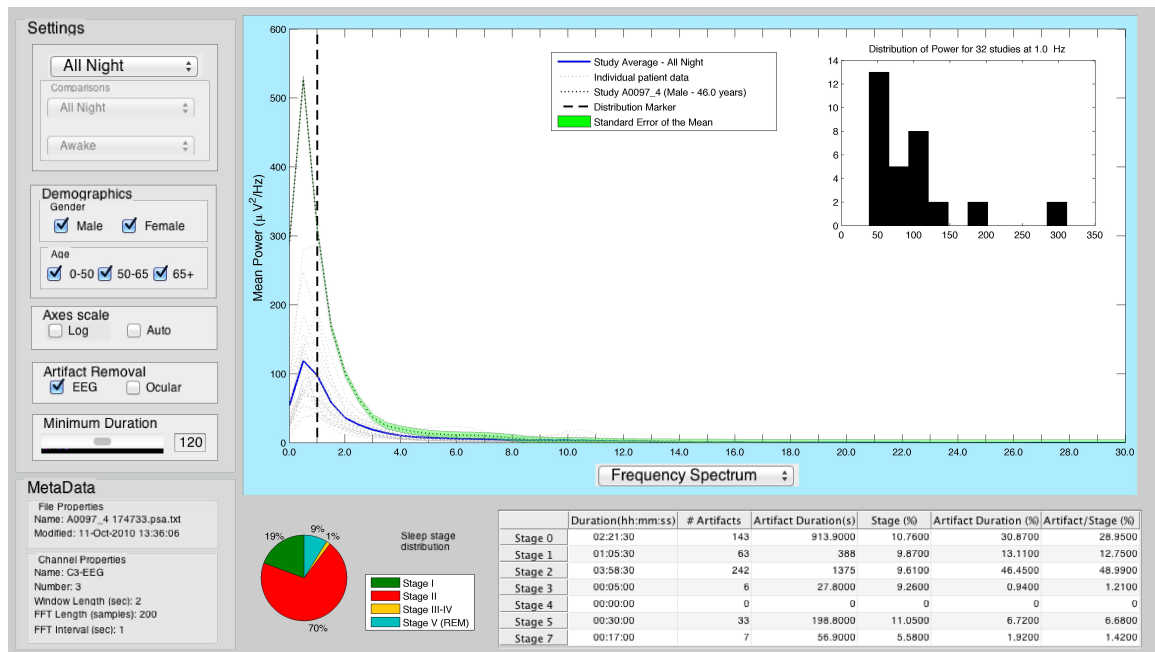
Figure 4.1: PhenoFinder's configuration here displays the power spectral density measured across all-subjects' data provided (n=32). Views can be further refined by age, gender, artifact exclusion, and minimum or maximum of the sleep category selected. Sleep categories include All Night, Wake, Non rapid eye movement sleep, rapid eye movement sleep, Stage 0, Stage 1, Stage 2, Stage 3-4, and comparisons between them. Phenotypes are queried using the drop down menus and checkboxes provided. Here the all night sleep PSD is shown across the frequency spectrum of interest, with the inset histogram revealing the patients PSD distribution at 1.0 Hz.

Users select the sleep stage of interest from the top left, drop-down menu seen in Figure 4.1where the entire night's sleep PSD is selected in the supplementary dataset (n=32). The relationship between sleep stages and groupings can also be visualized. Selecting *ratio*, *difference*, or *relative difference* from this menu activates two additional drop-down menus for choosing comparison categories. Figure 4.2 shows the ratio between all night power and wake. The ratio's decay across the frequency spectrum highlights the preponderance of lower frequency activity associated with sleep, and higher frequency activity seen in wake. REM sleep exhibits "wake like" EEG activity, and the ratio between NREM sleep and wake is even more pronounced. The relative difference phenotype normalizes the difference in power between two selected categories by the

sum of their power. Figure 4.3 shows the relative difference between the power measured during sleep to that measured in wake for 1,836 studies. This gives another, macro-level view of differences in frequency activity between sleep and wake.
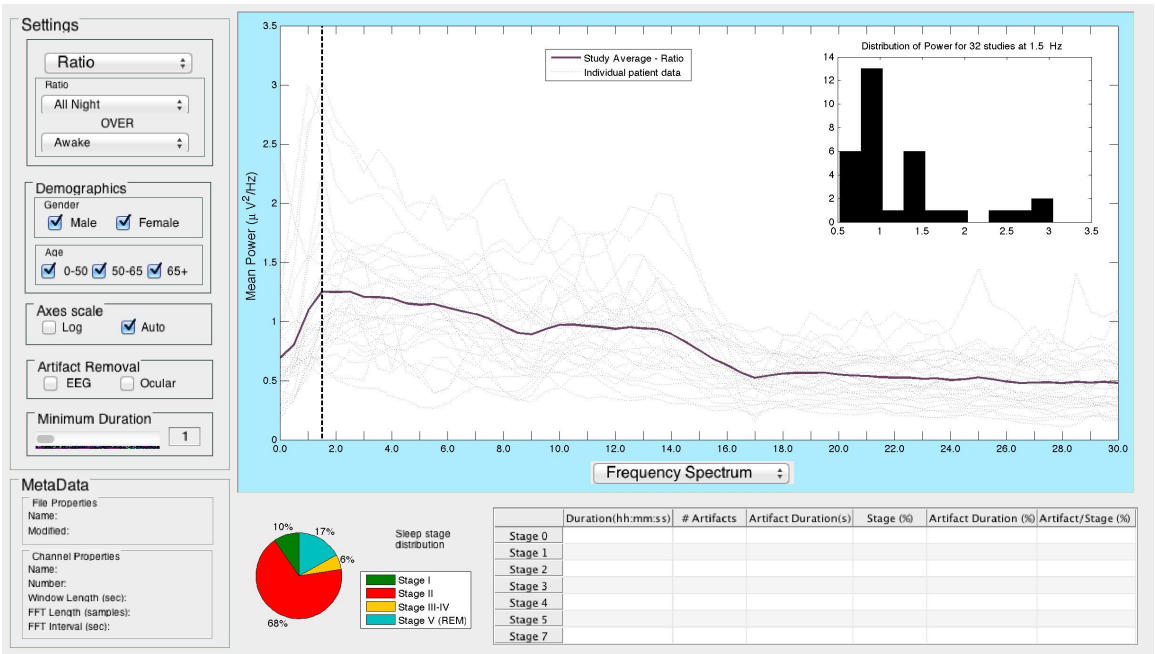
Figure 4.2: The ratio of EEG power during sleep (*All Night*) to wake (*Awake*) is shown for small sample of patients in the frequency range 0 to 30 Hz. The light gray, dotted lines represent this phenotype for each individual patient. The thick, purple line shows the group's average ratio. The inset histogram gives the power distribution for these patients at 1.5Hz. A vertical, black striped line is placed at this frequency in the main plot. The magenta line traces the average power for these patients across the spectrum and highlights the increased low frequency power common to sleep compared to the high frequency activity commonly associated with wake and rapid eye movement sleep.

Queried phenotypes are shown for each individual as small, gray-dotted lines while a thick, colored line shows the group mean. A unique color is assigned to each stage category, and the mean line color matches the stage. Stage colors were picked based on user preference and feedback. A sequential coloring scheme corresponding to sleep stage number and color frequency was avoided both because of the poor visual distinction of similar color mappings found by Ware [96] and Rogowitz, Treinish, and Bryson [97] and

because sleep transitions are rather sporadic. A separate color was chosen to specify stage comparisons (e.g. ratio or difference) and remind users they were not viewing a stage in isolation.

The PSD for known frequency bands may also be selected. The sigma band (i.e. 12 - 16.0 Hz) was added to accommodate the users' interest in spindle activity. The default sleep bands defined within the PhenoFinder are defined as follows:

- Delta: (0.5-4.0 Hz)
- Theta: (4.0-8.0 Hz)
- Alpha: (8.0-12.0 Hz)
- Sigma: (12.0-16.0 Hz)

The users were content with this breakdown and did not desire grouping the remaining frequencies into a band (i.e. remaining Beta band). Multiple bands may be selected for simultaneous display, and a dynamic legend updates relevant descriptors in the main window.

The mean power from previous stage selections can be retained for comparison to newer selections. The previous mean's line color, corresponding to its original stage selection, is retained but its thickness decreases with each subsequent query.

Fgure 4.4(b) shows Stage 2, delta band power. The mean power line is blue, corresponding to the tags Two's color category, and is thickest because it is the most current query. The number of phenotypes to retain for viewing is adjusted in the settings menu. This feature was requested during initial meetings, but not actively used.
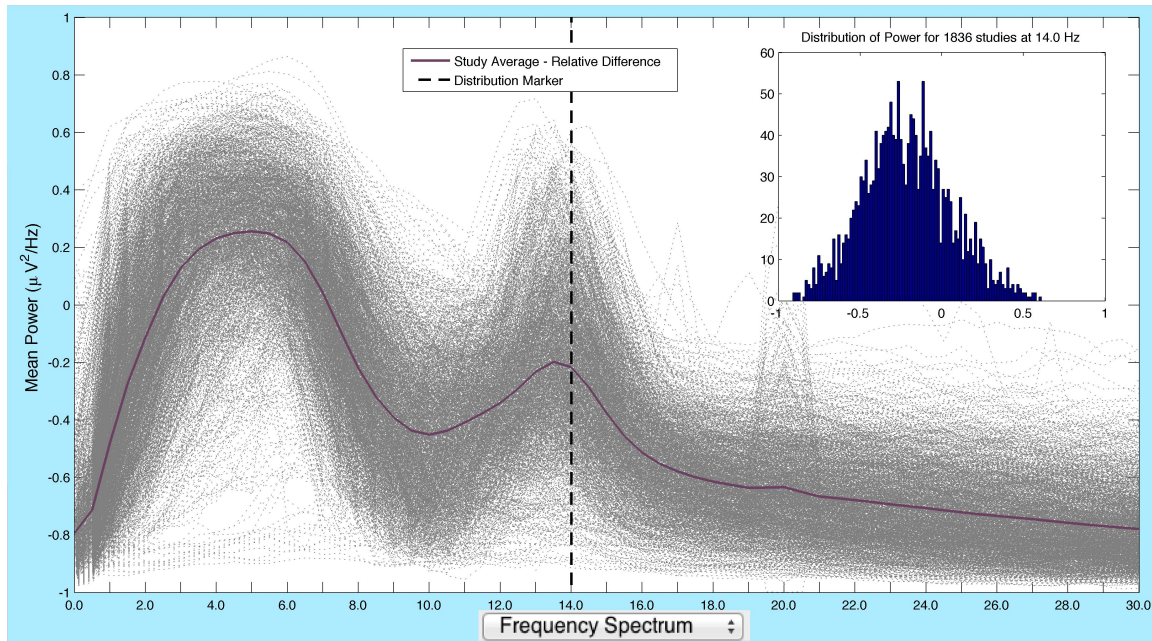
Figure 4.3: The relative difference phenotype here shows the difference in power measured during all night sleep to power measured during wake divided by their sum. The inset histogram shows a non-Gaussian distribution for this group (n=1836 studies) in the sigma power band (12-16Hz) at 14.0 Hz.

4.3.3 INTERACTIVE, INTUITIVE INTERFACE

Initial interface revolved using the described widgets and mouse interactions with the main display. The PhenoFinder highlights studies selected by the user with the mouse by increasing the lines thickness and contrast; darkening it while lightening others. After an informal software review, one user expressed interest in seeing the power distribution of patients for selected frequencies, which was hard to quantify from the individual power profiles. An inset histogram, with a context menu for hiding (

Fgure 4.4(b)), was added to the main plot window with a thick, black-striped vertical line placed at the frequency of interest. In Figure 4.2, this vertical striped line is placed at 5.5 Hz, which corresponds to the PSD distribution shown in the inset histogram for this frequency. Black is a relational color coder between the histogram and vertical line. The histogram's title also explicitly states the observed frequency.

During a follow-up session showing the added histogram, the same user was observed repeatedly pressing the left and right arrow keys in an effort to observe the distribution of adjacent frequencies instead of selecting them with the mouse. Rather than coach them out of their intuition, arrow key control was implemented and quickly became the standard practice for frequency transitions.

4.3.4 MULTIPLE WAYS OF EXAMINING QUERIED PHENOTYPES

PhenoFinder checkboxes, seen on the left side of Figure 4.2, let users view phenotypes on linear or logarithmic scales and retain or automatically resize vertical axes limits. The option to select between median or mean result views was originally placed on the left side, but later moved to a settings menu accessible from the *File* menu bar due to limited use. Checkboxes for age and gender provide further filtering of viewed data, as do artifact removal checkboxes that allow retention or removal of sections of data contaminated by ocular or muscle activity.

In addition to these checkboxes, MATLAB's standard, toolbar-accessible plot manipulation features are retained for viewing phenotypes. Despite the cognitive load claimed by some [98], the zoom-and-pan feature was used heavily by users who preferred exploring the frequency spectrum in this manner more than selecting between frequency bands from the drop-down widget along the x-axis.

4.3.5 REVEALING DATA INTEGRITY

Several steps are taken to address and visualize the data's integrity. First, the artifact removal checkboxes lets users see where the SEV's artifact detectors are working and where there is no effect. This option to examine the presence or absence of artifact led to an unexplored line of phenotype examination. The scientists began considering physiological differences, which could lead to excessive ocular or muscle artifact during sleep. Secondly, a slider widget seen at the bottom of the settings panel in sets the minimum or maximum observed epochs necessary for phenotype retention. PSD profiles are weighted equally in the PhenoFinder regardless of whether the profile is calculated

from one 30 sec epoch of sleep or two hours. The slider context menu, activated by right clicking, gives user the choice of interpreting the slider's value as a minimum or maximum value to bifurcate the group and remove unwanted studies. In Figure 4.1, the displayed studies have a minimum of 120 epochs (i.e. one hour) of sleep. The use of animation or blinking to highlight changes between sliders was suggested and discussed, but not implemented based on the users' belief that such features would be more distracting than helpful.

The pie chart and table, both below the PhenoFinder's main view window, provide further insight into patient sleep. The pie chart shows the sleep stage distribution, as a percent of total sleep, for the selected patient. The chart's colors correspond directly to stage color, and this color-stage pairing is shown in the legend next to the pie chart. Standard pie *exploding* gives visual redundancy to selected stages. The table to the right of the pie chart gives the exact time spent in each sleep stage and, in cases with a selected study like Figure 4.1, artifact statistics broken down by sleep stage.

Lastly, a green, transparent patch overlay, showing the standard error of the mean (SEM) highlights the selected PSD profile's variability. The narrow green patch in

Fgure 4.4(a) shows the selected line to be a stable representation of the patient's power profile.
Fgure 4.4(b), however, shows a large green patch that reveals a high SEM and greater uncertainty in the observation.
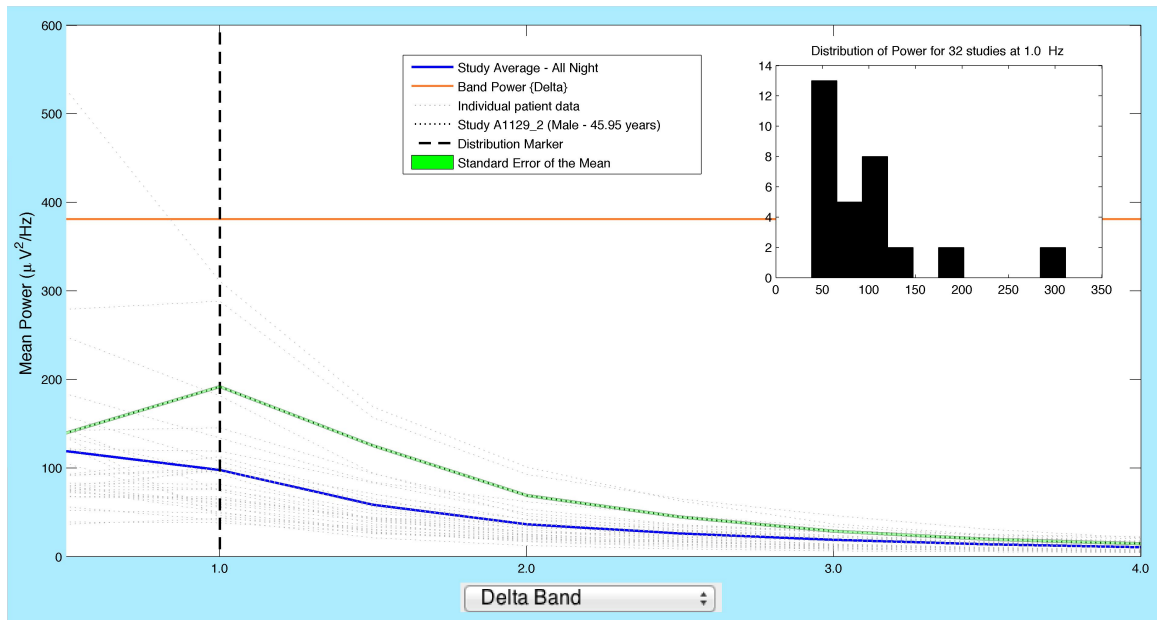
a. Delta band power for all sleep with power distribution shown (inset) for 1.0 Hz.
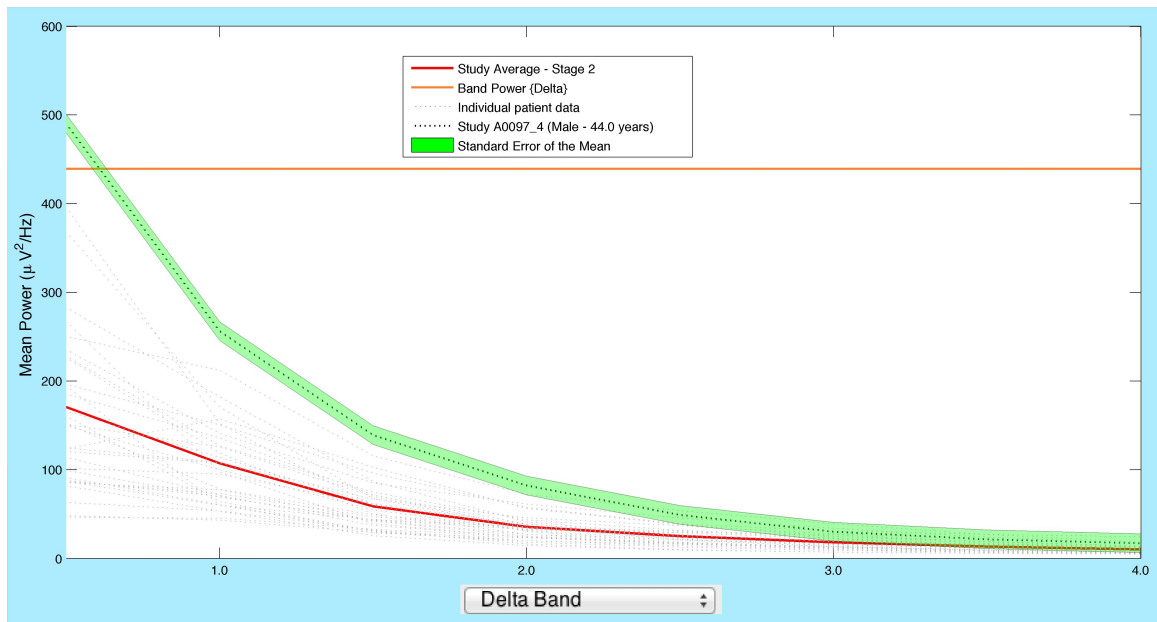
b. Delta band power for Stage 2 sleep without inset histogram.

Fgure 4.4: Data validation methods include the standard error of the mean (green patch), histogram distribution (inset), patient meta data, and stage distribution.

4.4 RESULTS

PhenoFinder is a MATLAB toolbox introduced here for visualizing EEG brain activity in large data sets for the purpose of exploring and selecting power spectral sleep phenotypes to associate with genotype.

PhenoFinder's design requirements and implementation follow the suggestions of Munzer (2008) [99] where, in our case, several informal user studies provided useful, productive interactions leading to the final software and selection of Stage 2 sigma power for GWAS.

Stanford's team of researchers were pleased with the PhenoFinder, especially with modifications based on their suggestions and feedback. Both the software itself and its collaborative design were instrumental in assisting the phenotype selection process. Initially, only power bands of specific sleep stages were available for selection. However, after several end-user discussions with early versions of the software, new possibilities began to emerge. Questions regarding the inter-stage relationships came up and led to comparison of the ratios and differences between sleep stages and also consideration of artifacts as phenotypes. Several new phenotypes were investigated including the log ratio of sigma power to all night sleep and muscle and ocular activity during sleep. Stage 2 sigma power was selected as one of the best phenotype to pursue for genetic analysis based on its stability across time and its correlation with spindle activity, a feature of interest to sleep researchers [78].

4.5 DISCUSSION

John Tukey's exploratory recommendation was taken to first examine the data and then develop hypotheses from the observations [100]. Examining human sleep data as a first step to understanding the physiologically mechanisms behind it has both merits and drawbacks.

128

Using PhenoFinder, researchers were able to see sleep EEG spectral profiles in nearly two thousand sleep studies at a glance. Hypotheses were quickly confirmed or rejected, and new hypotheses were often formed and examined with similar speed. The flexibility and efficiency by which different patterns and profiles could be viewed is a significant departure from previous studies in the literature, which have been based on smaller samples. The team also preferred staying *close* to the data, which they felt possible with the PhenoFinder where they had a degree of control in the process.

A drawback to this approach is the lack of clearly defined boundaries to focus the lines of investigation. While so many observations could be observed quickly, the number of phenotypes considered and evaluated grew, which led to increased discussion and formulating hypotheses on previously unconsidered phenotypes.

Toward the end, the number of questions to consider grew to such an extent that selecting a final phenotype became difficult given the larger number of choices. One phenotype suggested at the outset, sigma power during Stage 2 sleep, was selected for GWAS. The considerable amount of time spent examining this phenotype prior to the PhenoFinder's development, and subsequent to its visual support, made it a natural selection in the midst of so many choices.

4.5.1 ADDITIONAL CONSIDERATIONS

Data reduction of the time elapsed in each study is accomplished by collapsing the PSD into their manually scored sleep stage. Alternatively, sleep studies could be aligned to the point of sleep onset and compared using a linear, per-hour-elapsed basis with a heat-map to show power normalized by frequency across the night. This would provide more resolution into time-of-night changes in the data, but the users felt it would be harder to interpret than categorical sleep stage and was abandoned. Another possibility, representing sequential time with a third dimension was also considered and abandoned. Here the problems of occlusion and complexity were believed to outweigh any benefits. Arguments for the merits of 2D over 3D in data visualization can be found in [101].

Pattern classification methods are frequently employed to identify biomarkers of interest in the EEG. However, in this case, the users were not interested in automated approaches and declined offers to detect and highlight phenotypes of interest in the PhenoFinder, such as notable PSD distributions at certain bands or sleep stages. A primary concern expressed by the users was that the indicators would bias their judgment, making them less open to exploring alternatives. Another concern was the increased probability of making a poor classification (i.e. something artifactual which would not have been considered otherwise) associated with large number of phenotype configurations to test.

4.5.2 FUTURE WORK

The PhenoFinder can be extended in several ways. In addition to selecting frequency bands and panning through phenotypes with the mouse, the users wanted a feature that would allow the start and end frequencies to be set directly. Another request was for an overview window to see the outside their currently selected view (e.g. when using pan-and-zoom). And lastly, other users may want to see power distributions of interest highlighted for them automatically instead of manually selecting through the frequency spectrum. A heatmap or binary display along the bottom of the main view would likely be helpful here.

4.6 CONCLUSION

PhenoFinder is a software tool for visualizing the PSD of EEG recorded during sleep across large data sets. It was developed according to criteria generated from several informal interactions with four primary users investigating genetic markers for measurable brain wave activity in sleep using GWAS. The viewer represents a collaborative effort between members of Stanford's Center for Sleep Sciences and Departments of Electrical Engineering and Computer Science. Initial design requirements and follow-up interactions led to the consideration of new phenotypes and a specialized product targeted for researchers investigating brain activity in sleep as characterized by the EEG's PSD.

130

In summary, a data visualization tool, the PhenoFinder, was developed to help researchers find clinically meaningful or heritable patterns in brainwave activity during sleep in a cohort of 1,836 nocturnal polysomnography studies. The interactive software lets researchers quickly view and explore various electroencephalography power spectral density patient profiles and select desired phenotypes for genome wide association or sequencing. The design study addressed here evolved over an iterative process of informal studies conducted with end-users from Stanford's Center for Sleep Sciences and Behavioral Medicine and focused on highlighting the similarities and differences in patient data. Several new hypotheses were formed and new phenotypes considered during the design process. The software is primarily for domain experts, however a lay user may find exploration of demographic changes in sleep insightful. It was made available online with a small companion dataset at http://www.stanford.edu/~hyatt4/software/phenofinder/. A peer-reviewed publication was generated and is available in the journal Computer Methods in Biomechanics and Biomedical Engineering.

4.7   SUPPLEMENTARY MATERIALS

Supplementary materials include the following items that can be found online at http://www.stanford.edu/~hyatt4/software/phenofinder/.

- PhenoFinder for MATLAB (.zip): This package contains the MATLAB files (.m, .fig) necessary to run the PheonoFinder software described in this chapter.

- PhenoFinder tutorial (.pdf): Instructions on using the PhenoFinder.

- Sample Data (.zip): The .zip file (~250MB) includes MATLAB .mat file format for PSD data and .txt format for deidentified demographic data from 32 studies.

Developers wishing to contribute further to this project should use the open source repository at http://www.github.com/informaton/phenofinder.

CHAPTER 5   DESIGN AND VALIDATION OF A PERIODIC LEG MOVEMENT
DETECTOR

This chapter introduces and discusses the automatic periodic leg movement (PLM) detection algorithm that I built to score periodic leg movements (PLMs) in the general population and a clinical setting. I also investigated the interaction of cardiac activity and respiratory events in the context of PLMs with and without sleep-disordered breathing. The final detection algorithm, which was designed and validated using both the Wisconsin Sleep Cohort and the Stanford Sleep Cohort, is currently being implemented at Stanford's Sleep Disorders Clinic.

5.1 INTRODUCTION

Periodic Limb Movements (PLMs) are episodic, involuntary muscle contractions that occur during sleep. Restless Legs Syndrome (RLS) is often associated with PLMs, with four out of five patients diagnosed with RLS exhibiting PLMs[7]. However, PLMs can also occur without RLS symptoms. Because the movements may be accompanied by an arousal or sleep fragmentation, a PLM Index (PLMI) over 15 is likely to have an effect on an individual's overall health and wellbeing[8]. Little is known about the cause of PLMs or their impact on daytime sleepiness or insomnia symptoms. A study review of cardiac risk for RLS and PLMD found associations between PLMs and congestive heart failure[9]. Additionally, patients with RLS were at higher risk for heart disease and hypertension[9]. PLMs are known to be associated with several other disorders and pathologies such as depression, cardiovascular disease, rapid eye movement (REM) behavior disorder, narcolepsy, Parkinson's disease and multiple system atrophy[11-14].

PLM scoring rules have evolved over time and are based on the amplitude and duration of the event as well as the time between limb movements. In 1993, the Atlas Task Force, part of the American Sleep Disorders Association (ASDA, now AASM), defined leg movements (LMs) in polysomnograms (PSGs) as increased EMG activity lasting between 0.5 and 5.0 seconds (sec), in excess of 25% of the recorded voluntary flexion during calibration[102]. LMs associated with respiratory events were scored separately. The 2007 AASM Manual for Scoring Sleep, defines a significant LM as a period of 0.5-10 sec where EMG activity recorded from the left or right anterior tibialis (LAT/RAT) exceeds 8 μV above baseline and then falls below 2 μV from baseline for 0.5 sec or longer[2]. In both definitions, PLMs are defined by the consecutive sequence of four or more LMs whose inter-movement intervals are between 5 and 90 sec. The 2007 classification excludes all LMs 0.5 sec before, during or 0.5 sec after a respiratory event[2].

Three other automatic PLM detectors have been created to identify PLMs during sleep. These detectors were designed and validated using clean PSG data in relatively small groups of RLS patients versus controls. In 1996, Tauchmann proposed a PLM detection algorithm where optimal parameters were determined using a training-validation split of 1,671 and 1,740 visually scored LMs (respectively) from five PSGs[67]. Wetter adjusted this algorithm in 2004 and validated it using 8,300 visually scored LMs from PSGs of 10 patients diagnosed with RLS[68]. In 2006, Ferri proposed a detector with parameters optimized using ROC and validated against visually scored LMs in 15 patients diagnosed with RLS and 15 controls[66]. These detectors did not need to address LMs surrounding respiratory events because Sleep Disordered Breathing (SDB) was an exclusion criterion.

In this chapter, I report on the design and validation of a new PLM detector using large epidemiological and clinical samples totaling 1,073 subjects. The chapter is organized as follows. Section 5.2 presents the methods and design decisions taken to build the PLM detector in the presence of various problems related to signal quality and sleep disorders, such as SDB. Section 5.3 presents the final PLM detection algorithm and its validation by comparison to gold standard, manual scoring. Discussion of the study is provided in Section 5.4, followed by concluding remarks in Section 5.5.

5.2   METHODS

5.2.1 COHORTS USED IN THE ANALYSIS

Nocturnal polysomnograms (PSG) from the WSC[43] and the SSC were used in this study. Electroencephalography (EEG), electrooculography (EOG), and chin electromyography (EMG) were used to score sleep stages for each 30 second epoch using standard R&K criteria[1].

As mentioned in Chapter 1.4.1, volunteers were selected from the WSC with PSG files available between 2000 and 2004 (n=1,073). The timeframe was selected to be closest to a RLS survey mailing performed in 2003 [103]. Table 1.1 presents demographic data on

this selection of subjects all together, and with subjects stratified by SDB (i.e., apnea-hypopnea index > 15). In this cohort, LMs in PSGs were originally scored per ASDA 1993 criteria but using 50% instead of 25% of recorded voluntary flexion during calibration as the threshold. Additionally, the WSC scoring policy counts LMs within 4.0 sec of each other as one LM. LMs associated with respiratory events were variably scored. We considered these manually scored LMs as a preliminary standard.

Table 5.1: Wisconsin Sleep Cohort (WSC)

	RLS*, AHI≤15 (20)	AHI≤15 (20)	AHI>15 (20)
Demographics			
Age	54.7 ±0.67	55.1 ±0.63	54.5 ±0.72
Sex, Male (%)	50.0%	50.0%	50.0%
Clinical Data			
BMI (kg/m²)	31.1 ±1.25	31.1 ±1.25	35 ±1.43
AHI	5.64 ±0.96	5.18 ±0.80	31 ±3.97
AHI > 15 (%)	0.0%	0.0%	100.0%
Polysomnogram			
TST (hour)	6.33 ±0.23	6.41 ±0.19	6.21 ±0.24
Stage 1 (%)	9.2 ±0.96	9.1 ±1.44	11.9 ±1.52
Stage 2 (%)	69.5 ±1.85	70.5 ±1.69	68.6 ±0.99
SWS (%)	2.8 ±1.00	4.4 ±1.67	3.8 ±1.00
REM (%)	18.4 ±1.40	15.6 ±1.09	15.3 ±1.12

Data are mean ± Standard Error Mean, or percentage. The number of subject used for calculations are shown in parentheses. *= Presence of RLS symptoms, as described in the material and method section. AHI is the apnea-hypopnea-index calculated as the number of manually scored respiratory events per hour of sleep. Patients using CPAP were excluded from AHI categories. BMI is body mass index. TST is total sleep time.

Table 5.1 describes a subset of the WSC data identified to be re-scored to create a PLMI gold standard. The WSC gold standard sample contains sixty age and sex-matched patients selected from one of three groups: (1) RLS symptoms without SDB (n=20); (2) OSA without RLS symptoms (n=20); (3) neither RLS symptoms nor SDB (n=20). Patients in the WSC were identified as having "RLS symptoms" based on responses to a questionnaire sent to the entire parent cohort in 2003 as described in a previous study[103]. The questionnaire did not address all RLS diagnostic criteria put forth by the

135

National Institutes of Health[104], notably it did not ask for the symptoms to be worse at night. For this reason, we called patients positive for these questions as having "RLS symptoms". Patients with "RLS symptoms" reported they felt: (a) repeated urge to move legs and (b) strange and uncomfortable feelings in the legs weekly or more often plus that these feelings (c) got better when they got up and started walking and (d) disrupted their sleep. SDB was defined based on an AHI cut off of 15 events per hour. In this validation subsample, a Registered Polysomnographic Technician (EL) scored PLMs according to AASM 2007 criteria with minor modifications based on her professional judgment. Collectively, the WSC gold standard contains 5,387 PLMs.

The SSC, described in Chapter 1.4.2, is a naturalistic sample of 760 successive patients, including a wide range of sleep disorders, recruited to the Stanford Sleep Disorders Clinic and who had a nocturnal PSG from 1999-2007[44]. Table 1.2 reports on summary statistics for the SSC, broken down by diagnostic category.

A subsample of the SSC data, enriched in specific sleep pathologies was re-scored manually by a sleep physician (SL) as a secondary gold standard according to AASM 2007 criteria. The SSC gold standard contains 18 studies selected from age- and gender-matched patients with the following diagnoses (i.e. three patients per group): Insomnia, Narcolepsy, REM behavior disorder, RLS, SDB, and other (head trauma with excessive daytime sleepiness, depression, and night terrors).

5.2.2 DESIGN AND PRELIMINARY TESTING OF THE DETECTOR

The WSC was primarily used to established functionality of the detector, while the SSC sample was used both as a validation sample and to check whether the detector was robust to the presence of sleep disorders. To establish the detector, we first used the AASM 2007 rules for PLMs[2], but used a duration criterion of less than 5 sec (versus 10 sec). To meet the LM criteria, EMG activity recorded from the left or right anterior tibialis (LAT/RAT) needed to exceed 8μV for 0.5-10 seconds and then fall below 2 µV from baseline for 0.5 sec or longer[2]. PLMs are defined by the consecutive sequence of

136

four or more LMs whose inter-movement intervals are between 5 and 90 sec. The 2007
definition excludes all LM 0.5 sec preceding or following a respiratory event.

Table 5.2: Characteristics of previously published PLM automatic detectors

	Classification rules	Preprocessing	Postprocessing
All	Duration between 0.5 and 10.0 s (Amplitude criteria in reference to baseline)	Notch filtering for power line interference	
Tauchmann (1996)	Onset-offset amplitude > 7 μV Arithmetic mean > 5 μV	Low pass filter (f$_c$=16Hz) Rectify	Merge consecutive activity within 0.15 sec
Wetter (2004)	Continuous burst activity (standard deviation of amplitude>0.6 μV) for 0.4 s of 0.5 s segments.	High pass filter (f$_c$=16Hz) Rectify Truncate to 30 μV 16 ms standard deviation window	Merge consecutive activity within 0.5 sec
Ferri (2005)	Onset amplitude > 7 μV Offset amplitude < 2 μV	High pass filter (f$_c$=16Hz) 0.5 s moving average filter for offset amplitude	

To test a first iteration of these rules, an initial version of the detector was used on the
entire WSC dataset. Both the automated and manually scored results were examined to
identify outliers and potential artifact issues. We also compared different iterations of
our detector with other established detectors from Tauchmann[67], Wetter[68], and
Ferri[66] (Table 5.2).

5.2.3 TESTING OF OTHER DETECTORS

The Stanford's EDF Viewer (SEV) was used to visualize PSG epochs and LMs detected
by manual scoring and by individual detectors. SEV is a PSG processing MATLAB
toolbox with a batch mode that automates power spectral analysis along with
classification and detection algorithms across arbitrarily sized datasets[86] The original
PLM detection algorithms from Tauchmann[67], Wetter[68], and Ferri[66] were
implemented in the SEV as presented in their publications. Care was taken to follow the
same pre-filtering steps when possible. For example, differences in sampling rates and
electrical standards between countries were accounted for (e.g. 50 Hz interference vs 60

Hz power line interference). As mentioned earlier, LMs were manually scored as part of the original dataset according to WSC 1995 scoring guidelines[105], a modified version of ASDA 1993 criteria. To compare performance of the various detector, ROC curves were drawn against 119,277 originally scored LMs. Sensitivity, specificity and their quality indices or weighted Kappa statistics[46], Cohen's Kappa, positive and negative predictive values, and total LM count were explored using ROCs by leg movement and by subject (Table 5.3).

Ferri and Wetter's detectors looked the most promising in terms of sensitivity (see Table 5.3), but these detectors identified almost ten times more LMs than manual scoring. The abnormally high performance is a problematic feature of ROC analysis, where a large number true negatives (absence of leg movements detected by manual and automatic scoring) gives a false impression of high performance[106]. In this context, signal detection theory indicates that the positive predictive value (PPV) and Cohen's Kappa offer more insight into true performance. Using these metrics, our detector outperformed the other algorithms (Table 5.3). Examination of the data did not reveal a failure of the other detectors' methodologies, which were sound, but rather their lack of robustness to noisy EMG signal quality. Indeed, these detectors were typically evaluated and developed using high quality EMG data, with limited noise or artifact, from patients screened for sleep-disordered breathing (SDB).

We next examined outlier studies where PLMI scored data differed significantly from automatically detected scores on a correlation plot. Returning to individual studies, the following sources of problems were revealed: (1) ECG contamination in leg movement channels creating false detections; (2) variable baseline signal noise undermining the 8 μV rule and leading to over and under detections; (3) delayed LMs following sleep-disordered breathing events that were not omitted by the restrictive 0.5 sec window of the AASM 2007 criteria, while intermediate, respiratory unrelated LM were; and (4) fragmentary myoclonus.

Table 5.3: Preliminary detector performance compared to manually scored LM in the WSC, during sleep, according to WSC 1995 criteria.

	SE	SP	PPV	NPV	K(1,0)	K(0,0)	Cohen's Kappa	ACC	LM Count
Detector performance for all LM (n=1,073)									
Tauchmann	0.30	0.98	0.21	0.99	0.29	0.20	0.24	0.97	167,955
Wetter	0.86	0.95	0.20	1.00	0.85	0.19	0.31	0.95	503,211
Ferri	0.94	0.87	0.09	1.00	0.93	0.08	0.14	0.87	1,220,765
SNR+	0.78	1.00	0.73	1.00	0.78	0.72	0.75	0.99	124,351
ANC, SNR+	0.75	1.00	0.74	1.00	0.75	0.74	0.74	0.99	116,335
ANC, VAT	0.53	1.00	0.80	0.99	0.53	0.80	0.64	0.99	76,585
ANC, VAT, SNR+	0.72	1.00	0.79	1.00	0.72	0.78	0.75	0.99	106,397
WSC manually scored	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	119,277
Detector performance per patient (n=1,073)									
Tauchmann	0.26	0.98	0.37	0.99	-	-	0.25	0.97	156.53
Wetter	0.85	0.95	0.30	1.00	0.82	0.29	0.39	0.95	468.98
Ferri	0.93	0.89	0.31	0.96	-	-	0.42	0.89	1,137.71
SNR+	0.75	1.00	0.69	1.00	0.74	0.69	0.69	0.99	115.89
SNR+, ANC	0.71	1.00	0.71	1.00	0.71	0.70	0.67	0.99	108.42
ANC, VAT	0.44	1.00	0.68	0.99	0.44	0.67	0.50	0.99	71.37
SNR+, ANC, VAT	0.65	1.00	0.69	1.00	0.64	0.69	0.64	0.99	99.16
WSC manually scored	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	111.16

ANC: adaptive noise cancelling of cardiac interference; SNR+: signal-to-noise-ratio enhancement; VAT: Variable amplitude thresholding; SE: sensitivity; SP: specificity; PPV: Positive predictive value; NPV: negative predictive value; ACC: accuracy; LM count: total number of leg movements detected.

5.2.4 FINAL DETECTION ALGORITHM

As outlined in Figure 5.1, the final PLM detector presented in this manuscript incorporates salient features of previously published detectors and adds several key innovations, particularly in regard to handling noise and artifact. Its design followed a number of iterations not described in sequence for the sake of simplicity. The final

algorithm applies an adaptive filter to the leg EMG to remove ECG interference. A two-pass noise floor calculation is made of the cleaned signal to establish dynamic upper, lower, and falloff amplitude detection thresholds for each sample point. An optional 2 tap summing filter may be applied to the adaptively cleaned signal in places where the noise measurement falls below $2 \mu V$; referred to as the *SNR+* option. The 0.15 sec root mean square (RMS) value (i.e. similar to an integration time constant) of the adaptively cleaned signal is compared to these amplitude thresholds and a candidate LM is registered from the point the RMS first exceeds the upper threshold until the last point it is above the falloff amplitude before subsequently falling below the lower threshold for 0.05 sec (adjusting Ferri's dual threshold approach). Candidate classifications within 0.1 sec are next bridged (similar to Tauchmann), and candidate LMs subsequently less than 0.75 sec, rejected (i.e. 0.5 sec AASM criteria plus 0.25 sec to account for signal stretching caused by the filters). Remaining LMs within 2.0 sec are then merged a final time as done by Wetter [68], and candidate LMs greater than 10.0 sec are rejected. Merging within 2.0 sec is helpful in grouping alternating LMs recorded to a single channel.
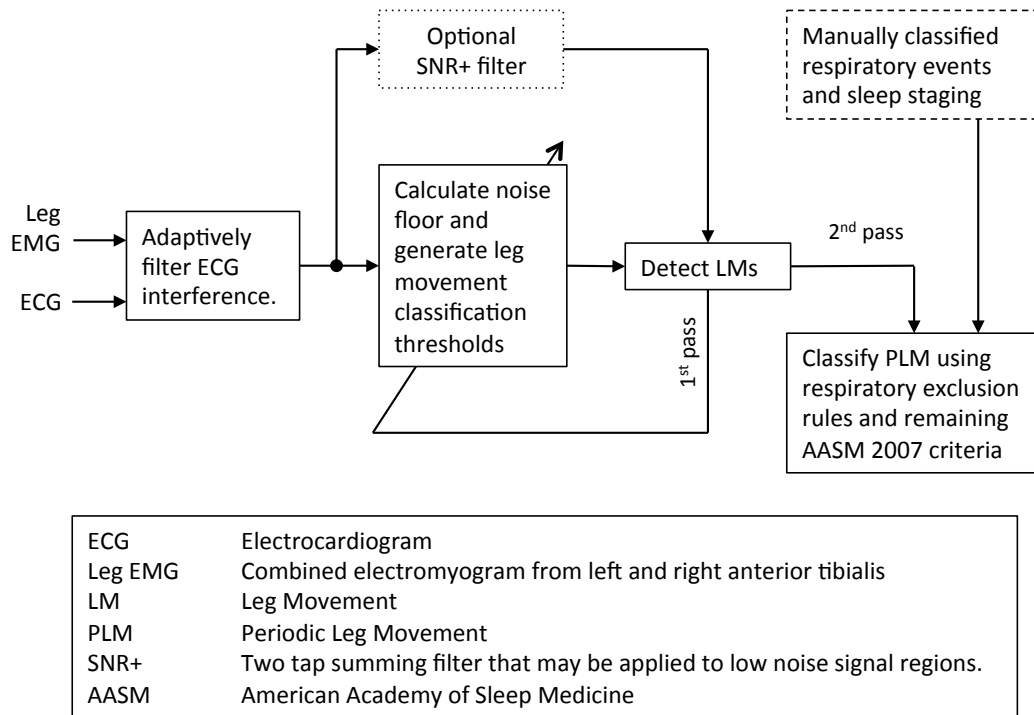
Figure 5.1: Flow chart of ten-step PLM detection algorithm. First pass LM detections are used to update the noise floor and generate the three amplitude thresholds a second time. The second pass LM detections are subsequently classified or rejected using AASM 2007 PLM scoring criteria, with our modification for respiratory exclusion.

The area under the curve (AUC), calculated by trapezoidal integration, is taken from cleaned data over the range of points corresponding to each candidate LM. If the candidate LM's AUC is greater than half the upper amplitude threshold it is retained as a LM, otherwise it is rejected. A final respiratory exclusion window is applied such that detected LMs that fall between 5.0 sec before until 0.5 sec following a manually scored respiratory event or from 0.5 sec before until 5.0 sec following its offset, are removed while any LMs during respiratory events which do not fall in these boundary exclusions are not. Remaining LMs are automatically reviewed by the final detection algorithm to determine whether they meet criteria for a string of PLMs according to AASM 2007's inter movement interval and frequency requirements (i.e. four or more consecutive LM

with inter-movement interval between 5 and 90 sec). Explanation of rationale for these design decisions is provided in the following sections.

5.2.5 ADAPTIVE NOISE CANCELLING TO REMOVE ECG ARTIFACTS FROM LEG EMG

Upon inspection, we found that interfering cardiac signal activity contaminated some studies and resulted in a large number of false positive detections of LMs. A low pass filter was initially used to remove this interference, however true muscle activity was indiscriminately attenuated which led to increased false negative detections. An adaptive recursive least squares filter as described by He et al for filtering EOG artifact from EEG signals[76] was deployed with great success.

To address possible cases where the ECG signal may be missing (e.g. some ambulatory configurations), we created a time-advanced copy of the LAT/RAT channel as the desired reference channel response as suggested by Widrow *et al*[107]. A small filter size and time advance allows the filter to adapt to the highly periodic cardiac interference and not LM. We found a five weight adaptive filter with a five-sample, duplicate source-channel, reference signal advance to effectively handle cardiac interference in a solitary input channel configuration. Although this second method was effective, it was less optimal and was not used in this particular dataset where ECG is always available.

Description of the filter and examples of its application to removing cardiac interference in the LAT/RAT channel without attenuating true leg muscle activity, as observed with conventional finite-impulse response filtering, are provided in Figure 5.2 and Figure 5.3.
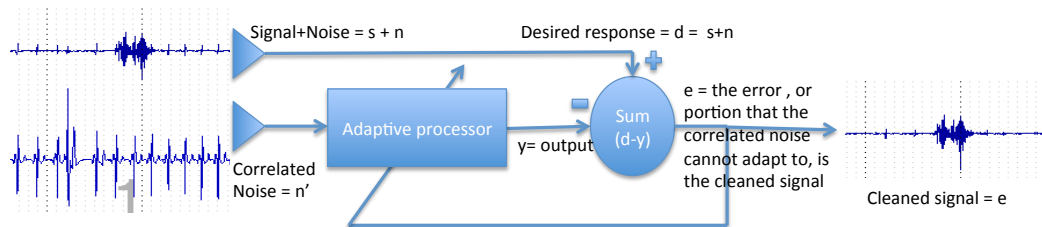
Adaptive noise cancellation using the Least Mean Square offers significant detection improvement over conventional preprocessing approaches
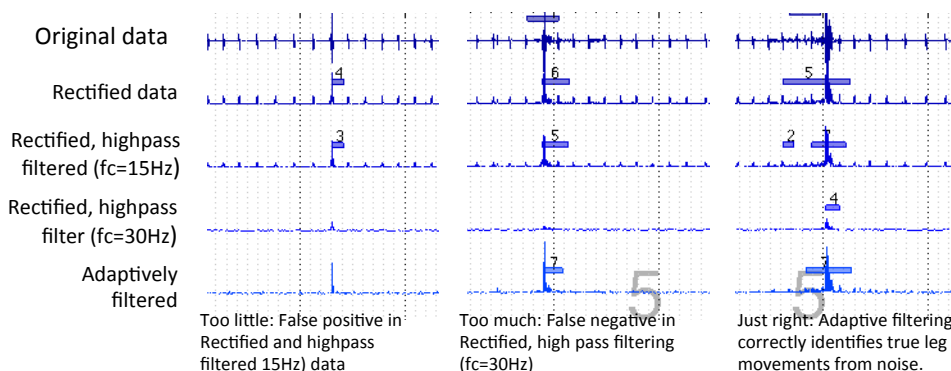
Figure 5.2: Adaptive filtering of cardiac interference compared to conventional finite-impulse-response filtering. Cardiac interference is adaptively cancelled from the leg EMG channel using a recursive least squares adaptive filter which continually updates its weights to minimize the least mean square difference between its output from filtering the correlated noise (i.e. the ECG channel) and the desired response (i.e. the leg EMG channel). The filter's output is tuned to the correlated noise, which when subtracted from the desired response of signal and noise leaves the clean signal behind as the error, which is the signal less the correlated noise (i.e. the leg EMG without cardiac interference). The lower section shows three examples of leg EMG activity. The original data is shown at the top of each example. Under the original data is the rectified version followed by high pass filtered outputs with cutoff frequencies of 15 Hz, then 30Hz, and finally the adaptively filtered data is shown at the bottom for each example. Blue horizontal bars show reveal detections from the classifier using the data shown. The adaptively filtered data provides the best results in each case.

A. Thirty-second epoch, ECG interference in the leg EMG, and no leg movement activity present.
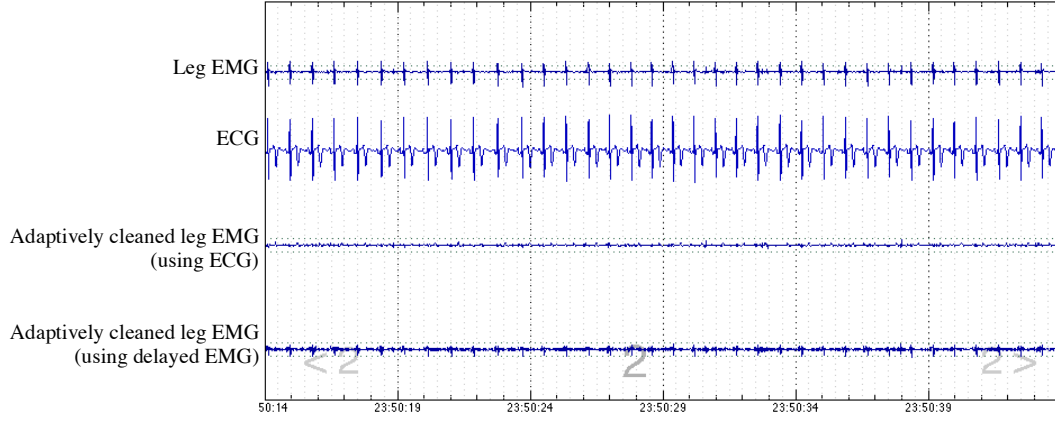
B. Thirty-second epoch, ECG interference in the leg EMG, leg movements present and noise in ECG channel. The self-reference adaptive configuration is not disrupted by the noise present in the ECG channel here.
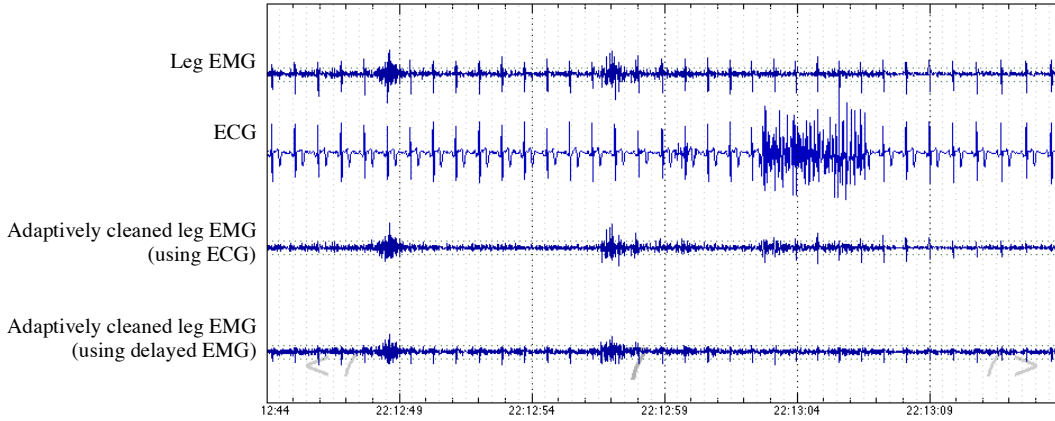
Figure 5.3: Examples of adaptive filtering to remove cardiac interference found in the leg EMG channel (continued on next page). Panels A, B, C, D show the leg EMG channel on top, followed by the ECG channel second, the adaptive noise cancelled EMG channel using the ECG as input, and finally the adaptive noise cancelled EMG channel using a time shifted copy of the EMG as input (i.e. single channel configuration). The time shifted, self-reference adaptive filter configuration is less effective in cleaning the data than the ECG configuration, but still better than the original data. It is not disrupted by noise in the ECG channel as seen in B. For methodological details, see text.

C. Fifteen-second epoch.

Leg EMG

ECG

Adaptively
cleaned leg EMG
(using ECG)

Adaptively
cleaned leg EMG
(using delayed

02:02:19   02:02:24   02:02

D. Ten-second epoch.

Leg EMG

ECG

Adaptively
cleaned leg EMG
(using ECG)

Adaptively cleaned
leg EMG (using
delayed EMG)

02:09:34   02:09:39

E. Thirty second epoch. Considerable cardiac noise in the leg EMG channel is effectively removed using adaptive filtering and allows leg movement detetion near the 20 s mark.

Leg EMG

Adaptively
cleaned leg EMG

:15:35   03:15:40   03:15:45   03:15:50   03:15:55   03:16:00

(Figure 5.3 continued)

5.2.6 Variable amplitude thresholding to address variable or excessive baseline noise

Patient movement or high electrode impedance levels deter automated detectors that rely on a statically calculated baseline noise levels. In these cases, the mean baseline signal may be high, for example 6 μV, and a vey small increase may be sufficient to trigger a false detection of a leg movement. The World Association of Sleep Medicine (WASM) standards for recording and scoring PLMs during sleep (PLMS) and wakefulness

(PLMW) recommend increasing the baseline in areas where EMG activity is raised due to prolonged muscle flexion – a related problem. The new baseline is taken from the average EMG amplitude during this time [108]. This recommendation is not well-suited for automated processing and does not help when calibration is not done (e.g. ambulatory studies) or is missing.

Variable amplitude thresholding is used to address this issue. We created a mathematical model of the noise floor from two passes of the LAT/RAT EMG channel in order to establish optimal thresholds for PLM detection. A first-pass measure of noise is taken as the 20 sec moving average filter of the data. An upper and lower threshold are calculated from this noise floor, and sections of data deemed LM on the first pass are then set to a lower value and a second noise floor measure is made using the adjusted data and additional LMs classified using the upper and lower thresholds calculated from the second pass noise floor measure.

Typically, the noise floor is measured during the preflexion calibration step in PSG studies, and the LM detection thresholds are adjusted accordingly (i.e. 8 μV above baseline followed by a drop below the 2 μV baseline offset). Unfortunately, establishing a constant offset of 8 μV for the entire study is insufficient in separating true signal from noise as the noise floor increases. As the baseline noise floor begins to increase, sporadic alterations in amplitude become common along with false detections from the frequent transitions between the upper and lower thresholds. Further noise floor elevation effectively shuts the detector off by raising the signal to the point where it either no longer falls below the lower threshold or does but will have exceeded the maximum LM duration (i.e. 10 sec). Therefore, a different model is needed to handle dynamic changes in the noise floor.

We developed the following system to determine the noise floor and corresponding upper and lower threshold for optimal LM detection. The noise floor, η, is defined for sample n as

$$x(n) = |\text{EMG amplitude at sample } n|$$

$$\eta(n) = \frac{1}{N} \sum_{i=-N/2}^{N/2} x(n-i)$$

where x(n) is the EMG voltage measured at sample n. Letting U be the original upper threshold scalar value (i.e. 8 μV) and L the lower threshold scalar (i.e. 2 μV), the PLM detection thresholds for each time sample are defined as

$$\alpha(n) = \left\{ \begin{array}{ll} \eta(n) \cdot \log(\eta(n)+1) + U & , \; \eta(n) \leq 50 \\ \infty & , \; \eta(n) > 50 \end{array} \right\}$$

$$\beta(n) = \frac{L}{U} \cdot \alpha(n)$$

$$\psi(n) = \frac{\alpha(n) + \beta(n)}{2}$$

where $\alpha(n)$, $\beta(n)$, and $\psi(n)$ correspond to the upper, lower, and offset thresholds at sample index n.

Figure 5.4 shows the three amplitude threshold functions dependent on the noise floor. The upper threshold, $\alpha(n)$, increases as a function of the noise floor $\eta(n)$, which is calculated for each sample point n of the leg EMG using a 20 s moving average filter. The lower threshold, $\beta(n)$, is scaled using the ratio of the AASM 2007 Scoring Manuals upper and lower constant threshold values (i.e. 2 μV over 8 μV or 0.25). The cutoff threshold, $\psi(n)$, is the average of the upper and lower thresholds and used to determine offset of candidate LMs. The detector shuts off whenever the noise floor exceeds 50 μV (not shown). Candidate LM detection onset occurs when the cleaned EMG signal first exceeds the upper threshold and terminates at the last point the cleaned EMG signal falls below the cutoff threshold prior to subsequently falling below the lower threshold for 0.05 s.
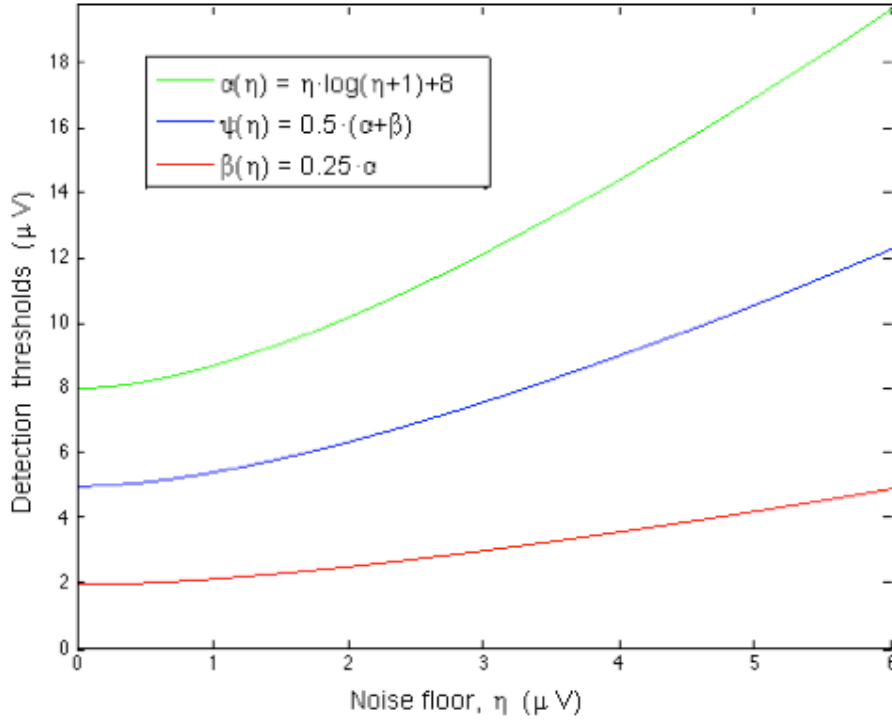
Figure 5.4: Upper, middle, and lower detection thresholds as a function of the noise floor. The noise floor is calculated in two passes using a 20-s moving average of the adaptively noise cancelled leg EMG channel on the first pass. Sections of the input signal classified with candidate LM using the first pass thresholds, are adjusted for the second and final noise floor and threshold calculations.

The EMG RMS, $y(n)$ is defined using a 0.15 sec window with sampling rate f_s as

$$M \;=\; 0.15 \cdot f_s$$

$$y(n) = \sqrt{\frac{1}{M} \sum_{i=-M/2}^{M/2} x(n-i)^2}$$

This is compared to the three thresholds to determine candidate LM onsets and offsets. LM onset is defined when $y(n)$ first exceeds $\alpha(n)$. LM offset is defined as the last point $y(n)$ exceeds $\psi(n)$ after onset and before subsequently falling below $\beta(n_1 + k)$ for 0.05 sec. Detections separated by 0.1 sec or less are merged.

148

A problem with a single pass noise floor calculation is that the contribution from true LMs are not taken into account. True LMs averaged into the noise floor may elevate the detection thresholds to the extent that nearby LMs are missed. Using a second pass, portions of the initial input signal $x(n)$ containing first pass LM are replaced with half the lower threshold for that same region (i.e. $\beta(n)/2$). The noise floor and corresponding thresholds are then calculated again before the second, final detection pass is made. Additional passes can be defined in the same manner, however doing so may overly reduce the noise floor and lead to false positive detections. Similarly, setting the first pass signal sections with detected LMs to zero, instead of $\beta(n)/2$, also overly reduces the noise floor and results in false positive detections.
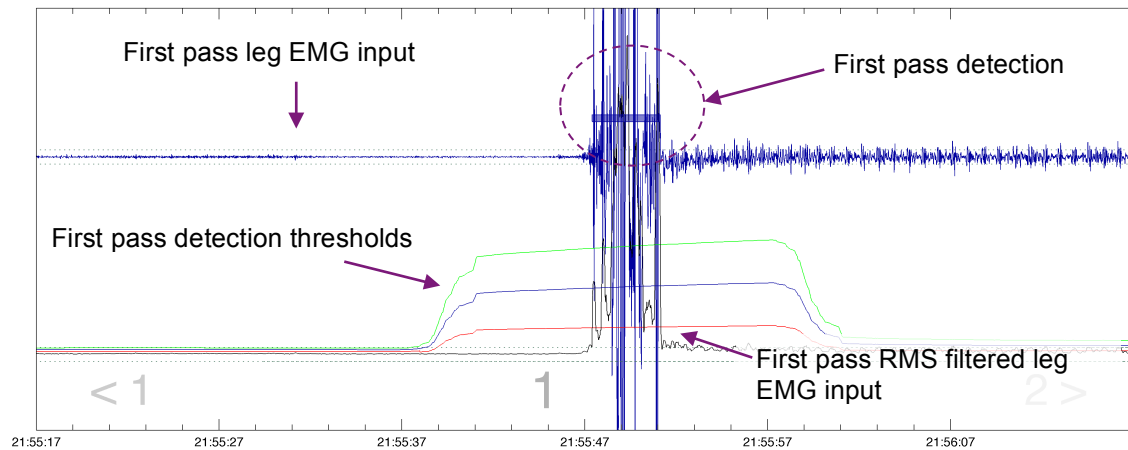
Figure 5.5: First pass thresholds. A large increase in noise floor is seen in the leg EMG channel (top) halfway through the one-minute epoch shown. The three amplitude detection thresholds are adjusted accordingly (bottom).

Figure 5.5 illustrates the first pass detection threshold adjustments in a one-minute epoch. High EMG activity is seen in the middle portion of the leg EMG input signal (top), which is likely due to movement loosening the electrode and increasing impedance. The three amplitude detection thresholds calculated from the noise floor are: upper (yellow), lower (red), and cutoff (blue). An LM detection, highlighted by the purple rectangle, is made in the first pass where the root mean square (RMS) filtered signal passes the upper threshold

149

until that last moment it falls below the cutoff threshold before subsequently falling below the lower threshold for the set duration.

The second pass threshold adjustments for this example are shown in Figure 5.6. The portion of signal identified as candidate LM from the first pass (Figure 5.5) is adjusted according to the previous threshold levels (i.e. half of the lower, first pass detection amplitude threshold) to produce the second pass leg EMG input. The remaining spike of activity, which passes the detection thresholds is too short, and only the first pass detection is retained. Had the spike in activity here been long enough to meet candidate LM detection, it would have fallen within the 2.0 sec merge window and been combined with the first pass detection to make a single, slightly longer candidate LM.
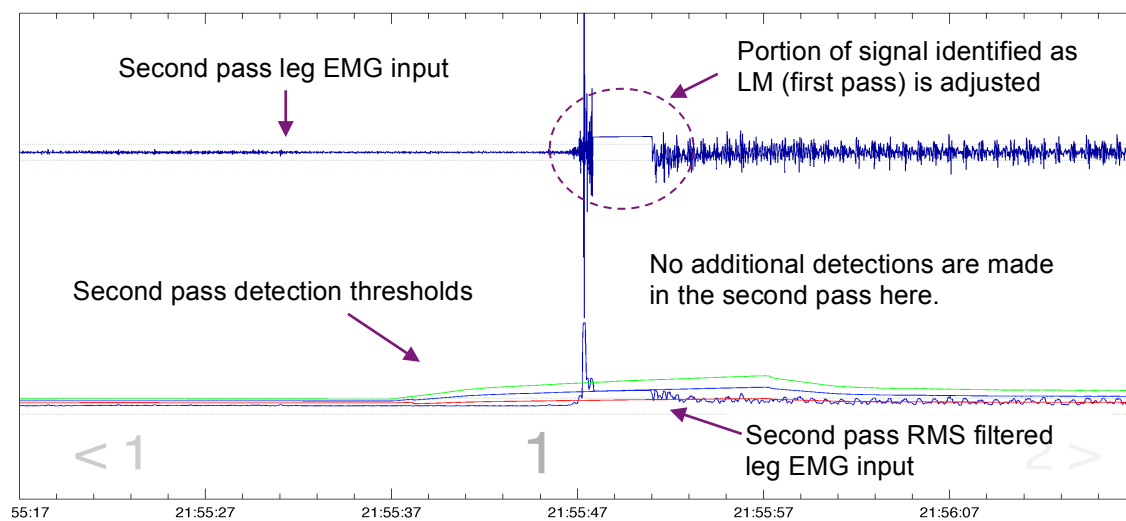
Figure 5.6: Second pass threshold adjustment. The portion of signal identified as candidate LM in the first pass is adjusted according to the previous threshold levels and the leg EMG input signal is adjusted for the second pass.

Alternatively, EMG signal strength may be attenuated at times from weak recording or physiological generation and can be corrected for using a signal-to-noise ratio enhancement (i.e. *SNR+*) option. We found that in some cases, very low noise at baseline allows the clear visualization of LMs that are not captured by our detector

because they are smaller than 8 μV. This was a particularly common occurrence in the WSC Gold Standard where the technician annotated PLM based more on visible periodicity seen on a one or two minute time scale than quantitative measures of change from baseline.

The SNR+ option increases signal strength when the noise floor falls below 2 μV by application of a two-tap finite impulse response (FIR) summing filter which increases signal strength from 0 to 33 Hz while attenuating its strength for frequencies above 33 Hz (i.e. high frequency activity commonly associated with noise). Figure 5.7 describes the filter in detail and motivating examples are given in Figure 5.8, Figure 5.9, and Figure 5.8 outlines the successive signal preparation steps using a one minute epoch of clean leg EMG signal. Figure 5.9 and Figure 5.10 further outline these steps using two minutes of noising leg EMG signal (one minute for each figure). In each case, the SNR+ option sufficiently separates the LM activity from the background noise to enable detection.
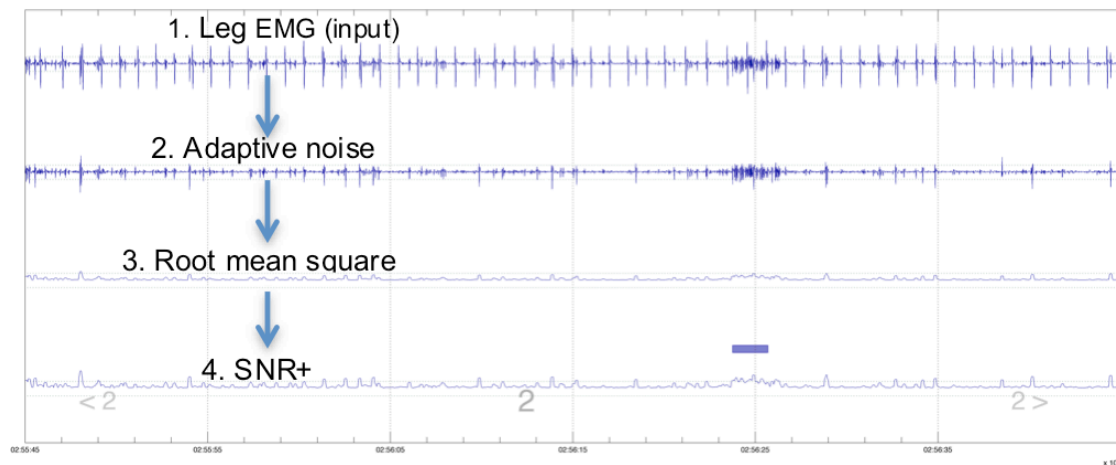
Figure 5.10. The rationale for providing the detector both with and without the SNR+ option is because the SNR+ is not in strict accordance to the 8 μV criteria of the AASM, although as it will be outlined below, it correlated better with technician trained scoring of PLMs.
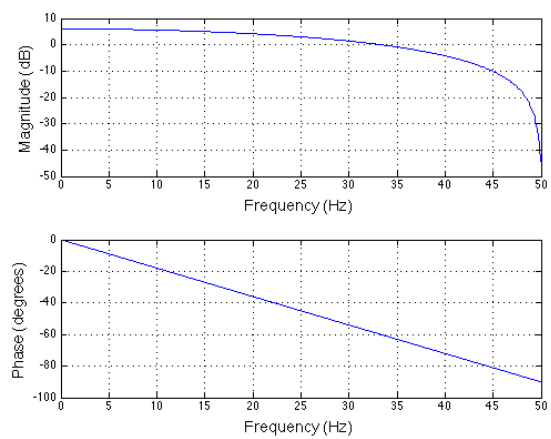
A.
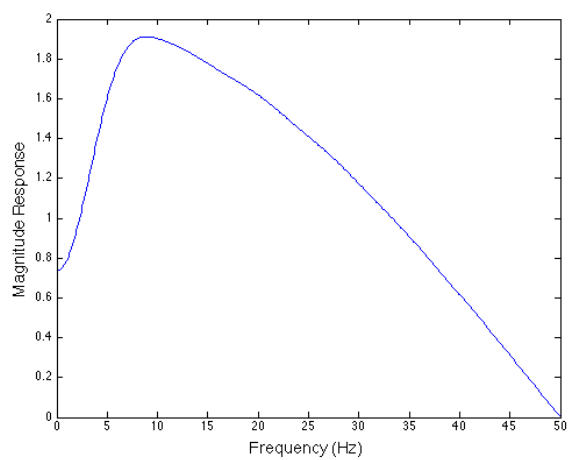
B.

Figure 5.7: Frequency response for SNR+ 2-tap summing filter (panel A) and normalized magnitude response in cascade with a 10 Hz high pass digital filter (panel B).

Figure 5.7 shows the 2-tap FIR summing filter used to raise the signal to noise ratio when the noise floor is small (i.e. less than 2 μV). The filter's frequency response, using a 100 Hz sampling rate, is shown in panel A. The filter has linear phase delay (A bottom) and increases signal amplitude at frequencies below 33 Hz while further attenuating signal strength above 33 Hz - high frequency activity commonly attributed to noise. Detection algorithms that apply low or high pass filters using at 16 Hz cutoffs remove relevant portions of the surrounding spectral activity. Panel B shows the normalized magnitude response of the 2-tap filter in cascade with a simulated 10 Hz high pass hardware filter that is applied before digitization.
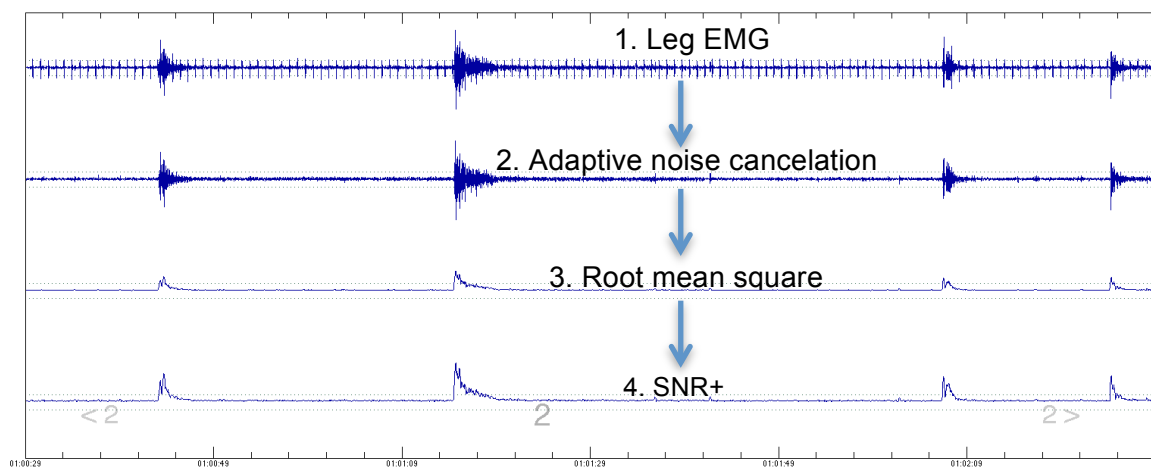
Figure 5.8: Clean signal example. The SNR+ option increases the signal to noise ratio so that the periodic leg movements can be properly identified.
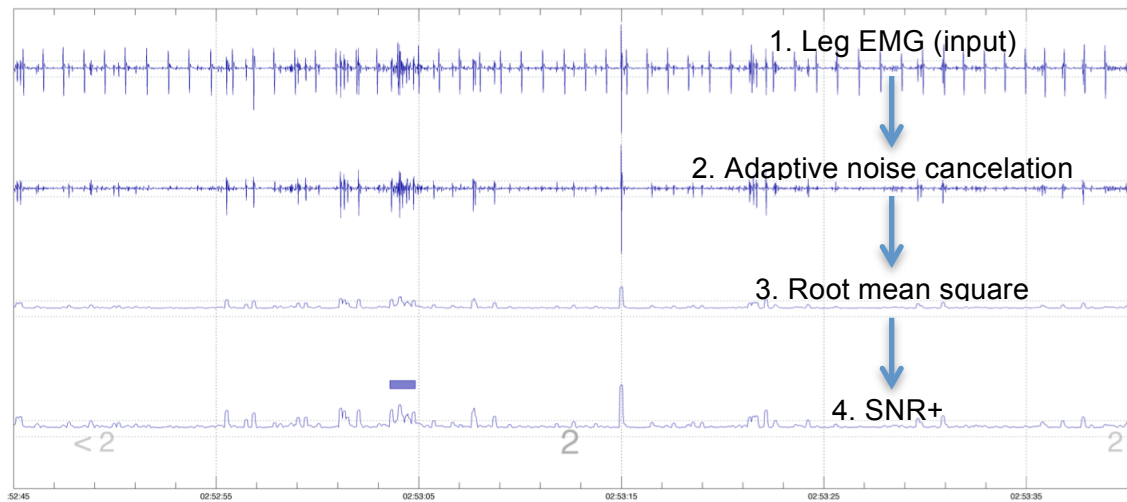
Figure 5.9: One minute, noisy signal example A.

Figure 5.8 outlines the successive signal preparation steps using a one minute epoch of clean leg EMG signal. Figure 5.9 and Figure 5.10 further outline these steps using two minutes of noising leg EMG signal (one minute for each figure). In each case, the SNR+ option sufficiently separates the LM activity from the background noise to enable detection.
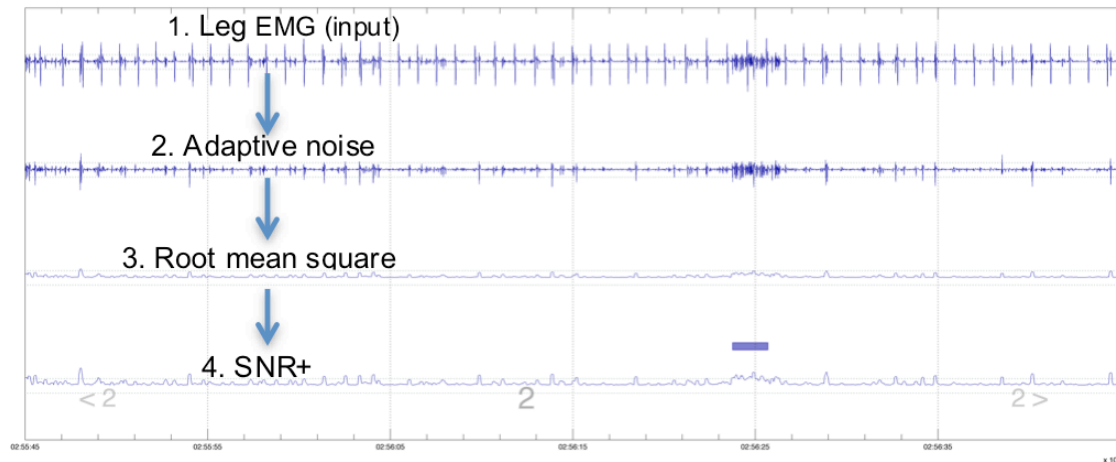
Figure 5.10: One minute, noisy signal example B.

In many cases, a low noise floor, defined as 2 μV or less, the leg EMG signal is attenuated and observed LM do not meet AASM amplitude criteria. In the study it was found that many LM marked for PLM were clearly visible when viewed on the one or two minute interval time scale commonly used by technicians. However, on close examination, these LM do not in fact meet the AASM 2007's amplitude criteria. Three different time scales of stage 2 sleep are shown below: 30 seconds (Figure 5.11), two minutes (Figure 5.12), and five minutes (Figure 5.13). The raw input leg EMG channel is shown along the bottom of Figure 5.11, Figure 5.12, and Figure 5.13 with the adaptively filtered versions passed as input to the detection configurations directly above. Detections made using the SNR+ option are shown as orange boxes along the top channel, while detections made without the SNR+ option are shown as magenta colored boxes above the middle channel.
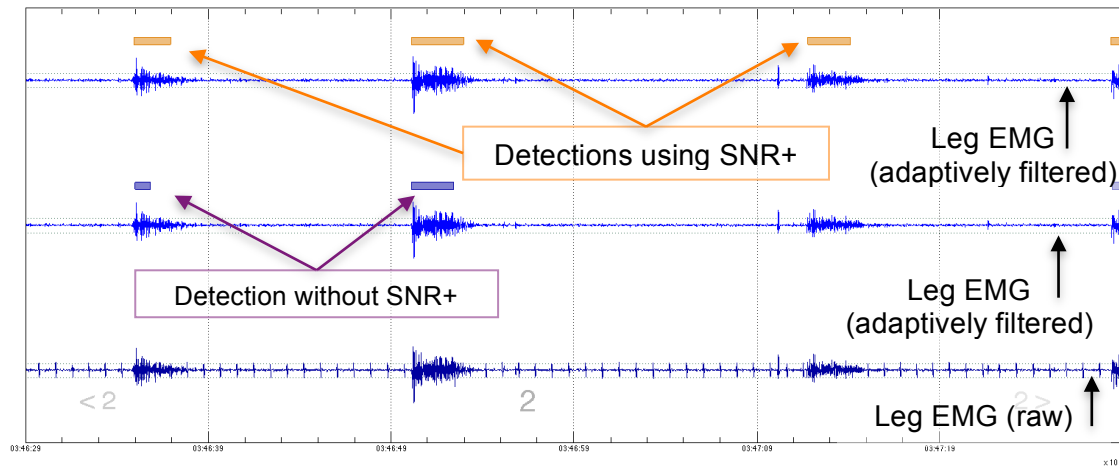
Figure 5.11: Thirty second epoch. The third LM is missed by not using the SNR+ option (middle channel).
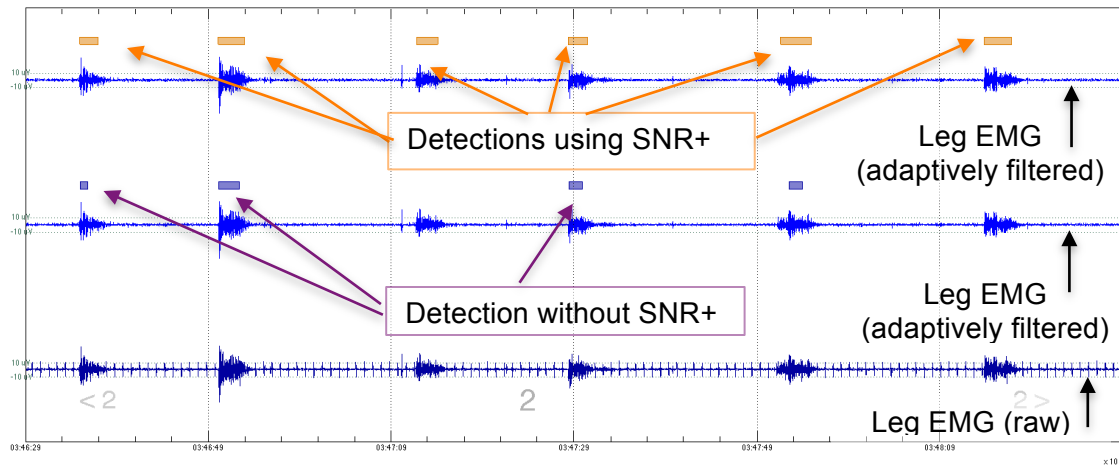
Figure 5.12: Two minute epoch. The third and sixth LM are missed (middle channel) by not using using the SNR+ option.
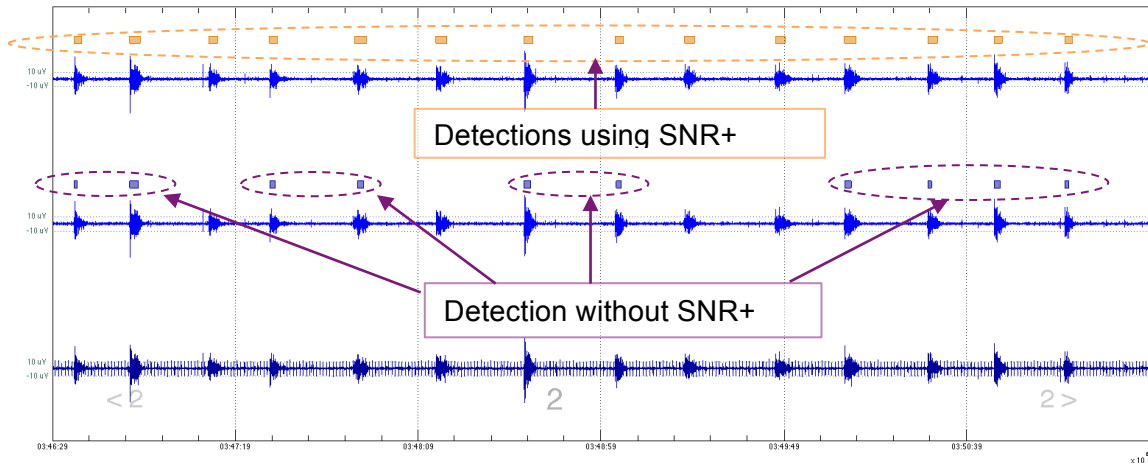
Figure 5.13: Five minute epoch. LMs are consistently detected using the SNR+ option (top, orange), though four small LM are missed occasionally missed here without it (middle, magenta).

5.2.7 FRAGMENTARY OR TRANSIENT MYOCLONUS AND SPURIOUS NOISE

Fragmentary myoclonus, brief (<200 msec) spikes of EMG activity[102], and other spurious noise are inadvertently *smeared* by the moving average filter applied to smooth the data, and produce false LM. Smearing effects are minimized by the short 0.15 sec RMS filter in combination with our 0.05 sec fall time and merging criteria.

The 0.05 sec fall time used to identify initial LM offset allows the detector to isolate short spikes in EMG activity and quickly *reset* to make new detections. Longer fall times leave the detector *on*, prolonging the opportunity to find movement, and making it susceptible to lingering, fragmentary activity. For example, a slowly declining LM followed by spurious noise becomes artificially long; possibly to the point of rejection. The subsequently applied 0.1 sec merge-window handles cases where the short fall time can finely split a true LM in parts, and falls in line with Tauchmann's approach [67].

Despite these tactics, fragmentary myoclonus or other sharp, short duration spikes in leg EMG activity may still raise the RMS signal to detection levels. As a final test against false detection, we calculate AUC of the original input signal's magnitude at each

157

candidate LM (i.e. onset to offset) and compare it to the upper detection threshold, $\alpha(n)$. LMs are relatively stable. They do not quickly rise above and then fall below the detection thresholds, but remain near or above it as characterized by the movement. The raw signal's AUC is used as measure of this stability in comparison to the detection threshold across the detected region. Candidate LMs with AUC less than half the upper noise threshold are rejected from PLM inclusion. Fragmentary myoclonus can be identified, instead of PLMs, by choosing to instead keep candidate LMs with AUC less than this value.

5.2.1 RESPIRATORY EVENTS AND ARTIFACT

Respiratory events, routinely annotated in nocturnal PSGs, include hypopneas (with 4% desaturation) and apneas (central, obstructive or mixed). The AASM 2007 Scoring Manual places a window from 0.5 sec prior *until* 0.5 sec after respiratory events in which LM activity is excluded from PLM criteria. Using this rule, we found PLMI to be greater in individuals AHI above 15, and lesser in those with AHI below 15. This seemed incongruent with the intention of excluding false LMs from PLM criteria due to respiratory events. Assuming PLMs and sleep disordered breathing are not strongly correlated at the epidemiological level, removing leg movements secondary to SDB should lead to similar PLMI values in subjects with and without sleep apnea. The AASM 0.5 sec exclusion rule does not meet its expressed intent of eliminating respiratory related LMs from PLM inclusion (Table 5.4).

We thought this requirement was overly conservative at the respiratory event boundaries where the 0.5 sec window is too small to catch the link between breathing event and leg movement. Therefore, we investigated the interaction by comparing manually scored respiratory events in the WSC studies (time of onset and offset of hypopneas with 4% desaturation and apneas) to leg EMG activity during *uninterrupted* sleep (i.e. activity surrounding respiratory events ending in or immediately followed by wake were not considered). Leg activity events surrounding respiratory events that were shorter than

Table 5.4: Several respiratory exclusion rules are applied to the PLM detector (adaptive filtering) to evaluate and compare their effect on median LM count, and mean and median PLMI in patients with AHI ≥ 15 and with AHI < 15 during sleep. The AASM 2007 exclusion window (indicated by *) reveals significantly higher median LM count and PLMI with increased AHI. Pathological PLMs should be similar in both groups, and onset and offset respiratory exclusion criteria are optimized to this end here. LM count shows the fluctuations caused by different windows prior to application of PLM criteria. Several suitable exclusion choices exist for removing activity associated with respiratory event that produce equal PLM detections in patients with and without SDB. We selected -5.0 to 0.5 around onset and -0.5 to 5.0 s around offset because of its good performance and relative symmetry.

Onset before (sec), after (sec)	Offset before (sec), after (sec)	Median LM count AHI<15, AHI≥15	Median PLMI AHI<15, AHI≥15	Mean PLMI AHI<15, AHI≥15
None	None	84,154	4.8, 12.6	14.1, 23.6
-0.5,~*	~, 0.5*	73, 107	3.43, 5.9	12.1, 12.7
-10, 0.5	-0.5, 5.0	70, 63	3.3, 2.6	12.0, 9.9
-10, 0.0	-0.5, 3.5	71, 66	3.3, 2.7	12.0, 10
-5.0, 0.0	-0.5, 3.5	76, 82	3.4, 3.8	12.5, 11.5
-7.5, 0.0	-0.5, 3.5	74, 72	3.3, 3.2	12.4, 10.7
-5.0, 0.5	**-0.5, 5.0**	**74, 76**	**3.34, 3.37**	**12.5, 11.9**
-5.0, 0.0	-0.5, 5.0	74, 77	3.3, 3.3	12.4, 11.2
-5.0, 0.0	0.0, 5.0	74, 79	3.3, 3.6	12.5, 11.3
-3.5, 0.0	-0.5, 3.5	76, 91	3.5, 4.4	12.6, 12.0
-6.0, 0.0	-0.5, 3.5	75, 77	3.4, 3.8	12.5, 11.1
-6.0, 0.5	-0.5, 3.5	75, 77	3.4, 3.8	12.4, 11.1
-6.0, 1.0	-1.0, 3.5	74, 74	3.4, 3.6	12.4, 10.9
-6.0, 1.0	-1.0, 4.0	73, 73	3.3, 3.3	12.4, 10.8
-6.0, 1.0	-1.0, 5.0	72, 72	3.3, 3.2	12.3, 10.6
-5.0, 1.0	-1.0, 5.0	73, 75	3.3, 3.2	12.4, 10.9
-5.0, 1.0	-1.0, 6.0	72, 72	3.3, 3.1	12.4, 10.7
-5.0, 1.0	-1.0, 7.5	72, 69	3.3, 2.8	12.3, 10.4
-5.0, 5.0	-5.0, 5.0	71, 70	3.3, 2.8	12.1, 9.9
-6.0, 5.0	-5.0, 4.0	71, 68	3.3, 2.9	12.1, 9.8
-6.0, 2.0	-2.0, 4.0	72, 72	3.3, 3.2	12.2, 10.5
-6.0, 2.0	-1.0, 5.0	79, 102	4.0, 4.7	13.0, 13.6
-6.0, 2.0	-0.5, 5.0	79, 102	4.0, 5.1	13.0, 13.8
-6.0, 2.0	-1.0, 7.5	77, 93	3.9, 4.0	12.9, 12.6
-6.0, 2.0	-2.0, 7.5	78, 92	3.9, 3.9	12.8, 12.4
-6.0, 1.0	-2.0, 7.5	79, 100	3.9, 5.8	13.1, 14.9
-7.5, 1.0	-2.0, 7.5	79, 100	3.9, 5.8	13.1, 14.9
-10.0, 1.0	-2.0, 7.5	77, 87	3.8, 4.2	12.8, 13.1

15 sec or were within 30 sec of another respiratory event were also rejected from analysis. Indeed, a leg movement preceding a respiratory event may be due to the nearby, previous respiratory event offset, and vice versa.
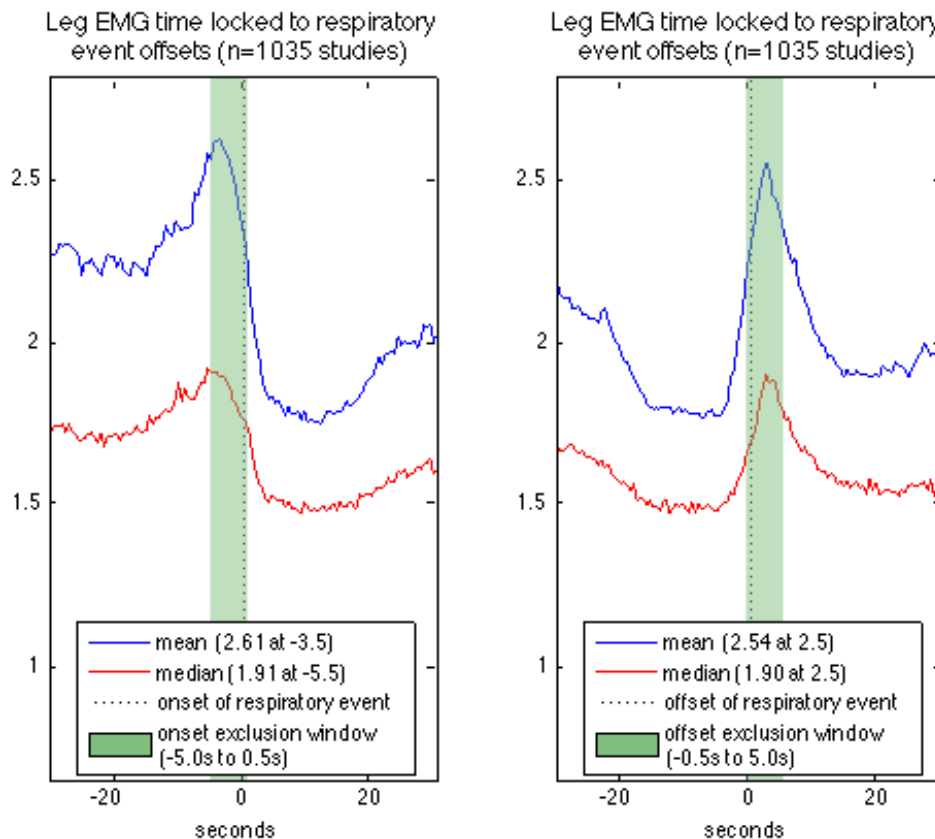
Figure 5.14: Leg EMG activity time locked to manually scored WSC respiratory events. Expected mean and median absolute leg EMG voltage is calculated in 0.5 s increments starting 30 s prior until 30 s after respiratory event onset (left figure) and each respiratory event offset (right figure). Respiratory events followed by wake, less than 15 s in duration, or which fall within 30 s of another event are removed. Leg EMG activity increases and then decreases prior to respiratory onset with a peak 4.5 s prior to onset, and later decreases before and increases after respiratory offset with a mean peak 2.5 s and median peak 3.5 s.

Figure 5.14 shows the expected mean and median leg EMG activity, calculated on 0.5 sec increments time-locked 30 sec prior until 30 sec following respiratory event onset and

offset as manually scored in the WSC. Leg EMG activity increases and then decreases prior to respiratory onset with a peak 4.5 sec prior to onset, and later decreases before and increases after respiratory offset with a mean peak 2.5 sec and median peak 3.5 sec. The effect of various respiratory exclusion windows on resulting LM and PLM metrics were evaluated using our initial detector with adaptive filtering and a new respiratory exclusion window was selected.

The window to exclude LMs due to respiratory events extends 5.0 sec prior to respiratory event onset until 0.5 sec after onset, and then again from 0.5 sec before to 5.0 sec after the respiratory event's offset. LMs occurring during the respiratory event that do not fall in the exclusion window may be considered for PLM, which is a departure from the AASM 2007 scoring guidelines.

Examination of respiratory events isolated by 60 sec reveals similar findings, as do examinations of time-locked EMG activity in patients with fewer than twenty respiratory events. Additional considerations and details of this analysis are provided in Figure 5.15 through Figure 5.21, and Table 5.4.

Figure 5.15 shows a preliminary analysis of leg EMG activity time locked to manually scored WSC respiratory events. Mean and median measures are taken of absolute leg EMG voltage at 0.5-second increments starting 30 s prior until 30 s after respiratory event onset (left panel) and each respiratory event offset (right panel). Leg EMG activity increases and then decreases prior to respiratory onset with a peak seen 5.5 s prior to onset. Leg EMG activity decreases and then increases at respiratory offset with a peak at 2.5 s (mean) or 3.5 s (median) following the exact point scored as offset. The pre-onset bump in EMG activity could be due to time-locking of respiratory events with short inter-event intervals (e.g. a five second lapse between the end of one respiratory event and the start of the next) or short duration (e.g. less than 15 s). Because of this uncertainty, the onset-to-onset and offset-to-onset interval distributions for successive respiratory events in the WSC were investigated.
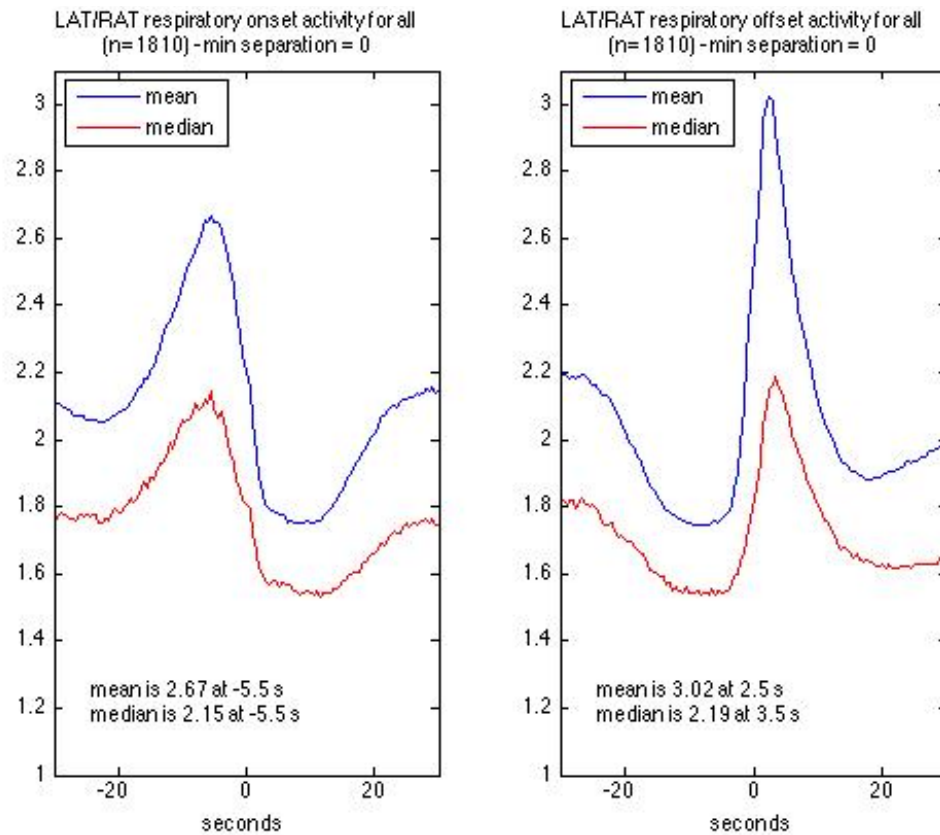
Figure 5.15: Preliminary analysis of leg EMG activity time-locked to manually scored WSC respiratory events. Mean and median measures are taken of absolute leg EMG voltage at 0.5-second increments starting 30 s prior until 30 s after respiratory event onset (left panel) and each respiratory event offset (right panel)
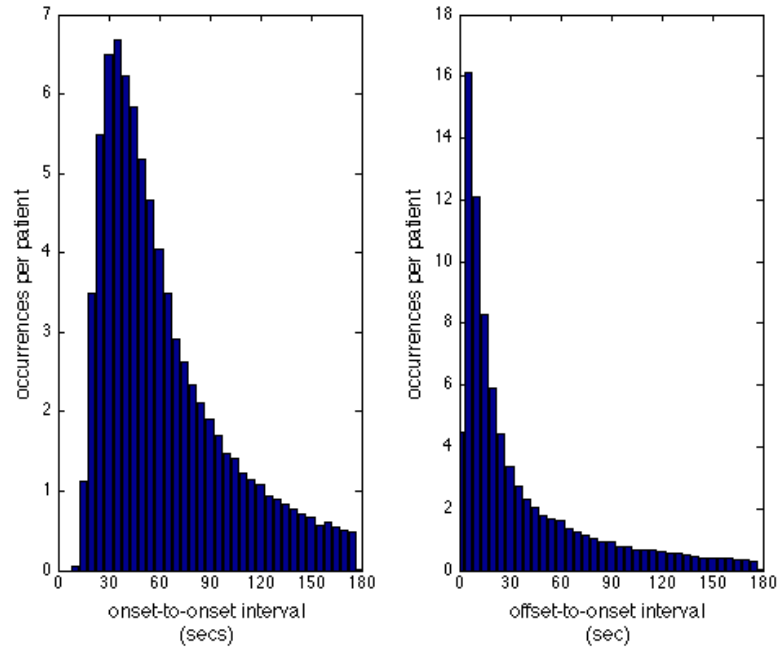
Figure 5.16: Interval distributions for all manually scored WSC respiratory events shown in 5 s increments. (a) Left plot is of onset to onset interval (b) Right plot is offset-to-onset interval (i.e. from the end of one respiratory event until the beginning of the next).

Figure 5.16 shows the interval distributions for all manually scored WSC respiratory events in 5 s increments. The onset-to-onset interval peak occurs at 35 s, while the offset-to-onset interval peaks around 5-seconds. These distributions do not clarify whether the rise in EMG activity seen at time-locked respiratory events is related to the offset of a respiratory event or the occurrence of the following respiratory event (which is often 5.0-10 s later in SDB). We removed this uncertainty by only considering respiratory events scored in excess of 15 sec and that are at least 30 sec from neighboring respiratory events. The onset-to-onset and offset-to-onset interval distributions for this configuration are shown in Figure 5.17. We used this setup to obtain the results shown in Figure 5.14, which in turn formed the basis of our newly proposed respiratory related PLM exclusion criteria.
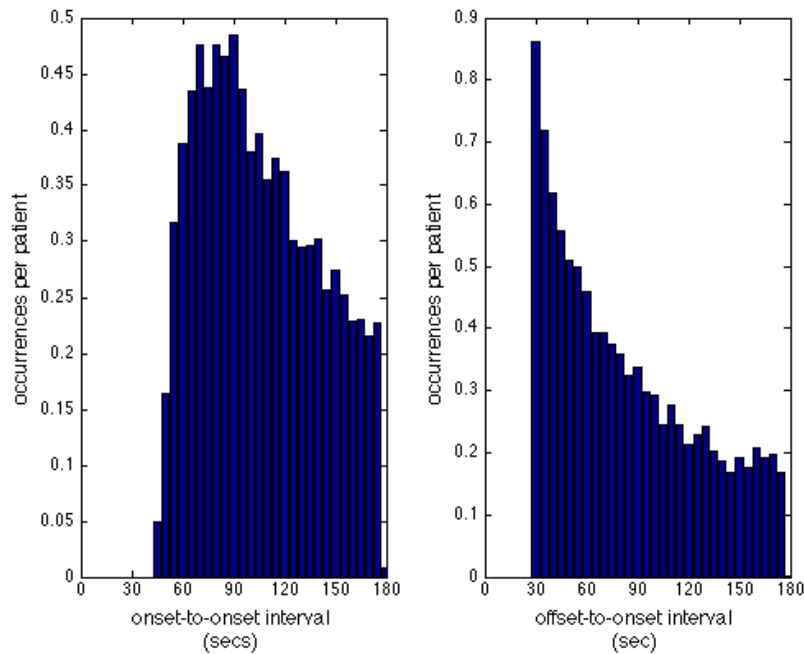
Figure 5.17: Interval distributions (5 s bins) for manually scored respiratory events greater than 15 s duration and separated from other by more than 30 s from other respiratory events are used for time-locking leg EMG activity. The left histogram shows the onset-to-onset interval distribution of these respiratory events, while the right histogram shows their onset-to-offset interval distribution.

Another concern with this analysis is that the scoring technicians may not report all respiratory events. In the case where a patient has an excessive number of apneas and or hypopneas the technician may become tired of scoring all of the events. I constrained my analysis to only those with twenty or fewer respiratory events (Figure 5.18), ten or fewer (Figure 5.19), five or fewer (Figure 5.20) and finally studies with only a single score respiratory event (Figure 5.21). The observed effect became less smooth and more dominant at offset, but is still visible around onset even in the case of a single respiratory event scored.

164
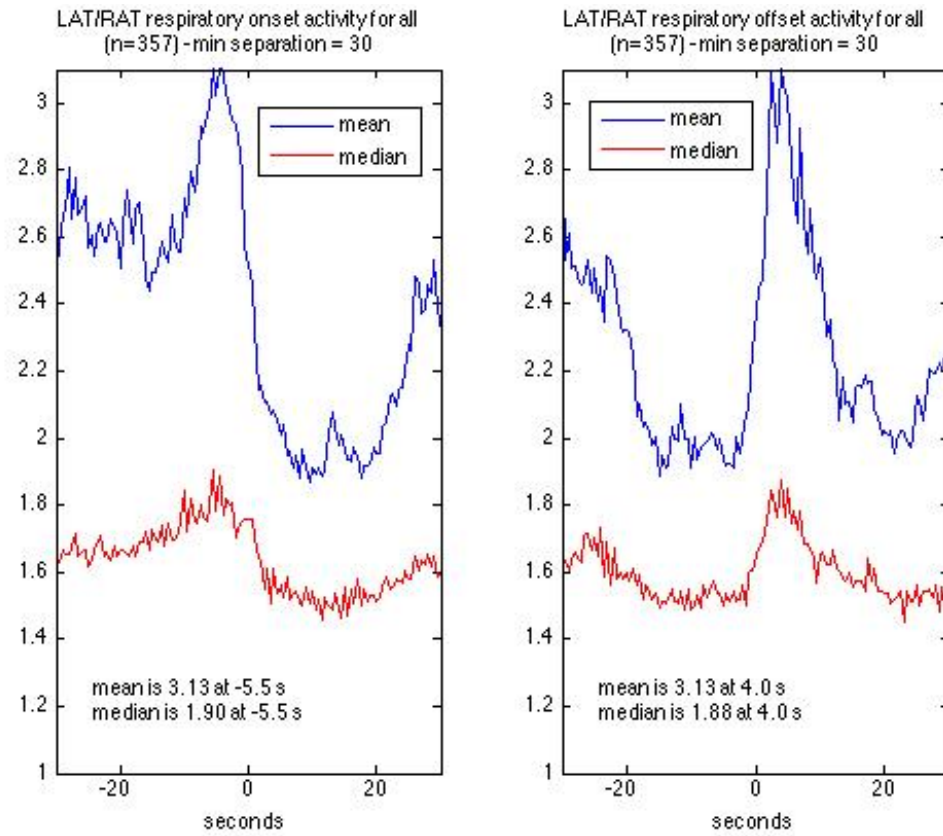
Figure 5.18: 30-second isolated, 15-s minimum duration respiratory time locked EMG activity in patients with between 1 and 20 scored respiratory events (after exclusion rules are applied).
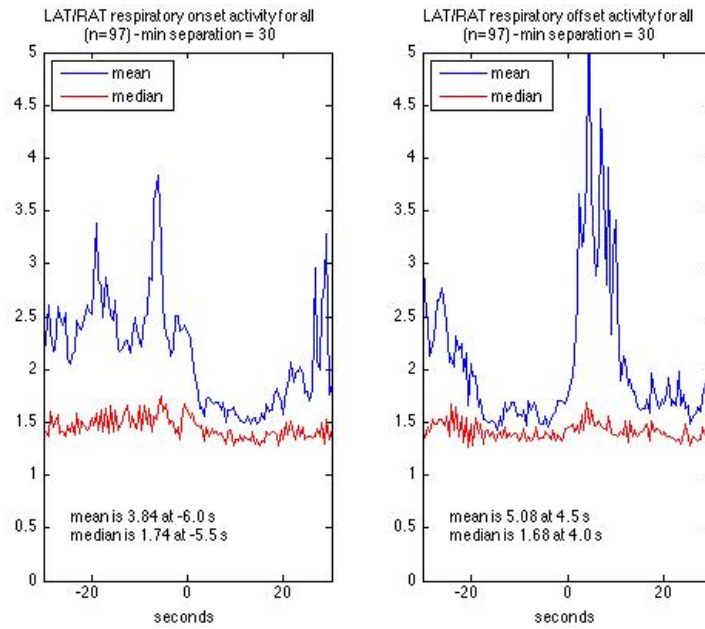
Figure 5.19: Per patient EMG average in 1 to 10 scored respiratory events.
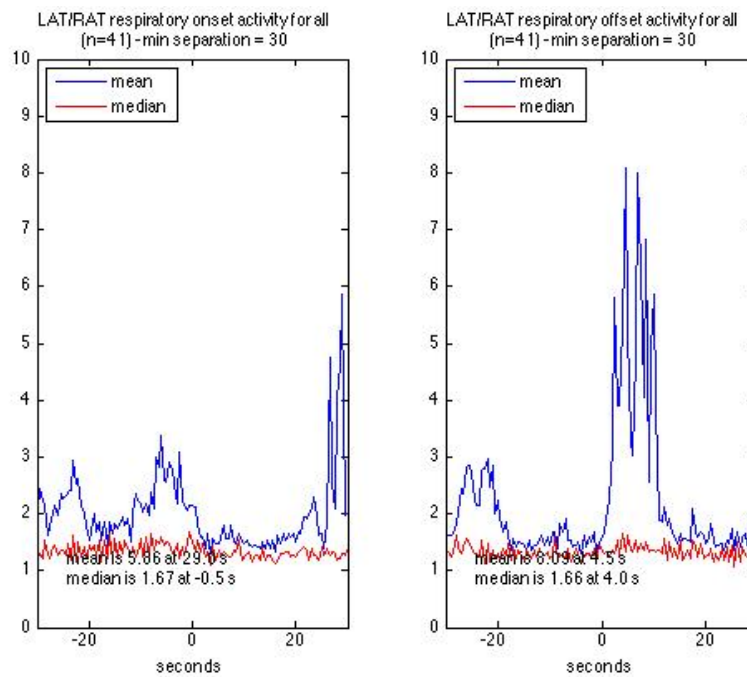
Figure 5.20: Per patient EMG average in studies with 1 to 5 respiratory events scored.
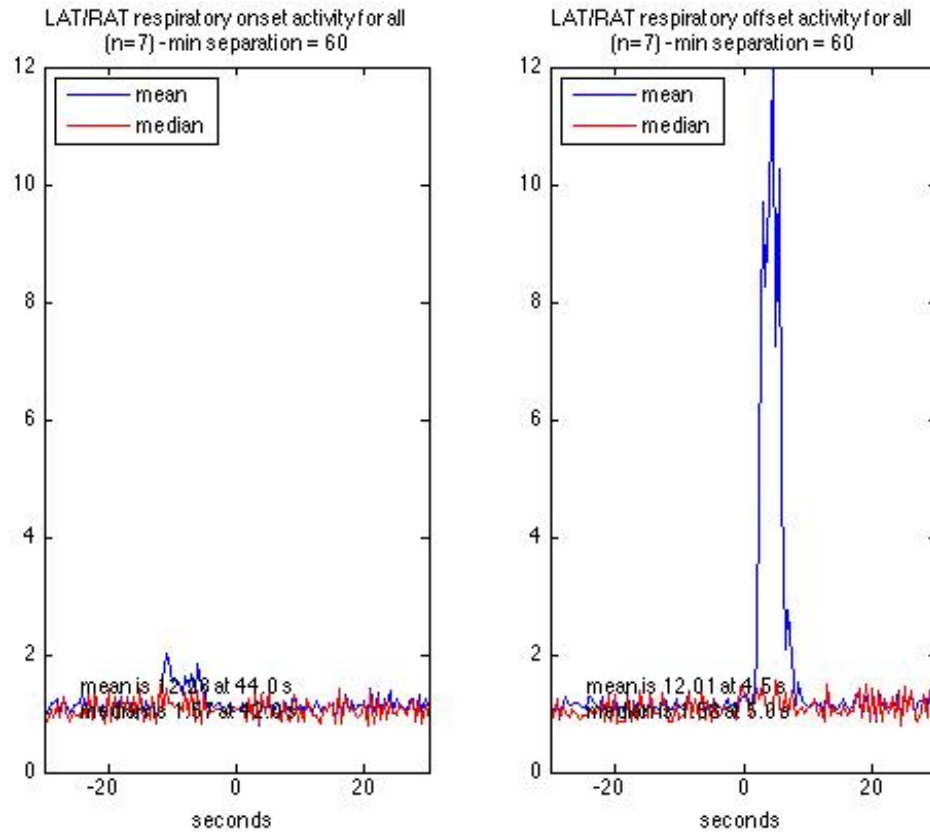
Figure 5.21: Per event EMG average in studies with only a single scored respiratory events (mean). Increased activity is still observed in this small patient sample, though the spike in activity following respiratory offset (right) is significantly higher than pre-onset activity (left).

5.2.2 PLM DISTRIBUTION AND CHARACTERISTICS

Beside PLMI and total PLM counts, we also reported LM counts (not necessarily periodic but after removal of any secondary to respiratory events). The Periodicity Index, known to differentiate RLS versus non RLS cases and first introduced by Ferri *et al* [7], was also reported as is PLM and LM night ratio, reflecting the ratio of leg

167

movements during the first part versus the second part of the night. Ratios were not calculated for individuals with zero event occurrences in either part of the night. These measures follow previous observations on the circadian effect and clustering of PLM [109, 110].

One advantage of computer-based methods is the flexibility with which it is possible to investigate events after those have been detected. For example, cardiac accelerations preceding PLM have been reported consistently with manual, or manually assisted detection methods of PLM [111-113]. For each event detected as a PLM in a given subject, we computed cardiac activation measures in the ECG surrounding PLM using the method presented by Winkelmann [112]. Heart rate (HR) for the ten cycles prior to and following each PLM is calculated using the R-R interval measured from the ECG at these points and using the heart rate directly before PLM onset as a baseline. That is, the heart rate calculated for the cycle just prior to PLM onset is subtracted from all twenty instantaneous HR calculations made for the PLM. We report two PLM associated cardiac measures: (1) *heart rate delta* which is the difference between the largest normalized HR of the ten cycles HR following PLM onset less the smallest normalized HR in the ten cycles preceding PLM onset, and (2) *heart rate slope* which is the heart delta divided by the number of cycles between the minimum and maximum detected HR values. Finally, these metrics were computed both during sleep and sleep plus intermittent wake (i.e. wake after sleep onset).

5.3 RESULTS

The complete flow chart of the final PLM detection algorithm is shown across Figure 5.22, Figure 5.23, and Figure 5.24 as three consecutive parts.

1. Adaptively clean ECG interference

Leg EMG → **Recursive Least Squares Filter (5 taps)** ← **ECG**

2. Two pass noise floor calculation and amplitude threshold generation

Cleaned leg EMG (first pass)

MA filter (20 f_s)

Noise floor (first pass)

3a. SNR+ option
Summing filter (0.02 f_s) for low noise sections (<2uV)

Generate amplitude thresholds

RMS filter (0.15 f_s)

5a. Apply amplitude threshold

Candidate LMs (first pass)

4. Adjust EMG. Regions of data classified with candidate LM are set to half the corresponding lower amplitude threshold

Cleaned leg EMG using SNR+ (to second pass)

Cleaned leg EMG (to second pass)

Figure 5.22: Complete PLM detection algorithm flow chart (part 1).
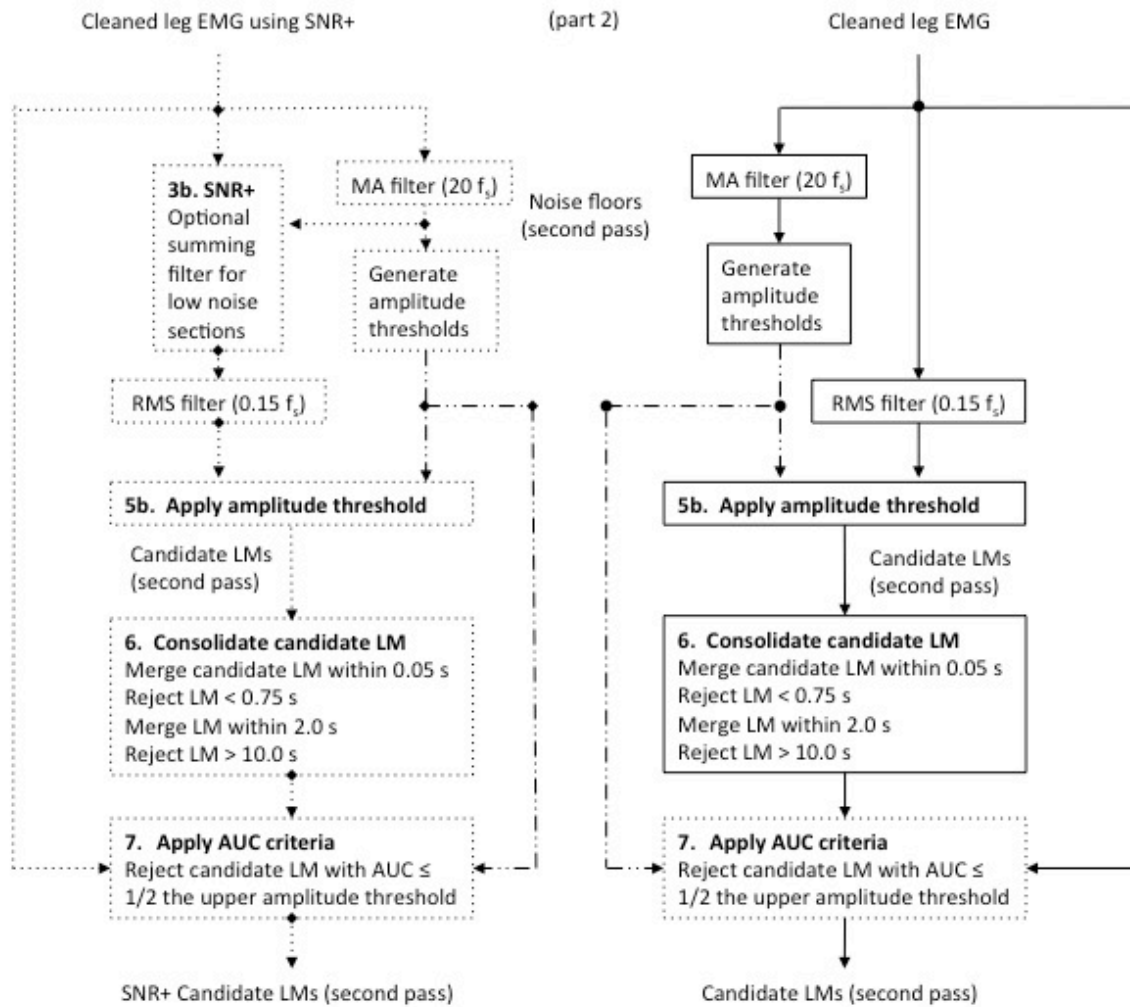
Figure 5.23: Complete PLM detection algorithm flow chart (part 2).
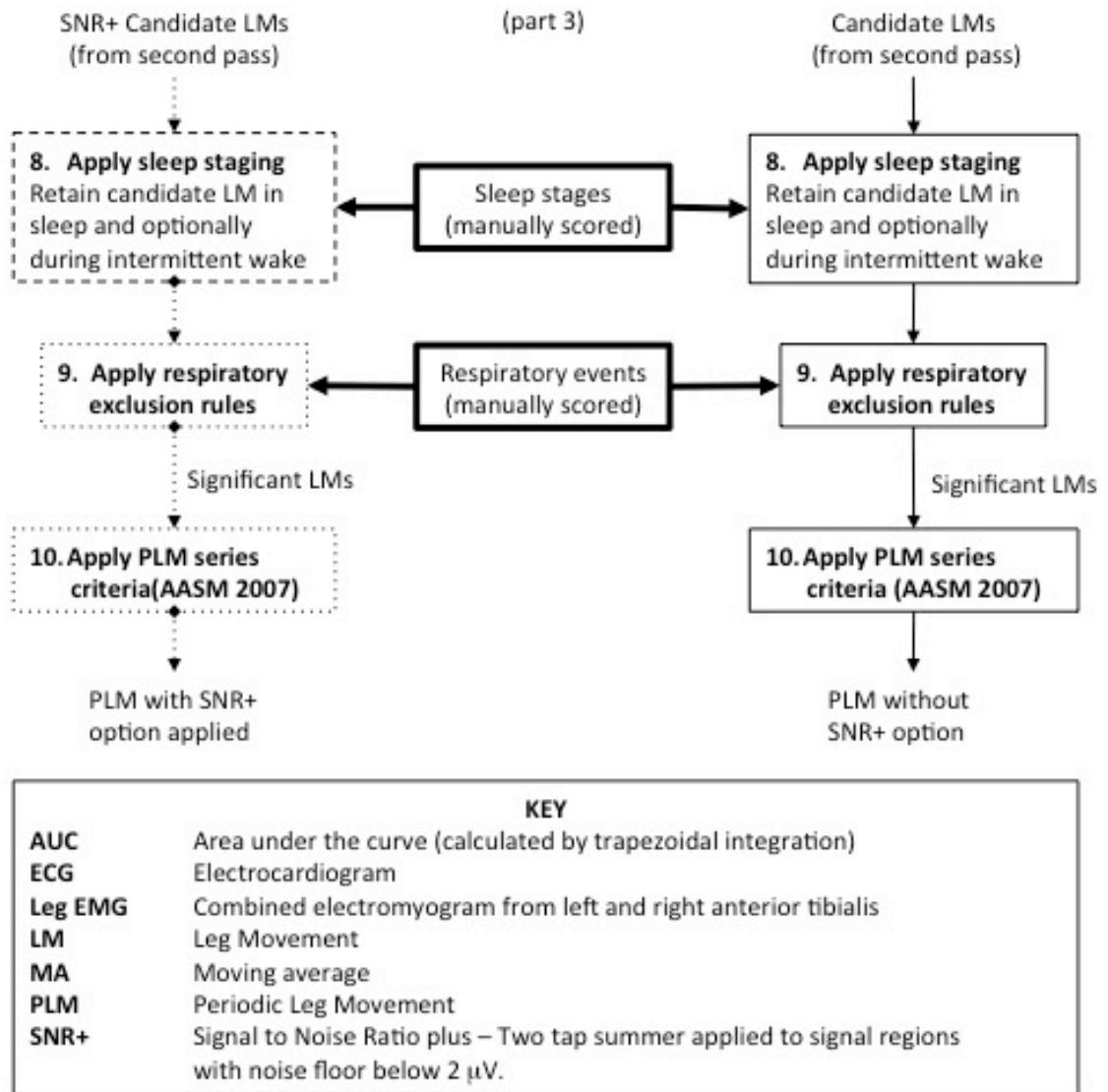
Figure 5.24: Complete PLM detection algorithm flow chart (part 3).

5.3.1 VALIDATION OF THE PLMI DETECTOR BY INDEPENDENT SCORERS

In Table 5.5, individual correlations of PLMI values during sleep as derived from AASM 2007 manual scoring (gold standard) are correlated with PLMI as computed with our detectors (with and without SNR). Table 5.5 also compares correlations to other previously published detectors. The tables give squared correlation coefficients (r^2) across

groupings of 20 subjects with and without RLS symptoms and SDB within the WSC, and overall for all 18 subjects selected from the SSC.

Table 5.5: PLMI comparisons between automatic methods and manually scoring. The squared correlation coefficient (r2) between PLMI determined automatically versus manually is shown in the table for previously published detectors and our PLM calculator with and without the SNR+ option. PLM are classified according to AASM 2007 scoring criteria with adjustment to LM classification for our classifier as described in the text.

	Wisconsin Sleep Cohort				Stanford Sleep Clinic
	RLS*, AHI≤15 (n=20)	AHI≤15 (n=20)	AHI>15 (n=20)	All (n=60)	All (n=18)
Tauchmann	0.64	0.42	0.54	0.55	0.49
Wetter	0.12	0.74	0.24	0.25	0.15
Ferri	0.93	0.66	0.85	0.86	0.41
PLM calculator	0.93	0.74	0.84	0.89	0.93
PLM calculator (SNR+)	0.95	0.78	0.94	0.93	0.94

As can be seen, our PLMI detectors had the best correlations overall and in individual subgroups, with a slight improvement when the SNR+ step was included in the detector. Overall correlation in Figure 5.25 and Figure 5.26 clearly shows that the detector did not produce outliers and was thus functioning well in all groups. The correlation seen in Figure 5.26 (SSC sample) is important as diagnoses were enriched for rare diagnosis that we felt could be problematic such as narcolepsy, hypersomnia, REM behavior disorder, obstructive sleep apnea, delayed sleep phase syndrome and insomnia.
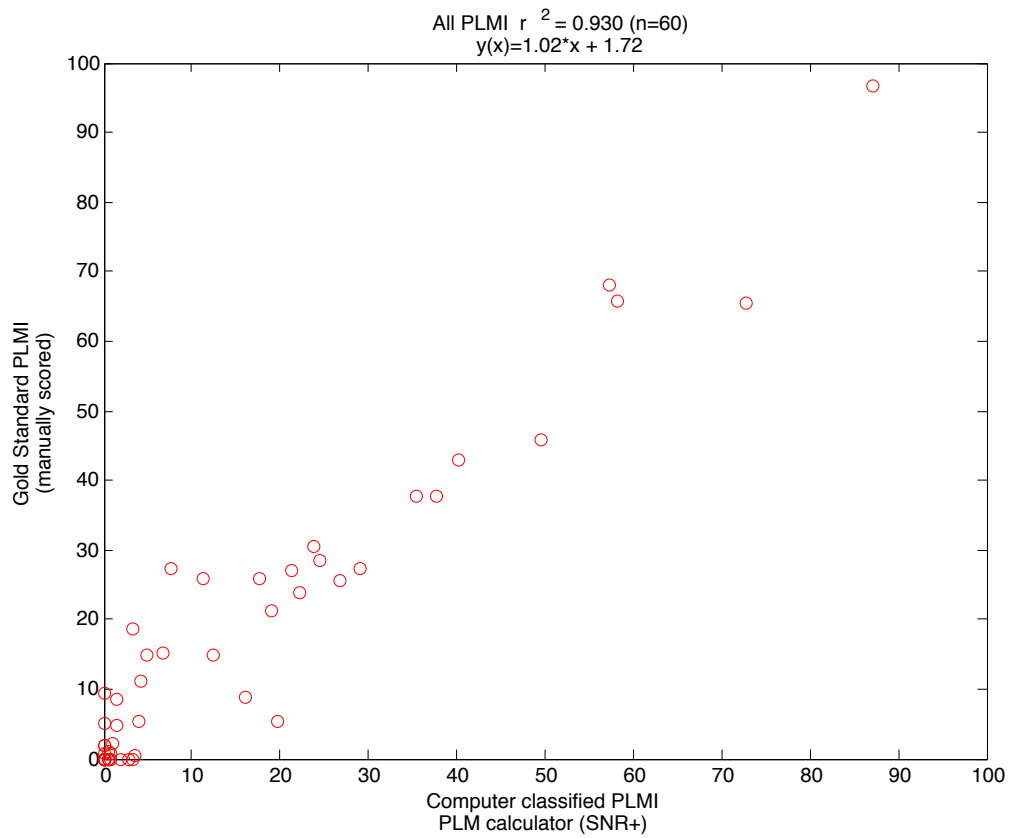
Figure 5.25: Gold standard PLMI evalution in the Wisconsin Sleep Cohort (n=60). Comparison of PLMI derived from manual scoring (y-axis) and our automatic PLM calculator using SNR+ (x-axis).
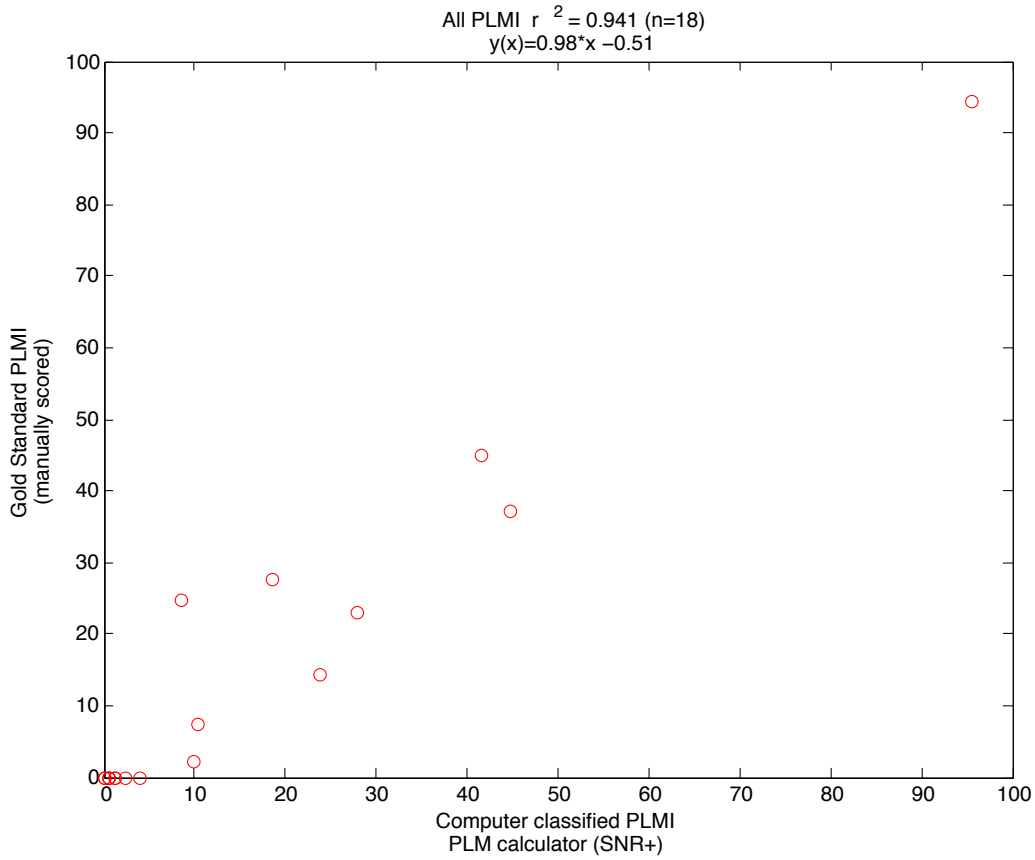
Figure 5.26: Gold standard PLMI evalution in the Stanford Sleep Clinic (n=18). Comparison of PLMI is derived from manual scoring (y-axis) and our automatic PLM calculator using SNR+ (x-axis).

5.3.2 PLM CHARACTERISTICS IN THE WSC AND SSC COHORTS

herwise continuous PLM series.

Table 5.6,

Table 5.7, and Table 5.8 show summary statistics for PLM metrics in the WSC and SSC, first using PLM detected during sleep, and then using PLM detected during sleep and intermittent wake. The motivation for including intermittent wake comes from the observation of PLM continuing, without change, during wake-scored arousals in our sleep studies. Intermittent wake sections last for 30 sec or longer and often fragment otherwise continuous PLM series.

Table 5.6: Automatically obtained PLM biomarkers in Wisconsin Sleep Cohort. PLM metrics are calculated from PLM classified using our detector (SNR+) in (A) sleep and (B) sleep with intermittent wake (wake occurring after sleep onset and before final arousal). Patients not using CPAP and having more two hours of evaluation were grouped according to their apnea-hypopnea index (AHI). A circadian effect is seen with PLMs which occur almost four times as frequently in the first half of the night compared to the second half, while LMs occur with equal probability across both halves of the night in patients with respiratory difficulty during sleep. The periodicity index is artificially affected by excluding intermittent wake and produces a statistically significant higher value in patients with AHI≤15 when examining sleep only.

A. PLM measures taken from sleep only

	AHI≤15 (738)	AHI>15 (264)	All (1069)	AHI≤15 vs AHI>15
PLMI	9.78 ±0.70	8.30 ±0.87	9.72 ±0.57	p=0.188
PLM count	60.0 ±4.3	49.5 ±5.2	58.0 ±3.3	p=0.118
LM count	102 ±5	101 ±6	103 ±4	p=0.924
Periodicity Index	0.41 ±0.01 (449)	0.33 ±0.02 (175)	0.39 ±0.01 (669)	p=0.000
Hours evaluated	6.28 ±0.04	6.08 ±0.06	6.17 ±0.03	p=0.004
Heart rate (delta)	21.2 ±0.7 (449)	20.1 ±1.12 (175)	20.6 ±0.6 (669)	p=0.401
Heart rate (slope)	2.08 ±0.07 (449)	1.98 ±0.12 (175)	2.02 ±0.06 (669)	p=0.502
PLM night ratio	3.65 ±0.61 (247)	4.95 ±1.26 (69)	3.77 ±0.52 (340)	p=0.356
LM night ratio	1.20 ±0.18 (706)	0.81 ±0.05 (254)	1.09 ±0.13 (1013)	p=0.035

B. PLM measures taken from sleep and intermittent wake

	AHI≤15 (738)	AHI>15 (264)	All (1072)	AHI≤15 vs AHI>15
PLMI	13.0 ±0.70	13.3 ±0.9	13.6 ±0.6	p=0.807
PLM count	94.5 ±5.0	97.6 ±6.8	98.5 ±4.1	p=0.721
LM count	158 ±5	171 ±8	165 ±4	p=0.163
Periodicity Index	0.36 ±0.01 (674)	0.36 ±0.01 (243)	0.36 ±0.01 (983)	p=0.981
Hours evaluated	7.37 ±0.03	7.38 ±0.05	7.35 ±0.02	p=0.911
Heart rate (delta)	38.84 ±0.93 (674)	30.92 ±1.27 (243)	36.45 ±0.74 (983)	p<0.001
Heart rate (slope)	3.97 ±0.10 (674)	3.11 ±0.14 (243)	3.72 ±0.08 (983)	p<0.001
PLM night ratio	4.58 ±0.86 (494)	4.18 ±0.47 (147)	4.34 ±0.62 (692)	p=0.685
LM night ratio	1.17 ±0.16 (723)	1.00 ±0.05 (258)	1.11 ±0.11 (1048)	p=0.325

Data are mean ± Standard Error Mean, or percentage. The number of subject used for calculations are shown in parentheses. Count is the total number of individual PLM or LM counted per study. Night ratio is the ratio of events classified in the first half of each study divided by the number of events classified in the second half. Heart rate is the normalized cardiac change (beats per minute) time locked to PLM as described in the text. P-values are calculated from the student t-test with significance level of 0.05. Periodicity index, heart rate, and PLM ratio is only calculated in the presence of PLM. PLM and LM night ratios are only calculated in cases where PLM or LM occur during both the first and second half of the study.

Comparing herwise continuous PLM series.

Table 5.6 (A and B) with

Table 5.7 and Table 5.8 and the two samples overall, PLMIs were higher in WSC versus SSC, likely reflecting age differences. Other differences appear (notably in PLM night ratio, a reflection of diurnal or circadian control of PLM) and could be explained by age or the nature of the cohort, population-based sample versus clinical sample. Further analysis, outside of the scope of this manuscript may reveal the source for these differences, which overall were relatively modest.

A key feature of our detector in comparison to others is removal of LM in association with SDB. The algorithm was successful as evidenced by PLMI that were not statistically different in subjects with AHI≤15 versus >15 in the WSC and between patients with and without SDB in the SSC. Problematically, the periodicity index varied with SDB status when PLMs were only assessed during sleep but not wake (herwise continuous PLM series.

Table 5.6 A versus B). This suggests that the periodicity index cannot be properly calculated if intermittent wake is removed, especially in cases with sleep apnea where arousals are common.

Another difference was in the HR slope and delta, which were significantly smaller in patients with SDB. A sub-analysis after exclusion of patients taking beta-blockers suggested this result was not secondary to medication effects. We also computed numbers of PLMs per sleep stages, and found increased occurrence of PLM during Stage 2 sleep (data not shown). This observation follows that of Ferri et al, who used an automatic detection followed by human confirmation and adjustment [111].

Table 5.7: Automatically obtained PLM biomarkers in Stanford Sleep Cohort from sleep only. PLM metrics are calculated from PLM classified using our detector (SNR+). Patients having more two hours of evaluation were grouped according to sleep pathology as determined by formal medical diagnosis. The circadian effect is reversed in narcolepsy, with PLM more likely to occur during the second half of the sleep study, and less extreme in insomnia where PLM are only slightly more frequent (i.e. 33%) in the first half of the study.

	All (756)	Delayed Phase Syndrome (14)	Insomnia (140)	Narcolepsy (17)
PLMI	8.1 ±0.65	4.48 ±1.80	6.06 ±1.20	12.9 ±4.18
PLM count	48.1 ±3.85	27.3 ±10.7	33.3 ±6.05	82.4 ±26.7
LM count	95.1 ±4.20	70.9 ±13.6	73.8 ±6.84	139 ±32.5
Periodicity Index	0.31 ±0.01 (478)	0.29 ±0.07 (10)	0.30 ±0.03 (76)	0.40 ±0.07 (12)
Hours evaluated	6.14 ±0.04	6.18 ±0.34	6.12 ±0.10	6.71 ±0.27
Heart rate (delta)	20.7 ±0.66 (478)	15.1 ±1.64 (10)	21 ±3.01 (76)	18.5 ±2.02 (12)
Heart rate (slope)	2.07 ±0.07 (478)	1.40 ±0.12 (10)	2.12 ±0.33 (76)	1.90 ±0.22 (12)
PLM night ratio	2.48 ±0.19 (188)	1.70 ±1.07 (3)	1.33 ±0.25 (25)	0.49 ±0.20 (7)
LM night ratio	0.54 ±0.10 (610)	0.66 ±0.41 (10)	0.41 ±0.08 (111)	0.32 ±0.06 (13)

	REM Behavior Disorder (4)	Restless Legs Syndrome (24)	Sleep Disordered Breathing (597)	Other (37)	p
PLMI	32.2 ±21.9	10.2 ±4.31	7.71 ±0.71	6.09 ±2.60	0.091
PLM count	186 ±131	53 ±20.8	46.3 ±4.29	34.5 ±14.25	0.073
LM count	254 ±140	106 ±22	93.1 ±4.67	82.1 ±17.16	0.021
Periodicity Index	0.43 ±0.14	0.28 ±0.06 (20)	0.30 ±0.01 (376)	0.22 ±0.05 (23)	0.564
Hours evaluated	6.01 ±0.66	5.97 ±0.25	6.15 ±0.05	5.97 ±0.19	0.573
Heart rate (delta)	7.94 ±1.86	26.6 ±7.17 (20)	20.2 ±0.95 (376)	21 ±2.09 (23)	0.663
Heart rate (slope)	0.74 ±0.17	2.89 ±0.90 (20)	2.03 ±0.11 (376)	2.1 ±0.22 (23)	0.556
PLM night ratio	1.47 ±0.78 (2)	9.36 ±8.07 (4)	2.49 ±0.46 (149)	2.84 ±1.43 (6)	0.301
LM night ratio	0.51 ±0.19 (3)	0.42 ±0.19 (15)	0.55 ±0.05 (487)	0.45 ±0.10 (26)	0.888

Table 5.8: Automatically obtained PLM biomarkers in Stanford Sleep Cohort measured in sleep with intermittent wake (wake occurring after sleep onset and before final arousal). Patients having more two hours of evaluation were grouped according to sleep pathology as determined by formal medical diagnosis.

	All (757)	Delayed Phase Syndrome (14)	Insomnia (140)	Narcolepsy (17)
PLMI	12.9 ±0.67	8.62 ±1.78	9.68 ±1.18	20.1 ±5.5
PLM count	96.4 ±5.0	63.4 ±13.8	70.5 ±8.2	167 ±46
LM count	167 ±5	132 ±20	136 ±9	240 ±51
Periodicity Index	0.32 ±0.01 (714)	0.29 ±0.04	0.28 ±0.02 (134)	0.43 ±0.06 (15)
Hours evaluated	7.44 ±0.04	7.29 ±0.32	7.39 ±0.08	8.19 ±0.22
Heart rate (delta)	27.4 ±0.73 (714)	18.9 ±2.11	30.3 ±2.07 (134)	22.4 ±2.17 (15)
Heart rate (slope)	2.78 ±0.08 (714)	1.78 ±0.19	3.06 ±0.23 (134)	2.38 ±0.25 (15)
PLM night ratio	1.83 ±0.20 (474)	2.14 ±0.72 (10)	1.09 ±0.14 (78)	0.603 ±0.16 (14)
LM night ratio	0.62 ±0.03 (702)	0.89 ±0.26 (13)	0.47 ±0.05 (126)	0.406 ±0.08 (16)

	REM Behavior Disorder (4)	Restless Legs Syndrome (24)	Sleep Disordered Breathing (597)	Other (38)	p
PLMI	40.9 ±20.2	17.9 ±4.4	12.5 ±0.7	11.3 ±2.5	0.006
PLM count	280 ±139	131 ±29	94.5 ±5.6	80.1 ±17.0	0.006
LM count	367 ±140	209 ±31	165 ±6	156 ±20	0.003
Periodicity Index	0.53 ±0.12	0.37 ±0.05 (23)	0.31 ±0.01 (563)	0.28 ±0.03 (37)	0.026
Hours evaluated	7.39 ±0.74	7.52 ±0.21	7.46 ±0.04	7.26 ±0.15	0.089
Heart rate (delta)	11.1 ±2.04	31.8±6.3 (23)	27.2 ±0.85 (563)	28.8 ±2.22 (37)	0.191
Heart rate (slope)	1.08 ±0.18	3.34±0.71(23)	2.76 ±0.10 (563)	2.94 ±0.26 (37)	0.240
PLM night ratio	0.72 ±0.23	1.35 ±0.42 (18)	1.83 ±0.22 (382)	1.37 ±0.39 (24)	0.772
LM night ratio	0.57 ±0.18	0.41 ±0.07 (21)	0.63 ±0.03 (553)	0.43 ±0.08 (31)	0.107

The data in

Table 5.7 and Table 5.8 is shown as the mean ± standard error of the mean, or as a percentage. The number of subject used for calculations are shown in parentheses. Count is the total number of individual PLM or LM counted per study. Night ratio is the ratio of events classified in the first half of each study divided by the number of events classified in the second half. Heart rate is the normalized cardiac change (beats per minute) time locked to PLM as described in the text. Probabilities (p) are calculated using one-way analysis of variance between groups with a 0.05 significance level. Periodicity index, heart rate, and PLM ratio is only calculated in the presence of PLM. PLM and LM night ratios are only calculated in cases where PLM or LM occur during both the first and second half of the study.

5.3.3 PLM BY DIAGNOSTIC GROUPS

Table 5.7 and Table 5.8 report on PLMs across diagnostic groups at the Stanford Sleep Clinic. As expected, PLMI (and periodicity index, when computed during sleep and wake, see above) was highest in narcolepsy and RLS; these pathologies are known to have the highest association with PLMs. Data from REM behavior disorder subjects was difficult to interpret, as our sample included only 4 subjects. Insomnia and delayed sleep phase syndrome had the lowest PLMI and periodicity indices, while sleep-disordered breathing was intermediate. A statistical difference was found using a one-way analysis of variance between groups for PLMI, PLM count, and periodicity index in examination of sleep with intermittent wake that was not seen in sleep only.

5.3.4 CODE FOR THE STANFORD PLM DETECTOR

The detector uses the SEV. Source code for both the SEV and the Stanford PLM detector are freely available online at http://www.stanford.edu/~hyatt4. The online repository for open source development is available at http://www.github.com/informaton/sev.

5.4 DISCUSSION

In this chapter, I described a novel PLM detector and its validation in two independent adult samples (patients-based and population-based cohorts). My goal was to create a robust detector that would closely approximate manual scoring by an experienced technician using AASM 2007 PLM scoring rules. The detector was optimized to remove false signals from leg channels, such as ECG contamination, or fragmentary myoclonus-like patterns. It may use one or two leg channels, although in our cases, two combined leg channels were used in both cohorts examined.

Although we started with the AASM 2007 criteria, two modifications of the rules were necessary to improve detection. First, we found that defining a LM as a having EMG signal exceeding 8 μV above baseline and then falling below 2 μV from baseline was difficult to use if background noise was either very high or very low. In the former case, leg movements may still be detectable above baseline, but if starting from a high

baseline, may be scored spuriously when the total signal crossed the 8-μV threshold even following a small rise in signal or an artifact. In the later cases, small leg movements may be obvious and periodic, but do not reach the 8 μV threshold. To accommodate these cases, we created the SNR+ option and continually account for changes to the noise floor as it varies from almost none to high (for example 6 μV, see Figure 5.4 to Figure 5.13).

Second, there has been much discussion regarding exclusion of LM secondary to SDB but no clear science based consensus. At one end of the spectrum, some authors have argued that almost all PLM events are secondary or connected to SDB. At the other end of the spectrum, investigators in the RLS field do not score SDB and remove LM events that could be secondary to SDB and associated arousal. To address this issue, the AASM 2007 Scoring Manual excludes LM activity 0.5 sec prior until 0.5 sec after respiratory events from PLM. Interestingly, however, we found this rule to be arbitrary and not reflecting reality. Indeed, using this rule, computed PLMI was greater in individuals with AHI above 15, and lesser in those with AHI below 15, suggesting false detections (see methods). Examination of EMG activity reflecting LM time-locked with respiratory events was next performed, and the 0.5 sec window applied to respiratory event boundaries inadequate to remove LM secondary to SDB, as EEG activity peaks approximately 3 sec following SDB. More surprisingly, we also found increased EMG a few seconds before the initiation of a SDB event. Whereas jerking activity of the body is commonly known to accompany the recovery breath at respiratory event offset, increased activity *prior* to respiratory event onset was not anticipated and several steps were taken to examine the results for error or explanation. For example, we only included isolated events without SDB immediately prior each event analyzed to avoid cofounding effects of prior events, but this did not change the signal. Not finding any artificial explanation, we conclude the finding to be physiological, perhaps reflecting a brief jerk when the airway is first obstructed or the leg jerk may precipitate a sharp inspiration sufficient to close or narrow a (pathologically) compliant airway. Alternatively, LM and SDB may be

connected in a more complex way in some cases, for example via changes in arousal threshold that are known to be associated with LM generation. It is notable that PLMs are associated with heart rate changes prior to the event, and that K-complexes and arousals associated with PLMs have been shown to occur periodically without the motor event, for example when patients with RLS are treated with dopamine agonists.

These results notwithstanding, the new exclusion window proposed, 5 sec before initiation (plus 0.5 sec within the SDB event) and 5 sec after termination (plus 0.5 sec before termination), is based on empirical examination of leg EMG activity time locked to manually scored respiratory events that shows the strongest association at the ends (i.e. onset and offset) (see Figure 5.14). Using this rule, PLMI did not differ between subjects with and without SDB, indicating success of the algorithm for this purpose.

A second controversial issue pertains to the inclusion of leg movements during intermittent wakefulness. PLM are known to continue during intermittent wake, thus removing these movements may break a PLM series. On another hand, wake may be contaminated with LM secondary to voluntary movements that can also confound the issue. Studying cases with various sleep disorders in the SSC, results were similar, but the periodicity index, a measure that is known to best differentiate true PLM, was less reliably measured, thus we advocate use of PLMI during sleep plus intermittent wake.

In the course of this study, we found that heart rate activation in association with PLMs was lower in subjects with SDB versus without. This preliminary result may reflect decreased cardiovascular response in these subjects or increased arousal threshold due to sleepiness. Indeed, it is striking to note that most patients with SDB do not recall the severe sleep disruption associated with breathing events and have attenuate respiratory occlusion related evoked potentials. Still there are several nuances with measuring cardiac activity with PLM that affect the outcome as presented by Ferri's excellent commentary [111]. Our goal here is not to provide definite explanation on the matter, but

rather how our detector may be used to further explore these and other avenues of research related to PLM.

In comparison with other detectors, the Stanford PLM detector fared best. Validation of the detector in two independent cohorts in comparison with gold standard, manual scoring using AASM 2007 criteria, revealed a very high correlation (Figure 5.25 and Figure 5.26). The fact the validation samples were enriched in complex associations such as unusual pathologies (REM behavior disorder, narcolepsy etc.) or SDB gives confidence that this detector should perform well in multiple clinical settings. We hope that dissemination of the detector will ensure widespread use and help comparative studies of PLMs.

5.5 CONCLUSION

This chapter presented a robust, automated algorithm for scoring PLMs in both a general population and also a sleep clinic sample. The PLM detector was designed and validated using NPSG-based sleep studies taken adult participants from the WSC (n=1,073, 2000-2004) and a sample of successive SSC patients undergoing baseline nocturnal polysomnography (n=760, 1999-2007).

NPSGs from the WSC were first used to develop the algorithm to identify PLMs using AASM 2007 criteria (e.g. 8 μV above baseline, 0.5-10 seconds (sec) duration, exclusion of LMs adjacent to respiratory events), and minor modifications were made to optimize performance. LMs (n=119,277) manually scored using the 1995 AASM guidelines were used as a first guide to test performance of this detector and other published algorithms using ROCs. Outlier scores were examined, and rules were modified when appropriate to accommodate problems, notably ECG interference, exclusion of movements adjacent to respiratory events and adjustment of noise floor/amplitude ratio.

To further test the algorithm, the PLMI mean and range during sleep as well as other metrics were compared in subjects reporting RLS symptoms or SDB (i.e., an AHI > 15).

A registered PSG technologist scored PLMs using the AASM 2007 criteria in 60 randomly selected WSC subjects with RLS symptoms (n=20), SDB (n=20), and normal sleepers (n=20). The manually scored PLMI was correlated with the automatically computed PLMI. To extend on these findings, the same detector was applied to the SSC, and validated in 18 subjects with various sleep disorders (SDB, RLS, REM sleep behavior disorder, narcolepsy-cataplexy, insomnia, delayed sleep phase syndrome, 3 of each) by parallel blind human scoring (n=1,733 PLMs). The detector's performance was also compared using all manually scored studies with that of other known detectors.

The final detector provided a PLMI for sleep and sleep and wake, plus periodicity index and other LM metrics (if PLMI>0). The optimized PLM detector incorporated adaptive noise cancelling of cardiac interference, noise-floor adjustable detection thresholds, removes leg movements secondary to SDB within 5 sec of respiratory events and deletes popping electrodes or fragmentary myoclonus. The detector exhibited higher sensitivity and specificity than other detectors and correlated well with gold standard calculated PLMI in both cohorts (r^2=0.93 in WSC and r^2=0.94 in SSC). Its performance was much improved in comparison to other existing detectors, and showed that automated detection of PLM is possible in controls and patients with various sleep disorder diagnosis. The following chapter will use the PLM detection algorithm introduced here to identify single nucleotide polymorphisms with increased susceptibility for PLMs.

As mentioned in the introduction, the contents of this chapter are taken from the manuscript, "Design and Validation of a Periodic Leg Movement Detector," which has been submitted for publication.

CHAPTER 6   SINGLE NUCLEOTIDE POLYMORPHISM SUSCEPTIBILITY FOR
PERIODIC LEG MOVEMENT

In this chapter, I investigate thirteen single nucleotide polymorphisms (SNPs) previously associated with Restless Legs Syndrome (RLS) for susceptibility of periodic leg movements (PLMs) in 1,090 adult participants from the Wisconsin Sleep Cohort (WSC) (2,394 observations, 2000-2012). PLMs were enumerated using the automated detector discussed in the previous chapter.

6.1 INTRODUCTION

As mentioned in the prior chapter, periodic leg movements (PLMs) are episodic, involuntary muscle contractions that occur during sleep. Restless Legs Syndrome (RLS) is often associated with PLMs, with four out of five patients diagnosed with RLS exhibiting PLMs[7]. However, as discussed in Section 1.3.2, PLMs can also occur without RLS symptoms.

In this chapter, I report on the comparative frequencies of these SNPs across persons with PLMs with and without RLS symptoms drawn from the Wisconsin Sleep Cohort (WSC), The SNP and sleep data were obtained between 2000 to 2012 as described in Chapter 1.4.1.

Recent investigations have revealed genetic associations with RLS. Most notably, 13 single nucleotide peptides (SNPs) were identified as RLS susceptibility markers in a genome-wide association study (GWAS) described in 2011[15]. The study examined 922 RLS cases and 1,526 controls with replication in 3,935 RLS cases and 5,754 controls (all of European ancestry). Association of the BTBD9 and MEIS1 genes with RLS and end stage renal disease was reported in another study of Germans (200 RLS cases, 443 controls) with replication in a Greek sample (141 RLS cases, 393 controls)[114]. Another team, using an Icelandic discovery sample with replications in Icelandic and U.S. samples found association of BTBD9 in PLM in sleep (PLMS)[115] apart from RLS diagnosis as adopted from a 1995 criteria[116]. PLMs were measured using a small tri-axial accelerometer worn on the ankle. The ambulatory device could not distinguish sleep from wake and so the researchers classified participants with PLMs if during any of the five nights of sleep, the ambulatory device detected five or more PLM during an hour while recumbent during their major rest period[115]. These studies did not use PSG measures of PLMs or account for additional effects on PLMs like use of medication that can aggravate or inhibit factors contributing to RLS.

In this chapter I examine the 13 SNPs mentioned above for susceptibility to PLMs with and without RLS symptoms. Section 6.2 presents the cohort, the process for selecting RLS symptoms and PLM phenotypes, and the statistical models to evaluate SNP susceptibility to the phenotypes. Section 6.3 gives the results of this analysis, and is followed by a discussion of the results and concluding remarks in Section 6.4.

6.2 METHODS

6.2.1 COHORT USED IN THE ANALYSIS

A total of 2,394 Nocturnal PSG (NPSG) studies from 1,086 WSC participants were used in this study. Electroencephalography (EEG), electrooculography (EOG), and chin electromyography (EMG) were used to score sleep stages for each 30 second epoch using standard R&K criteria[1]. The WSC and its NPSG collection montage are described in Chapter 1.4.1.

6.2.2 RLS SYMPTOMS PHENOTYPE

WSC participants were stratified according to RLS symptoms based on questionnaire responses from a previous study of the cohort in 2003[103]. Figure 6.1 shows the survey questions used to identify RLS symptoms. Patients were asked to provide the frequency with which they felt (a) repeated urge to move legs, (b) strange and uncomfortable feelings in the legs, and/or (c) periods of several leg jumps or jerks. The response answers were never, less than once a month, monthly, weekly, and nightly. Two additional yes/no questions were (d) Do you these feelings just mentioned get better when you get and start walking? (e) Do these feelings just mentioned disrupt your sleep? The questionnaire did not completely address all RLS diagnostic criteria put forth by the National Institutes of Health (NIH)[104]. Notably it did not ask for the symptoms to be worse at night; hence our use of the term RLS *symptoms* rather than RLS.

Q13. **How often, when you are sitting or lying down, do you have any of the following feelings in your legs?**
   a. Repeated urge to move your legs
   b. Strange and uncomfortable feelings in your legs
   c. Periods of several leg jerks or jumps in a row
   d. Do these leg feelings just mentioned get better when you get up and start walking? (Yes; No)
   e. Do these leg feelings just mentioned disrupt your sleep? (Yes, some; Yes, a great deal; No)

Figure 6.1: Question 13 from WSC survey mailing in early 2000.

Table 6.1: Questionnaire response criteria RLS symptoms category.

RLS symptoms category	Questionnaire response
Category A: Definite RLS symptoms	(a) Weekly or more often (d) Yes (e) Yes.
Category B: Possible RLS symptoms	(a) Monthly or more frequent (d) Yes. Category B could not include members already in Category A
Category C: No RLS symptoms	(a) Less than monthly (b) Less than monthly or missing
Category D: Unknown or uncertain	Responses not categorized as A, B, or C to include missing responses.

RLS symptoms were split into three categories using the response criteria listed in Table 6.1. Category A (n=186, observations=360), definite RLS symptoms, is defined as response to (a) as weekly or more often, (d) yes, and (e) yes. Category B (n=190, observations=384), moderate RLS symptoms, is defined as response (a) monthly or more frequent, and (d) yes. Category B does not include members already in Category A.

Category C (n=523, observations=997), no RLS symptoms, is defined by responses to (a) as less than monthly and (b) as either missing or less than monthly. Remaining subjects (n=171, observations=321) were excluded from RLS symptom stratification (i.e. missing responses or responses that did not fit into Category A, B, or C). (See Table 6.1.)

Categories A and B are merged to form Category *AB*, the positive RLS symptoms phenotype, *RLS+*. Category C defines the control group, *RLS-*.

Table 6.2 shows demographic, polysomnograms sleep measures, and medication use in the WSC stratified by RLS symptoms. The difference in choosing Category AB as the RLS+ phenotypes instead of Category A only is negligible.

Table 6.2 : Demographic, polysomnograms, and medication information for Wisconsin Sleep cohort stratified by RLS symptoms.

	RLS(A) (n=188, o=417)	RLS(B) (n=192, o=443)	RLS(AB) (n=380, o=860)	RLS(C) (n=535, o=1169)
Demographics				
Age	58.8 ±0.62 (188)	58.3 ±0.64 (191)	58.6 ±0.44 (379)	59.0 ±0.34 (535)
Sex, Male (%)	47.9% (188)	51.8% (191)	49.9% (379)	55.2% (534)
Clinical Data				
Body Mass Index (kg/m²)	33.7 ±0.57 (186)	30.3 ±0.44 (191)	32.0 ±0.37 (377)	31.3 ±0.30 (532)
Apnea Hypopnea Index (AHI)	15.6 ±1.36 (179)	13.2 ±1.15 (187)	14.4 ±0.89 (366)	12.4 ±0.59 (516)
AHI > 15 (%)	33.5% (179)	28.9% (187)	31.1% (366)	27.5% (516)
PLMI	18.7 ±1.79 (188)	15.3 ±1.20 (191)	17.0 ±1.07 (379)	10.8 ±0.64 (535)
Polysomnogram				
TST (hour)	5.98 ±0.07 (188)	6.13 ±0.07 (191)	6.06 ±0.05 (379)	6.11 ±0.04 (535)
Stage 1 (%)	11.4 ±11.39 (418)	9.86 ±9.86 (444)	10.6 ±10.62 (861)	10.7 ±10.65 (1170)
Stage 2 (%)	65.5 ±65.47 (418)	65.7 ±65.73 (444)	65.7 ±65.68 (861)	66.0 ±66.01 (1170)
SWS (%)	7.01 ±7.01 (418)	7.92 ±7.92 (444)	7.49 ±7.49 (861)	6.81 ±6.81 (1170)
REM (%)	15.6 ±15.64 (418)	16.1 ±16.10 (444)	15.9 ±15.89 (861)	16.3 ±16.25 (1170)
WASO (hour)	1.35 ±0.06 (188)	1.19 ±0.04 (191)	1.27 ±0.04 (379)	1.22 ±0.03 (535)
Medication				
RLS symptom aggravators	40.00%	38.10%	39.10%	33.00%
RLS symptom inhibitors	15.10%	9.70%	12.30%	7.50%

Table 6.3: Student t-test comparisons between RLS symptoms categories in the Wisconsin Sleep Cohort.

	RLS(A) vs RLS(C)	RLS(B) vs RLS(C)	RLS(AB) vs RLS(C)	p (ANOVA)
Demographics				
Age	p=0.700	p=0.330	p=0.382	p=0.592
	OR=0.74	OR=0.87	OR=0.81	X²=3.17
Sex, Male (%)	p=0.081	p=0.417	p=0.109	p=0.075
Clinical Data				
Body Mass Index (kg/m²)	**p<1e-3**	p=0.064	p=0.144	p<1e-5
Apnea Hypopnea Index (AHI)	**p=0.032**	p=0.535	p=0.065	p=0.050
	OR=1.33	OR=1.07	OR=1.19	X²=2.33
AHI > 15 (%)	p=0.128	p=0.723	p=0.242	p=0.127
PLMI	**p<1e-4**	p<1e-3	**p<1e-6**	p<1e-6
Polysomnogram				
TST (hour)	p=0.126	p=0.819	p=0.391	p=0.214
Stage 1 (%)	p=0.962	p=0.957	p=0.998	p=0.997
Stage 2 (%)	p=0.995	p=0.998	p=0.997	p=1.000
SWS (%)	p=0.983	p=0.915	p=0.946	p=0.995
REM (%)	p=0.978	p=0.994	p=0.987	p=1.000
WASO (hour)	p=0.028	p=0.621	p=0.231	p=0.030
Medication				
RLS symptom aggravators	**OR=1.36** **p=0.010**	OR=1.25 p=0.053	**OR=1.30** **p=0.005**	X²=8.25 p=0.004
RLS symptom inhibitors	**OR=2.19** **p<1e-5**	OR=1.32 p=0.153	**OR=1.73** **p<1e-3**	X²=20.44 p<1e-5

198

Table 6.3 shows student t-test comparisons between A and C, B and C, and AB and C. The gender difference between A and C is closest to RLS epidemiology, where approximately two-thirds are female. PLMI and medication use decrease from category A to B to C and show statistically significant (i.e. $p<0.05$) differences between categories. The difference in PLMI is greater between AB and C than A and C because of the additional power gained by adding B. Prescription medications are classified as RLS symptom aggravators and RLS symptom inhibitors. Aggravators include antidepressants (e.g. selective seritonin reuptake inhibitors, tricyclics), antipsychotics, antiseizure medication, and antihistamines. Inhibitors include benzodiazepines, opiates, and medication for Parkinson's disease. Body mass index (BMI) and apnea-hypopnea index (AHI) are statistically significant between Category A and Category C, but not Category AB vs. Category C.

6.2.3 PLM PHENOTYPE

PLMI, the average number of PLMs per hour of sleep plus wake after sleep onset (WASO), was calculated automatically using the detection method we described in Chapter 5. First, the algorithm removes cardiac interference from the EMG channel using an adaptive filter that reduces false positive detections. Next, a two-pass noise floor calculation is performed and variable amplitude thresholds generated to account for changes in the baseline noise. Candidate LMs are identified where the filtered signal (i.e. adaptively cleaned and root-mean-square filtered) meets the amplitude and duration which fall in line with AASM 2007 criteria for *significant LM* as described in the original manuscript [62]. LMs whose area under the curve is too small or that fall within the respiratory exclusion window of respiratory events described above[1] are removed from candidacy. Remaining LMs are deemed *significant* and scored for PLM using AASM 2007 criteria.

---

[1] LM that occur 5.0 seconds before to 0.5 seconds after an apnea or hypopnea starts or which occur 0.5 seconds before until 5.0 seconds after the respiratory events end are

WSC PLM metrics for the RLS symptoms phenotype (i.e. RLS+) are given in Table 6.4. Statistically significant (p<0.05) differences are observed for PLMI, periodicity index, PLM distribution by sleep stage and WASO, and cardiac changes measured from the ECG channel time locked to all PLM detections. The periodicity index is the number of inter-PLM-movement intervals greater than 10 s divided by the total number of inter-movement intervals (i.e. the elapsed time between consecutive LM onsets (i.e. the total LM count less one). It is based on the observation that RLS patients exhibit longer inter-movement intervals than non-RLS patients[7]. The *PLM night ratio* quantifies the

Table 6.4: Wisconsin Sleep Cohort PLM metrics stratified by RLS symptoms.

	RLS(+) (860)	RLS(-) (1169)	All (2394)	RLS(+) vs RLS(-)
Hours evaluated	7.32 ±0.03	7.33 ±0.02	7.32 ±0.02	p=0.743
PLMI	20.89 ±0.86	13.94 ±0.55	16.96 ±0.45	p<1e-10
Periodicity Index	0.44 ±0.01 (807)	0.35 ±0.01 (1098)	0.39 ±0.01 (2244)	p<1e-14
PLM stage 1	11.43 ±0.73 (807)	7.63 ±0.52 (1098)	9.05 ±0.38 (2244)	p<1e-4
PLM stage 2	82.67 ±4.31 (807)	53.11 ±2.76 (1098)	66.92 ±2.31 (2244)	p<1e-8
PLM SWS	8.53 ±0.82 (807)	5.45 ±0.50 (1098)	6.64 ±0.41 (2244)	p=0.001
PLM REM	5.88 ±0.47 (807)	3.93 ±0.31 (1098)	4.95 ±0.27 (2244)	p<1e-3
PLM WASO	54.28 ±2.38 (807)	37.78 ±1.29 (1098)	44.32 ±1.15 (2244)	p<1e-8
PLM night ratio (first half/second half)	2.69 ±0.25 (663)	2.28 ±0.16 (854)	2.44 ±0.13 (1784)	p=0.166
Attrition (PLM/hour)	-4.05 ±0.23	-2.80 ±0.15 (1167)	-3.30 ±0.12 (2392)	p<1e-5
Heart rate (delta)	29.55 ±0.78 (807)	30.86 ±0.67 (1098)	30.48 ±0.48 (2244)	p=0.199
Heart rate (slope)	3.00 ±0.09 (807)	3.13 ±0.08 (1098)	3.09 ±0.05 (2244)	p=0.263
PLMI > 5%	63.2	49.45	55.79	X^2=35.538 p<1e-8 OR=1.755
PLMI > 10%	55.14	40.16	46.88	X^2=41.947 p<1e-10 OR=1.831
PLMI > 15%	45.11	31.6	37.25	X^2=36.251 p<1e-8 OR=1.778
PLMI > 30%	25.65	15.12	19.52	X^2=32.767 p<1e-7 OR=1.937

circadian or diurnal effect observed in PLM[26, 27] and is presented here as the PLM count from the first half of the PSG sleep study divided by the PLM count measured in the second half. The PLM ratio uses sleep onset to define a study's start and total sleep time

plus WASO to determine the study's mid point. *PLM attrition* is the per hour change in PLM as determined by linear regression. Because PLMI is readily accepted and understood in the sleep community we examined several PLMI cut-points in selecting the PLM phenotype.

The positive PLM phenotype (*PLMI+*) is defined as PLMI≥15 and the control group or negative PLM phenotype (*PLMI-*) is PLMI<15. We also considered PLMI>5 and PLMI<15 vs. PLMI≥15 as our control vs. phenotype grouping (not shown), but the results were not as significant as PLMI≥15 vs. PLMI<15.

Table 6.5 compares PLM+ and controls (PLMI-) in the WSC. PLM+ observations are predominantly male (60.0%) and older (62.2 years) than controls (54.0% male and 59.7 years of age). PLMI+ shows less total sleep time (5.85 vs. 6.20 hours, $p<10^{-13}$) and greater WASO duration (1.45 hours vs. 1.13 hours, $p<10^{-inf}$) than controls. Lastly, RLS+ is more prevalent in PLM+ (44.3% vs. 31.7%, $p<10^{-8}$) and RLS- less prevalent (40.3% vs. 53.1%, $p<10^{-8}$).

Table 6.5: Student t-test comparisons between PLM+ phenotype and controls in the Wisconsin Sleep Cohort.

	PLMI>=15 (797 observations)	PLMI<15 (1597 observations)	All (2394 observations)	PLMI≥15 vs PLMI<15
Demographics				
Age	62.2 ±0.30	58.4 ±0.20	59.7 ±0.17	p<1e-Inf
Sex, Male (%)	60.70%	50.70%	54.00%	OR=1.51 p<1e-5
Clinical Data				
Body Mass Index (kg/m²)	31.5 ±0.24 (762)	31.7 ±0.19 (1570)	31.6 ±0.15 (2332)	p=0.538
Apnea Hypopnea Index (AHI)	12.8 ±0.56 (728)	12.5 ±0.41 (1506)	12.6 ±0.33 (2234)	p=0.625
AHI > 15 (%)	29.0% (728)	27.5% (1506)	28.0% (2234)	OR=1.08 p=0.461
PLMI	32.0 ±0.95	1.99 ±0.08	12.0 ±0.43	p<1e-Inf
RLS(A) (%)	21.20%	15.50%	17.40%	OR=1.46 p<1e-3
RLS(B) (%)	23.10%	16.20%	18.50%	OR=1.55 p<1e-4
RLS(AB) (%)	44.30%	31.70%	35.90%	OR=1.71 p<1e-8
RLS(C) (%)	40.30%	53.10%	48.80%	OR=0.60 p<1e-8
Polysomnogram				
TST (hour)	5.85 ±0.04	6.20 ±0.02	6.08 ±0.02	**p<1e-13**
Stage 1 (%)	11.5 ±0.24	10.1 ±0.17	10.6 ±0.14	p<1e-5
Stage 2 (%)	66.9 ±0.33	65.6 ±0.23	66.0 ±0.19	p=0.002
SWS (%)	6.33 ±0.25	7.54 ±0.20	7.13 ±0.16	p<1e-3
REM (%)	15.1 ±0.22	16.5 ±0.16	16.0 ±0.13	p<1e-6
WASO (hour)	1.45 ±0.03	1.13 ±0.02	1.24 ±0.02	p<1e-Inf
Medication				
RLS symptom aggravators	39.40%	35.30%	36.70%	OR=1.19 p=0.051
RLS symptom inhibitors	10.70%	9.20%	9.70%	OR=1.18 p=0.255

6.2.4 GENOTYPES

Thirteen SNPs from six genes are examined. SNPs from BTBD9 on chromosome six include rs935271, rs9296249, and rs392809. LOC643714 (chromosome 16) SNPs include rs3104767, rs3104774, and rs3104788. Two genes from chromosome 2 are

examined: MEIS1 (rs6710341, rs12469063, rs2300478), and "no gene" (rs6747972). Also examined are chromosome 15 gene MAP2K5 (rs649469) and chromosome 9 gene PTPRD (rs4626664 and rs1975197).

6.2.5 ANALYSIS

PSGs were processed for PLM using our described method that uses manually scored staging and respiratory event files in determining PLMI. The PLM detector was implemented in the SEV, a MATLAB toolbox for automating pattern recognition and biomarker classification algorithms in NPSG sleep cohorts[86, 117].

The OR of each SNP is determined for the RLS symptoms phenotype (i.e. RLS+/-), the PLM phenotypes (i.e. PLMI+/-), and combinations of both (e.g. PLMI+ and RLS+) using generalized estimating equations (GEEs) with repeated measures as implemented in GEEQBOX, a MATLAB toolbox[118]. Statistically significant differences between phenotypes and controls (see Table 1 and Table 2) are included as covariates in each model.

6.3 RESULTS

SNP OR's, calculated by risk allele frequency, are greater and more significant for PLM+ (Table 2) than RLS+ phenotype (Table 6.6).

6.3.1 RLS(+) PHENOTYPE

BTBD9 SNP rs392809 shows borderline significance for RLS+ in Table 6.6 (OR=1.24, p=0.051). Estimating the OR with repeated measure logistic regression (i.e. GEEs) and accounting for RLS inhibiting and RLS aggravating medication as covariates increases its significance to p=0.029; however, this association was rendered insignificant (p=0.053) after adjustment for PLMI. Table 6.7 gives the ORs and p-values of each SNP. None were statistically significant.

Table 6.6: Risk allele frequency for SNPs in RLS+ phenotypes.

	RLS(+) (n=380,o=860)	RLS(-) (n=535,o=1169)	RLS(+) vs RLS(-)
BTBD9(6)			
rs9357271(T)	0.796 (357)	0.765 (493)	OR=1.20 p=0.132
rs9296249(T)	0.792 (336)	0.761 (464)	OR=1.19 p=0.145
rs3923809(A)	0.721 (356)	0.680 (490)	OR=1.22 p=0.071
LOC643714(16)			
rs3104767(G)	0.586 (359)	0.569 (495)	OR=1.08 p=0.466
rs3104774(G)	0.588 (359)	0.575 (481)	OR=1.05 p=0.596
rs3104788(T)	0.588 (360)	0.575 (495)	OR=1.05 p=0.598
MEIS1(2)			
rs6710341(G)	0.132 (357)	0.148 (495)	OR=0.87 p=0.325
rs12469063(A)	0.772 (359)	0.750 (494)	OR=1.13 p=0.303
rs2300478(T)	0.767 (346)	0.753 (476)	OR=1.08 p=0.507
no gene(2)			
rs6747972(A)	0.460 (359)	0.436 (494)	OR=1.10 p=0.338
MAP2K5(15)			
rs6494696(G)	0.692 (356)	0.693 (492)	OR=1.00 p=0.976
PTPRD(9)			
rs4626664(A)	0.155 (358)	0.163 (489)	OR=0.95 p=0.675
rs1975197(A)	0.157 (356)	0.161 (492)	OR=0.98 p=0.856

Table 6.7 Genotype $\hat{\beta}$ values and odds ratios for RLS+ (n=860) phenotype versus controls (RLS-, n=1169) calculated by generalized estimating equations with PLMI and use of RLS effecting medications as covariates.

	$\hat{\beta}$	OR	p
BTBD9(6)			
rs9357271(T)	0.16	1.17 (0.97, 1.38)	0.132
rs9296249(T)	0.14	1.15 (0.95, 1.36)	0.171
rs3923809(A)	0.18	1.20 (1.01, 1.39)	0.061
LOC643714(16)			
rs3104767(G)	0.02	1.02 (0.85, 1.19)	0.819
rs3104774(G)	0	1.00 (0.83, 1.17)	0.981
rs3104788(T)	0	1.00 (0.83, 1.17)	0.973
MEIS1(2)			
rs6710341(G)	-0.12	0.89 (0.65, 1.14)	0.352
rs12469063(A)	0.16	1.17 (0.97, 1.37)	0.121
rs2300478(T)	0.12	1.13 (0.93, 1.33)	0.233
no gene(2)			
rs6747972(A)	0.07	1.07 (0.90, 1.25)	0.43
no gene/MAP2K5(15)			
rs6494696(G)	-0.01	0.99 (0.80, 1.19)	0.954
PTPRD(9)			
rs4626664(A)	-0.01	0.99 (0.77, 1.22)	0.964
rs1975197(A)	-0.06	0.94 (0.72, 1.17)	0.622

6.3.2 PLM(+) PHENOTYPE

Table 6.8 shows statistically significant ORs, by allelic frequency, for PLM+ in BTBD9 (rs3923809: OR=1.57, p<10⁻⁴), MEIS1 (rs12469063: OR=0.77, p=0.009), and PTPRD (rs1975197: OR=1.29, p=0.048). Table 6.9 shows SNP ORs for PLM+, calculated with GEEs, covarying for age, gender, RLS+, RLS-, total sleep time, and WASO duration. SNPs in BTBD9, MEIS1, and MAP2K5 remain statistically significant with lower p-values, and LOC643714 SNPs become statistically significant (e.g. rs104788: OR 1.35, p<10⁻⁴). Interestingly, the MEIS1 genotypes show opposite effect for the RLS+ and PLM+ phenotypes. For example, rs12469063 allele A shows increased odds of RLS+ (*not statistically significant*), but decreased, protective odds for PLM+.

Table 6.8: Risk allele frequency for SNPs in PLM+ phenotype and controls.

	PLMI>=15 (797 observations)	PLMI<15 (1597 observations)	All (2394 observations)
BTBD9(6)			
rs9357271(T)	0.796 (357)	0.765 (493)	OR=1.20 p=0.132
rs9296249(T)	0.792 (336)	0.761 (464)	OR=1.19 p=0.145
rs3923809(A)	0.721 (356)	0.680 (490)	OR=1.22 p=0.071
LOC643714(16)			
rs3104767(G)	0.586 (359)	0.569 (495)	OR=1.08 p=0.466
rs3104774(G)	0.588 (359)	0.575 (481)	OR=1.05 p=0.596
rs3104788(T)	0.588 (360)	0.575 (495)	OR=1.05 p=0.598
MEIS1(2)			
rs6710341(G)	0.132 (357)	0.148 (495)	OR=0.87 p=0.325
rs12469063(A)	0.772 (359)	0.750 (494)	OR=1.13 p=0.303
rs2300478(T)	0.767 (346)	0.753 (476)	OR=1.08 p=0.507
no gene(2)			
rs6747972(A)	0.460 (359)	0.436 (494)	OR=1.10 p=0.338
MAP2K5(15)			
rs6494696(G)	0.692 (356)	0.693 (492)	OR=1.00 p=0.976
PTPRD(9)			
rs4626664(A)	0.155 (358)	0.163 (489)	OR=0.95 p=0.675
rs1975197(A)	0.157 (356)	0.161 (492)	OR=0.98 p=0.856

Table 6.9: Genotype $\hat{\beta}$ values and odds ratios for PLM+ (n=797) phenotype versus controls (PLM-, n=1597) calculated by generalized estimating equations with gender, RLS+, RLS-, TST, and WASO duration as covariates.

	$\hat{\beta}$	OR	p
BTBD9(6)			
rs9357271(T)	0.36	1.43 (1.27, 1.59)	<1e-4
rs9296249(T)	0.39	1.47 (1.31, 1.64)	<1e-5
rs3923809(A)	0.43	1.54 (1.39, 1.69)	<1e-7
LOC643714(16)			
rs3104767(G)	0.26	1.30 (1.16, 1.43)	<1e-3
rs3104774(G)	0.28	1.33 (1.19, 1.46)	<1e-4
rs3104788(T)	0.3	1.34 (1.21, 1.48)	<1e-4
MEIS1(2)			
rs6710341(G)	0.04	1.04 (0.85, 1.23)	0.664
rs12469063(A)	-0.31	0.74 (0.58, 0.89)	<1e-4
rs2300478(T)	-0.26	0.77 (0.61, 0.92)	<1e-3
no gene(2)			
rs6747972(A)	0.03	1.03 (0.89, 1.16)	0.687
MAP2K5(15)			
rs6494696(G)	0.23	1.26 (1.10, 1.41)	0.003
PTPRD(9)			
rs4626664(A)	-0.03	0.97 (0.78, 1.15)	0.721
rs1975197(A)	0.29	1.34 (1.16, 1.51)	0.001

6.3.3 COMBINED PHENOTYPES

RLS+ and PLM+ phenotypes are combined into three phenotypes: (1) RLS+ with PLM+ (RLS+/PLM+), (2) RLS+ with PLM- (RLS+/PLM-), and (3) RLS- with PLM+ (RLS-/PLM+). Table 6.10 lists genotype ORs calculated with GEEs using RLS- with PLM- (RLS-/PLM-) as the control group. Covariates include age, gender, TST, WASO duration, use of RLS inhibiting medication, and RLS aggravating medication. BTBD9 and LOC643714 SNPs show increased OR for RLS+/PLM+, while MEIS1 SNPs show increased OR for RLS-/PLM+. BTBD9 SNP rs392809 shows increased risk for RLS-/PLM+ (OR=1.31, p=0.021) and also RLS+/PLM+ (OR=1.71, p<10^{-3}).

Finally, SNP ORs for PLM+/RLS+ vs. PLM-/RLS+ and for PLM+/RLS- vs. PLM- are estimated by GEEs, covering for age, gender, TST, WASO duration, and use of RLS effecting medications. Table 6.11 shows the results. Two statistically significant differences arise from these results that are not captured by Table 6.8 or Table 6.9. PTRD SNP rs1975197 is a significant risk for PLM+ given RLS+ (OR=1.50, p=0.006), as is MAP2K5's SNP rs6494696 (OR=1.32, p=0.032).

Table 6.10: Genotype $\hat{\beta}$ values and odds ratios for (A) PLM+ RLS+ (n=353), (B) RLS+ PLM- (n=507), and (C) RLS- PLM+ (n=321) versus controls (PLM- RLS-, n=848). Results are calculated using generalized estimating equations with age, gender, TST, WASO duration, and RLS effecting medication use as covariates.

	$\hat{\beta}$	OR	p
BTBD9(6)			
rs9357271(T)	0.57	1.77 (1.46, 2.08)	<1e-3
rs9296249(T)	0.57	1.76 (1.46, 2.07)	<1e-3
rs3923809(A)	0.53	1.70 (1.43, 1.97)	<1e-3
LOC643714(16)			
rs3104767(G)	0.29	1.33 (1.10, 1.57)	0.015
rs3104774(G)	0.3	1.35 (1.12, 1.58)	0.012
rs3104788(T)	0.28	1.32 (1.08, 1.55)	0.021
MEIS1(2)			
rs6710341(G)	-0.03	0.97 (0.64, 1.29)	0.836
rs12469063(A)	-0.07	0.93 (0.67, 1.20)	0.612
rs2300478(T)	-0.08	0.93 (0.66, 1.20)	0.581
no gene(2)			
rs6747972(A)	0.16	1.17 (0.93, 1.41)	0.192
MAP2K5(15)			
rs6494696(G)	0.09	1.09 (0.83, 1.35)	0.512
PTPRD(9)			
rs4626664(A)	-0.14	0.87 (0.55, 1.18)	0.378
rs1975197(A)	0.07	1.07 (0.77, 1.37)	0.655

(A) RLS+ PLM+ (n=353) vs RLS- PLM- (n=848)

	$\hat{\beta}$	OR	p
BTBD9(6)			
rs9357271(T)	0	1.00 (0.78, 1.23)	0.974
rs9296249(T)	0.01	1.01 (0.79, 1.24)	0.898
rs3923809(A)	0.1	1.11 (0.90, 1.32)	0.336
LOC643714(16)			
rs3104767(G)	-0.13	0.88 (0.69, 1.07)	0.194
rs3104774(G)	-0.15	0.86 (0.67, 1.05)	0.117
rs3104788(T)	-0.13	0.88 (0.69, 1.07)	0.172
MEIS1(2)			
rs6710341(G)	-0.14	0.87 (0.59, 1.15)	0.339
rs12469063(A)	0.12	1.12 (0.89, 1.35)	0.325
rs2300478(T)	0.08	1.08 (0.85, 1.31)	0.515
no gene(2)			
rs6747972(A)	0	1.00 (0.80, 1.20)	0.972
MAP2K5(15)			
rs6494696(G)	-0.06	0.94 (0.72, 1.16)	0.573
PTPRD(9)			
rs4626664(A)	-0.03	0.97 (0.71, 1.23)	0.816
rs1975197(A)	-0.19	0.83 (0.56, 1.10)	0.171

(B) RLS+ PLM- (n=507) vs RLS- PLM- (n=848)

	$\hat{\beta}$	OR	p
BTBD9(6)			
rs9357271(T)	0.04	1.04 (0.80, 1.27)	0.764
rs9296249(T)	0.15	1.16 (0.92, 1.39)	0.231
rs3923809(A)	0.28	1.32 (1.09, 1.54)	0.017
LOC643714(16)			
rs3104767(G)	0.07	1.07 (0.88, 1.27)	0.487
rs3104774(G)	0.11	1.11 (0.91, 1.31)	0.294
rs3104788(T)	0.13	1.14 (0.94, 1.34)	0.194
MEIS1(2)			
rs6710341(G)	0.01	1.01 (0.73, 1.29)	0.947
rs12469063(A)	-0.34	0.71 (0.49, 0.94)	**0.004**
rs2300478(T)	-0.3	0.74 (0.51, 0.98)	**0.013**
no gene(2)			
rs6747972(A)	-0.13	0.88 (0.67, 1.09)	0.221
MAP2K5(15)			
rs6494696(G)	0.04	1.04 (0.81, 1.27)	0.74
PTPRD(9)			
rs4626664(A)	-0.06	0.95 (0.68, 1.21)	0.687
rs1975197(A)	0.18	1.19 (0.93, 1.45)	0.18

(C) RLS- PLM+ (n=321) vs RLS- PLM- (n=848)

Table 6.11: Genotype $\hat{\beta}$ values and odds ratios for (A) PLM+ RLS+ (n=353) versus controls (PLM- RLS-, n=507) and (B) PLM+ RLS- (n=321) versus controls (PLM-RLS+, n=848). Results are calculated using generalized estimating equations with age, gender, TST, WASO duration, and RLS effecting medications as covariates.

	PLM(+) RLS+ (n=353) vs Controls (n=507)			PLM(+) in RLS- (n=321) vs Controls (n=848)		
	$\hat{\beta}$	OR	p	$\hat{\beta}$	OR	p
BTBD9(6)						
rs9357271(T)	0.68	1.97 (1.68, 2.25)	<1e-5	0.04	1.04 (0.80, 1.28)	0.763
rs9296249(T)	0.67	1.96 (1.67, 2.25)	<1e-5	0.15	1.16 (0.92, 1.39)	0.23
rs3923809(A)	0.5	1.64 (1.39, 1.89)	<1e-4	0.28	1.32 (1.09, 1.54)	0.017
LOC643714(16)						
rs3104767(G)	0.51	1.67 (1.45, 1.90)	<1e-5	0.07	1.07 (0.88, 1.27)	0.486
rs3104774(G)	0.53	1.70 (1.48, 1.92)	<1e-5	0.11	1.11 (0.91, 1.31)	0.294
rs3104788(T)	0.51	1.67 (1.44, 1.89)	<1e-5	0.13	1.14 (0.94, 1.34)	0.195
MEIS1(2)						
rs6710341(G)	0.08	1.08 (0.77, 1.39)	0.618	0.01	1.01 (0.73, 1.29)	0.947
rs12469063(A)	-0.15	0.86 (0.61, 1.10)	0.224	-0.34	0.71 (0.49, 0.94)	0.004
rs2300478(T)	-0.09	0.92 (0.67, 1.16)	0.487	-0.3	0.74 (0.51, 0.98)	0.013
no gene(2)						
rs6747972(A)	0.17	1.18 (0.96, 1.40)	0.138	-0.13	0.88 (0.67, 1.09)	0.226
MAP2K5(15)						
rs6494696(G)	0.27	1.32 (1.07, 1.57)	**0.032**	0.04	1.04 (0.81, 1.27)	0.741
PTPRD(9)						
rs4626664(A)	0.03	1.03 (0.73, 1.33)	0.834	-0.06	0.94 (0.67, 1.21)	0.675
rs1975197(A)	0.4	1.50 (1.21, 1.79)	**0.006**	0.18	1.19 (0.93, 1.45)	0.181

6.3.4 TREND ANALYSIS

Genotypes are evaluated for a linear trend of increasing PLMI by linear regression. Table 6.12(A) shows each SNP's effect coefficients for PLMI in the unadjusted model, while Table 6.12(B) shows the coefficients in the adjusted model which accounts for age, gender, use of RLS inhibiting medication, use of RLS aggravating medication, AHI, BMI, TST, WASO duration, RLS+, and RLS-.

BTBD9 and MEIS1 SNPs are significant before adjustment and increase in statistical significance after adjustment. LOC643714 SNPs are statistically significant after covariate adjustment. MEIS1 shows the greatest trend (rs21469063 $\hat{\beta}$ =-3.5, p<0.001), followed by BTBD9 (rs3923809 $\hat{\beta}$=2.85, p<0.001) and LOC643714 (rs3104767 $\hat{\beta}$=1.78, p=0.022).

Table 6.12: Genotype linear trend for PLMI calculated by regression (A) without adjustment, and (B) adjusting for age, gender, RLS effecting medication use, apnea hypopnea index, body mass index, TST, WASO duration, RLS+, and RLS-.

	(A) PLMI for All (n=2394), unadjusted model			(B) PLMI for All (n=2394) adjusted for covariates		
	$\hat{\beta}$	OR	p	$\hat{\beta}$	OR	p
BTBD9(6)						
rs9357271(T)	2.73	15.28 (13.25, 17.30)	0.008	2.92	18.52 (17.16, 19.89)	<1e-4
rs9296249(T)	2.38	10.80 (8.80, 12.80)	0.02	3.06	21.31 (19.97, 22.66)	<1e-5
rs3923809(A)	2.83	16.98 (15.13, 18.82)	0.003	3.01	20.19 (18.96, 21.43)	<1e-5
LOC643714(16)						
rs3104767(G)	1.16	3.18 (1.51, 4.85)	0.175	1.62	5.04 (3.92, 6.17)	0.005
rs3104774(G)	1.02	2.78 (1.09, 4.47)	0.235	1.67	5.30 (4.17, 6.44)	0.004
rs3104788(T)	1.17	3.23 (1.54, 4.91)	0.172	1.74	5.68 (4.55, 6.82)	0.003
MEIS1(2)						
rs6710341(G)	-0.35	0.71 (-1.69, 3.10)	0.776	-0.42	0.66 (-0.98, 2.30)	0.618
rs12469063(A)	-3.12	0.04 (-1.90, 1.99)	0.002	-3.17	0.04 (-1.28, 1.36)	<1e-5
rs2300478(T)	-2.9	0.05 (-1.92, 2.03)	0.004	-2.81	0.06 (-1.31, 1.43)	<1e-4
no gene(2)						
rs6747972(A)	0.69	2.00 (0.27, 3.72)	0.431	0.96	2.60 (1.42, 3.79)	0.113
MAP2K5(15)						
rs6494696(G)	1.31	3.71 (1.81, 5.61)	0.177	1.74	5.67 (4.39, 6.95)	0.008
PTPRD(9)						
rs4626664(A)	-1.44	0.24 (-2.08, 2.55)	0.221	-0.45	0.64 (-0.92, 2.20)	0.574
rs1975197(A)	0.7	2.02 (-0.24, 4.27)	0.542	1.73	5.65 (4.11, 7.20)	0.028

6.4 DISCUSSION

6.4.1 ALLELE MEASURES

The allelic OR and frequencies in RLS cases and controls for rs3923809 fall in line with those shown previously by Steffansen et. al [115], but do not reach the statistical significance necessary to state replication without model adjustment. Population attributable risk is also much lower in our sample than theirs (0.104 vs. 0.46).

The allelic frequencies and odds ratios for RLS symptoms are lower than those put forth by Winkelmann et. al for RLS [15]. And in the case of rs2300478, the risk allele is different (T instead of G). It may appear that these differences are due to the definition of *RLS symptoms*, which is less stringent than the criteria used by Winkelmann. However, the similarities between Winkelmann's SNP measures for RLS and ours for PLMI≥15 (Table 4) are striking. The OR and allele frequencies closely follow each other now for the two phenotypes (i.e. PLMI≥15 in our study and RLS in Winkelmann [15]), and the rs2300478 risk alleles match (i.e. G). PLMI is an accurate manifestation of several SNPs identified for RLS and will likely prove a more stable, and physiologically verifiable measure for RLS than current assessments (e.g. interview response). If we used PLMI≥15 as the definition of RLS, our assessments would replicate the genetic findings of Winkelmann et al [15] for SNPs in BTBD9, LOC643714, MEIS1, MAP2K5, and PTPRD.

6.4.2 SUMMARY

In summary, thirteen SNPs known to increase susceptibility to RLS were examined for susceptibility of PLM with and without RLS symptoms. PLM+ was defined as PLMI≥15 where PLMI was automatically scored during sleep and WASO using the PLM detector described in Chapter 5. RLS symptoms were determined by mailed survey response with questions that closely followed RLS diagnostic guidelines, though not specifying if symptoms worsen at night [104]. ORs calculated by risk allele frequency and GEE were examined in several comparisons of PLM+ and RLS+ while taking into account

214

statistically significant differences between phenotypes (e.g. age, gender, TST, WASO duration, and use of medication). SNPs in BTBD9, LOC643714, MEIS1, MAP2K5, and PTPRD were associated with heightened risk for PLMs. Several SNPs in MEIS1 were associated with heightened risk for PLMs without RLS symptoms, while SNPs in BTBD9, LOC64714, MAP2K5, and PTPRD9 were associated with higher risk for PLMs in the presence of RLS symptoms (i.e. RLS+). A linear increase of PLMI by risk allele is observed in these SNPs as well. A publication discussing these findings is currently in revision.

CHAPTER 7   SLEEP IN COMBAT VETERANS WITH POSTTRAUMATIC STRESS DISORDER

This chapter investigates two aspects of sleep in a cohort of Vietnam era combat veterans diagnosed with and without posttraumatic stress disorder. Rapid eye movements are investigated first and a new technique for measuring and characterizing ocular position is introduced. Second, inter-hemispheric EEG coherence is measured during a sleep and a significant reduction is seen in combat veterans with PTSD.

Steve Woodward was instrumental in getting this research going and providing exceptional help with the dataset and having the vision for investigating EEG coherence in this cohort. Bernard Widrow was very helpful in talking through many of the questions relating to rapid eye movements and cognitive activity during sleep in subjects with PTSD. Also, Emmanuel Mignot was very supportive in letting me pursue this line of research under such guidance.

7.1   INTRODUCTION

There has been limited characterization of *phasic* eye movement activity during rapid eye movement (REM) sleep in patients diagnosed with posttraumatic stress disorder. Sleep disturbances are a core component of post-traumatic stress disorder (PTSD), a syndrome that develops when a person is unable or fails to recover from the stress induced by a specific traumatic event [17, 119]. Common symptoms of PTSD, such as nightmares, insomnia, and hyper-vigilance or hyper-arousal can aggravate sleep problems, which may in turn perpetuate PTSD [120-122].

Emotional memory consolidation [123], as well as dreams and nightmares [124, 125], are thought to occur most during REMS. Inability to recover from trauma-induced stress has been viewed as a failure to learn from and re-integrate trauma-related events during

sleep[120, 126-128]. And, without positive intervention, emotional memory attached to the traumatic event may be maintained rather than stored during sleep [123].

Nocturnal polysomnography (NPSG) based-sleep studies involving PTSD have historically focused either on global power spectral analysis of the EEG or on sleep architecture changes (e.g. time spent in REM sleep versus non-REM) [30]. Despite numerous published and validated algorithms for quantifying phasic eye movement activity [69-73], the sleep community has been slow to incorporate such measures in their investigations of PTSD.

This chapter investigates phasic eye movements (EM) and hemispheric correlation in the EEG of sleep in Vietnam era combat veterans diagnosed with PTSD and matched controls. Chapter 1 describes this *Combat Veteran Cohort (*CVC) and its collection procedure in detail. Background information on PTSD, and different electrooculogram (EOG) montages are mentioned in section 7.2. Section 7.3 presents several rapid eye movement (REM) detectors that are optimized for analysis in the data set using a physiologically based criteria. A new approach to measuring EMs which takes advantage of the EOG montage used in this dataset is given in Section 7.4. Section 7.4 shows, for the first time, the hemispheric breakdown in sleep EEG of combat veterans diagnosed with PTSD. Alternating hemispheric EEG activity is found in some prey animals, and is found in combat veterans as well when examining the magnitude squared coherency of left and right EEG electrodes. Concluding remarks are given in Section 7.7.

7.2 BACKGROUND

This section provides background information on PTSD and motivation for analysis of REM and inter-hemispheric EEG coherency.

7.2.1 POSTTRAUMATIC STRESS DISORDER

Sleep disturbances are a core component of posttraumatic stress disorder (PTSD), a syndrome that develops when a person is unable or fails to recover from a traumatic

experience as discussed in Section 1.3.3. Common symptoms of PTSD, such as nightmares, insomnia, and hyper-vigilance or hyper-arousal can aggravate and impose sleep problems, which may in turn perpetuate PTSD. NPSG based-sleep studies involving PTSD have historically focused on global power spectral analysis of the EEG or on sleep architecture changes.

7.2.2 EYE MOVEMENTS IN SLEEP

Aserinsky and Kleitman discovered groupings of rapid eye movements with wake-like electroencephalograhy (EEG) readings and large heart rate variations in sleep [129, 130], which led to sleep's two state division: rapid eye movement sleep (REMS) and non-rapid eye movement sleep (NREMS). REMS may be further divided based on ocular activity. Phasic activity in REMS consists of REM, muscle twitches, and increased heart rate variability. These behaviors are absent in tonic REMS [69]. Tonic activity is described in terms of its timing and duration of REM sleep, while *activity* or total counts, mean duration, incidence, and rotation of EMs is useful for describing phasic activity [71].

7.2.3 ELECTROOCULAGRAPHY

EMs are measured by electrooculagraphy (EOG) during nocturnal polysomnography (NPSG) sleep studies. The EOG models the eye as a dipole with positive polarity at the pupil and negative polarity at the retina. Placing electrodes near the eye allow changes in pupil position to be measured as changes in voltage with respect to a distal cephalic reference.

Electrooculagraphy montages
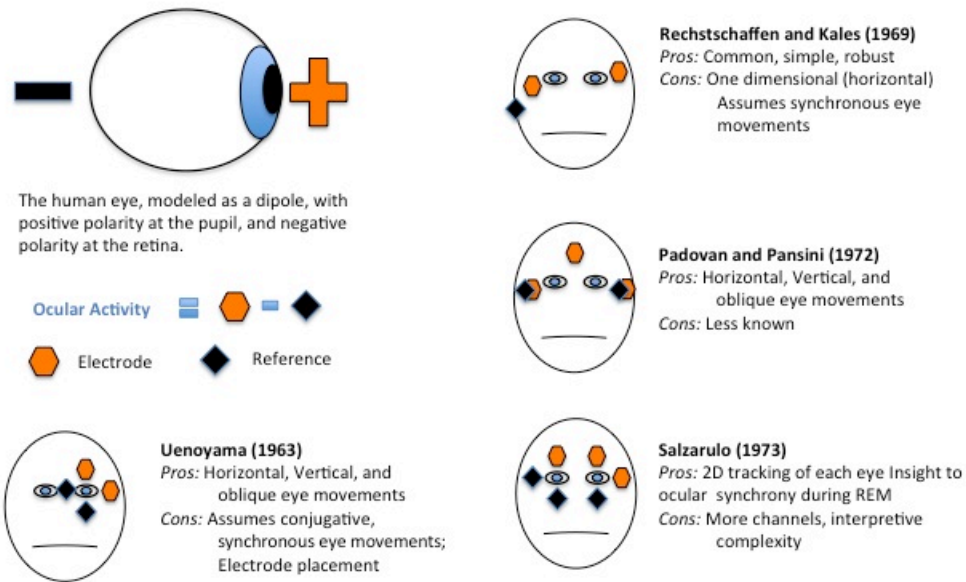
Figure 7.1: Electrooculagraphy montages for measuring eye movements.

Figure 7.1 shows several EOG montages used in NPSG. A two channel, EOG montage, following that of Uenoyama's VEOG configuration [131], was used on the left eye as shown in Figure 7.2. The *Woodward* configuration allows detection of horizontal, vertical, and oblique EMs by treating the horizontal and vertical input channel voltages as X and Y vectors from which ocular direction can be determined. The setup provides greater fidelity than the traditional montage setup proposed by Rechtschaffen and Kales [1], but less than the double eye, double channel montage proposed by Schneider [132, 133] which takes into account the possibility of non-conjugate EMs during REMS [134, 135].

Woodward montage
Pros: (1) Measures horizontal, vertical, and oblique eye movements.
(2) Horizontally robust to EEG contamination.
Cons: Conjugative synchrony assumed

Tonic REM – no eye movements

Vertical, slow eye movements

Horizontal eye movement

Oblique, rapid eye movement

Phasic REM with possible non-conjugative eye movements
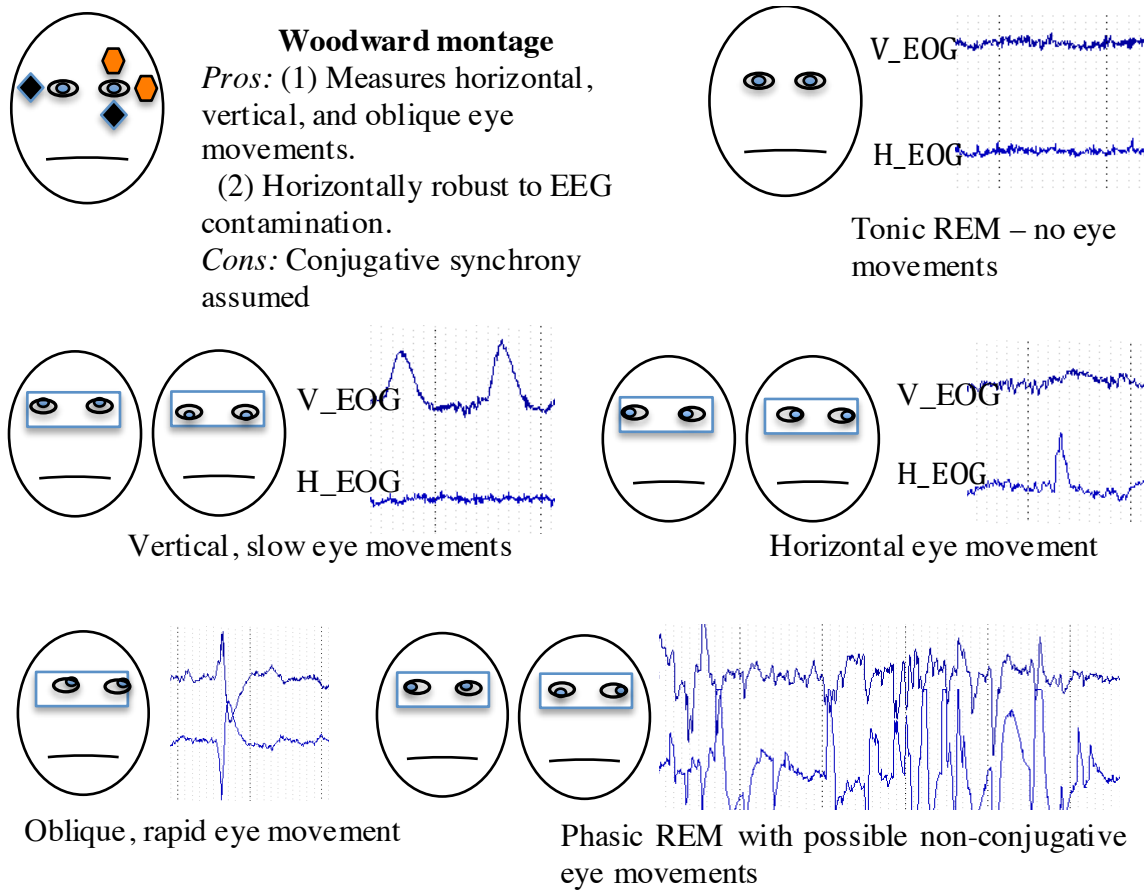
Figure 7.2: Woodward EOG montage. Ocular position can be estimated from the vertical and horizontal EOG signal data as shown here.

7.3 PHASIC RAPID EYE MOVEMENT DETECTORS

REMS abnormalities are considered a core component of PTSD and likely reflect a patient's hindered ability to process traumatic memory. Four previously published, and three unpublished eye movement detection algorithms are considered for phasic REMS analysis here. Published eye movement detection methods are from Gopal and Haddad [69], Takahashi and Atsumi [71], McPartland and Kupfer [70], and Tan, Campbell, and Feinberg [72]. These methods fall into the categories of power spectral analysis (*Tan*), slope and amplitude detection (*Takahashi*, *Gopal*), and ocular synchrony (*McPartland*). Two new methods, *dual threshold*, and *simple slope* are developed as baseline measures

of amplitude and slope base detection methods. A third method, *quadrant detection*, takes advantage of the Woodward montage and combines both vertical and horizontal EOG inputs to determine ocular activity in a two dimensional plane.

7.3.1 POWER SPECTRAL ANALYSIS

Tan implements a power spectral method for detecting eye movement potentials using a single channel EOG with the lead placed on the left outer canthus and reference placed mid-forehead. Spectral analysis by Welch averaging was performed on 5.120s blocks at 2.620 s increments. The majority of spectral power of EM activity during REM was found in the 0.3 - 2.0 Hz frequency range [72].

7.3.2 AMPLITUDE

Detector Dual Threshold applies a lower and upper threshold to the EOG channel followed by a duration criteria to determine eye movement. Voltage must exceed an upper threshold of 30 μV and then fall below 10 μV within 0.1 seconds for detection to occur. Detections less than 0.05 seconds apart are combined.

7.3.3 SLOPE AND AMPLITUDE

Detector Simple Slope uses a four-point, first derivative slope of the EOG signal to obtain peaks of interest. Peaks whose raw data exceeds 30 μV are classified as eye movement.

Gopal determines a series of decision points based on local maxima and minima and then applies a series of rules – slope threshold and amplitude threshold – to determine presence of "eye movement waveforms (EMW)". The method was applied to the EOGs of babies with leads placed on the outer canthus of each eye and referenced to the forehead [69]. The optimal thresholds discovered by Gopal were adjusted to more appropriately fit our adult dataset. A 25 μV/sec threshold was set for the slope and a 30 μV threshold for amplitude. Their low-pass frequency cutoff of 6.5 Hz was changed to 8.0 Hz to match the other detection algorithms.

Takahashi applies three rules to first and second order derivatives of smoothed EOG data to determine EMs. Smoothing is done via a seven point moving average filter, and EMs, recognized by abrupt changes in slope, are detected if the signal amplitude exceeds 30 μV, the slope exceeds 248.3 μV/second, and the duration is less than 0.5 seconds.

7.3.4 SIGNAL ESTIMATION AND DENOISING

The eye movement detection algorithms described above were implemented in the SEV and visually spot-checked for correctness. The previously published methods performance appeared much poorer in our dataset than those described the authors. False positive detection of EMs were frequently seen in NREM sleep due to interfering EEG activity, such as K-complexes (waveforms that transition sharply once between a negative and positive deflection), or transient noise picked up in the EOG channel. Similar to the PLM detectors tested in Chapter 5, the EM detectors here were biased by the specific patient data sets and montage configurations they were developed for and not robust to our dataset. The EM detectors appeared logically sound given clean data, so I resolved to clean the data and use the distribution of detected EMs per hour in REM and NREM sleep as a measure of performance. A data cleaning method that results in higher REM to NREM detections per hour sleep is more suitable than one without or with a smaller REM/NREM classification ratio. Wavelet denoising and adaptive filtering are considered for preparatory EOG cleaning.

7.3.4.1 Adaptive Filtering

The least mean squares (LMS) algorithm adaptively updates filter weights to minimize the mean square error between two signal inputs. Let $s(n)$ be defined as

$$s(n) = x(n) + \eta(n)$$

where $x(n)$ is the true signal of interest and $\eta(n)$ is the unwanted noise. Given a reference signal $r(n)$ that is correlated with $\eta(n)$ but independent of $x(n)$ an adaptive filter may be constructed to remove the unwanted noise $\eta(n)$ from the signal $s(n)$ by

222

minimizing the mean square error. The block diagram in Figure 7.3 shows the adaptive filter with the vertical EOG channel as the signal with noise $(s(n))$ and the EEG channel as the correlated noise reference signal $(r(n))$. The error, $e(n)$, is the difference between the primary input signal and the adaptively filtered reference signal output $y(n)$

$$y(n) = \sum_{m=1}^{M} h(m) \cdot r(n+1-m)$$

where h is an M length vector of filter coefficients. Mathematically, the error is

$$e(n) = s(n) - y(n)$$
$$= x(n) + \eta(n) - y(n)$$

Dropping the (n) sample notation, the expected value of the error squared is

$$E[e^2] = E[(x+\eta-y)^2]$$
$$= E[x^2 + 2x \cdot \eta - 2 \cdot \eta \cdot y + \eta^2 - 2y \cdot x + y^2]$$
$$= E[x^2] + 2E[x \cdot \eta] - 2E[\eta \cdot y] + E[\eta^2] - 2E[y \cdot x] + E[y^2]$$

Assuming x to be a zero-mean, stationary random variable, uncorrelated with η the mean-squared-error (MSE) can be reduced to

$$E[e^2] = \sigma_x^2 + 2E[x] \cdot E[\eta] - 2E[\eta \cdot y] + \sigma_\eta^2 - 2E[y] \cdot E[x] + \sigma_y^2$$
$$= \sigma_x^2 - 2E[\eta \cdot y] + \sigma_\eta^2 + \sigma_y^2$$
$$= \sigma_x^2 + E[(\eta-y)^2]$$

Minimizing the mean-squared error involves minimizing $\sigma_x^2 + E[(\eta-y)^2]$. The adaptive filter has no effect on the signal $x(n)$, and determining the least mean squared (LMS) error is found by minimizing $E[(\eta-y)^2]$. Adjusting the filter coefficients to bring y toward η.

Here, the EOG channel is the first signal and consists of true eye movement and noise and the EEG channel is the reference signal that correlates with *noise* in the EOG. The recursive least squares (RLS) implementation presented by He [76] to remove unwanted EOG signal from EEG signal is used here with the opposing purpose of removing unwanted EEG from EOG data by simply swapping the input and reference channel in the adaptive algorithm. Figure 7.3 shows a K-complex from the EEG removed from the vertical EOG channel by adaptive RLS filtering.
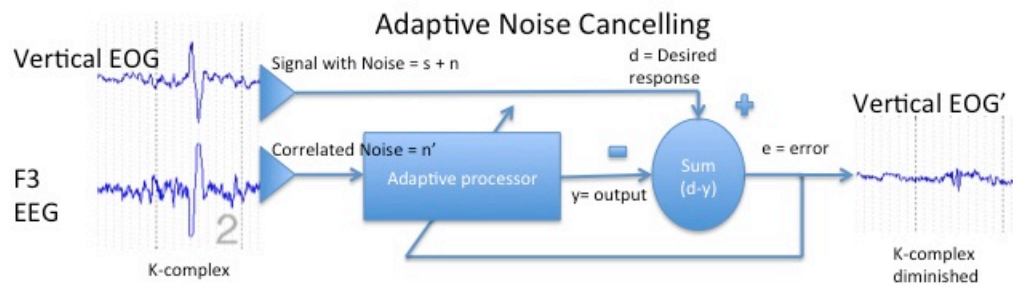
Figure 7.3: Adaptive noise cancelling of EEG interference in the EOG channel.

7.3.4.2 Wavelet Denoising

Wavelets are used to de-noise a signal by decomposing the signal into low and high frequency components using a specified kernel. The components, details and approximate, represent different frequency bands of the signal that may individually cleaned by zeroing unwanted parts (e.g. places where the components fall below some threshold) and then resynthesizing with cleaned (zeroed) components.

The SEV's wavelet denoising function, *denoise_wavelet*, allows users to specify the number of wavelet decompositions to use, set the threshold *T*, and determine whether *soft-thresholding* should be used. I used five decomposition levels (four details and one approximation) and selected soft-thresholding at 50 μV. The SEV function pads the input signal data, if necessary, so its final length N is evenly divisible by 2^J where J is the number of decomposition levels. MATLAB's *swt* function performs the stationary

224

wavelet decomposition of the padded input data and returns *W*, a *JxN* matrix of wavelet coefficients. The *denoise_wavelet* method then goes through the first *J-1* levels (i.e. the detail coefficients) and compares each sample to the threshold and changes the value according to the soft-threshold selection. *W(j,n)*, the nth wavelet coefficient of detail level *j*, is updated as

$$W(j,n) = \begin{cases} 0 & \text{hard thresholding} \\ \text{sgn}(W(j,n)) \cdot (|W(j,n)| - T) & \text{soft thresholding} \end{cases}$$

wherever *T* is exceeded. MATLAB's *iswt* function resynthesizes the denoised coefficients into a single *N* length vector, which *denoise_*wavelet crops to the original input signal length as necessary. The *Haar* kernel is used for wavelet decomposition and synthesis. Further information on wavelet de-noising can be found in [136] and [137].
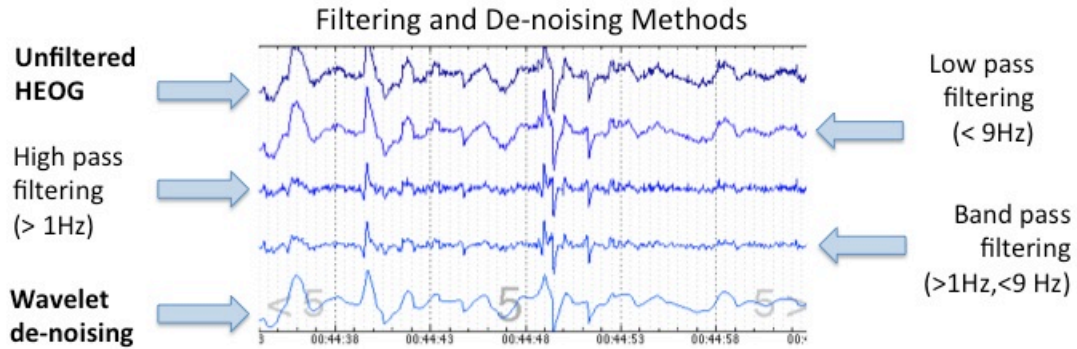
Figure 7.4: Comparison of finite impulse response filtering and wavelet denoising for cleaning EOG signal.

Figure 7.4 shows wavelet denoising as a suitable option compared to standard low, high, and band pass finite-impulse-response filtering techniques. Decomposing the EOG signal by wavelet transform, thresholding its detail components, and finally synthesizing the selected components together produces a smoothed version of the original signal that is less susceptible to false detections due to noise.

225

7.3.5 PHYSIOLOGICAL BASED PATTERN RECOGNITION OPTIMIZATION

One of the most challenging aspects of machine learning is obtaining a gold standard data set to train with. Having an expert to identify different categories as we did for PLMs in the previous chapters is often a rare occurrence or an expensive undertaking. A popular alternative to this currently is to outsource the task to a general population at a smaller cost as recently done with spindle detections using Amazon's Mechanical Turks (AMT) service [138]. While this is often a good and acceptable solution, a considerable amount of *overhead* is required – setting up the service, creating instructions for the users, obtaining the training dataset, and cleaning the results to remove errors caused by poor or nefarious performers. Various strategies exist for removing *bad results* from the training sample, and many utilize information from the other scorers to determine some measure of collective reliability. However, rather than pursue these lines of investigation, we consider a physiological based approach that takes advantage of the information we have.

REMs, the namesake feature of REMS, are required for identifying and scoring REMS by R&K criteria [1]; the criteria used to score the CVC. While we do not have a gold standard of the exact positions or number of EMs in each study, we can infer their distribution from the record of manual scored sleep stages. Because REM occur predominantly during REMS (and also wake) a REM detector's performance can be roughly gauged from the distribution of its detections across manually scored sleep stages. A random, naïve detector will have a flat distribution with events scored equally across all sleep stages. A detector that is better than random will favor REMS over Stage 1, Stage 2, and Slow Wave Sleep. This approach is particularly helpful in determining which preprocessing steps are best for each REM detector.

Two ratios, the REM:NREM activity ratio and the REM:NREM density ratio, are used to quantify distribution performance of the adaptive noise cancellation and wavelet denoising preprocessing techniques as applied to each REM detector.

7.3.5.1 REM:NREM activity ratio

The REM:NREM activity ratio is the total number of events detected in REM divided by the total number of events detected in NREM sleep (NREMS). It is defined as

$$\text{REM:REM activity ratio} = \frac{\text{REM activity}}{\text{NREM activity}} = \frac{\sum_{k=0}^{K-1} stage(v[k]) \in \{5\}}{\sum_{k=0}^{K-1} stage(v[k]) \in \{1,2,3,4\}}$$

for K detections with discrete epoch start locations stored in the K element vector v. In practice, the vector v is obtained using a sample to epoch conversion formula based on the underlying signal's sample rate and the defined epoch length (e.g. 30 s). The function $s=stage(n)$ returns the manually scored sleep stage s for the input epoch n. Possible values for s include 0 for wake, $1,2,3$ or 4 for NREM, 5 for REM, and 7 for unknown cases. In the above, and following expressions, $stage(y) \in S$ is a Boolean operation which is true, and given a value of 1, if $stage(y)$ is in the set S. Otherwise the expression is false and returns 0. The summation of Boolean values is done using the decimal, not binary, system (i.e. true + true + false = 1 + 1 + 0 = 2).

7.3.5.2 REM:NREM density ratio

The REM:NREM density ratio is the ratio of the detection density in REMS to the detection density in NREM sleep. The density ratio accounts for unequal distributions of REM and NREM sleep during a study and is the activity ratio normalized by the duration of REM and NREM sleep measured in epochs.

$$\text{REM:NREM density} = \frac{\text{REM density}}{\text{NREM density}} = \frac{\text{REM activity} \Big/ \sum_{n=0}^{N-1} stage(n) \in [5]}{\text{NREM activity} \Big/ \sum_{n=0}^{N-1} stage(n) \in [1,2,3,4]}$$

7.3.6 OPTIMIZATION RESULTS

Table 7.1 and Table 7.2 show the REM:NREM density ratios for each EM detector preprocessing configuration as applied to the horizontal and vertical EOG channels respectively. The Tan filter does not improve with preprocessing of either HEOG or VEOG signals. The dual threshold method improves with wavelet denoising of the VEOG channel, but does not improve with preprocessing of the HEOG channel. Takahashi's detector improves with adaptive filtering of the VEOG channel, and wavelet denoising of the HEOG. The remaining detectors improve with wavelet denoising of the VEOG and HEOG channels.

Gopal's EM detector shows the highest REM:NREM density for both EOG channels with wavelet denoising, followed by the simple slope and amplitude method, then the dual threshold, and lastly Tan's detector. The Tan detector shows the lowest (i.e. worst) REM:NREM densities for every EOG configuration evaluated and does not warrant further investigation here. The remaining methods are still hindered by their separate application to each EOG channel (i.e. vertical and horizontal), which leaves the question of combining EMs. For example, if a horizontal EM is detected shortly after, and also during a vertical EM detection, should it be counted as a single EM, or two separate EMs? I wanted to address this problem in a way that combines the EOG channels so that EMs can be detected directly and not require a post-combination synthesis of EMs detected from the separate channels. The approach taken is discussed in the next section.

Table 7.1: Horizontal EOG REM:NREM density ratios for EM classification methods under preprocessing configurations: none, wavelet denoising, adaptive filtering of EEG (F-4 position), and both wavelet denoising and adaptive filtering. Results are shown as the mean, plus or minus the SEM, and number of subjects in parenthesis.

	None	Wavelet denoising	Adaptive filtering of F-4 EEG	Wavelet denoising with Adaptive filtering
Tan	2.41 ±0.14 (120)	1.08 ±0.02 (120)	2.34 ±0.15 (120)	1.07 ±0.02 (120)
Simple Slope + Amplitude	16.6 ±2.60 (119)	19.9 ±3.9 (113)	13.3 ±2.2 (118)	18.1 ±4.6 (115)
Gopal	22.3 ±2.9 (118)	25.0 ±3.4 (114)	18.7 ±2.8 (119)	18.0 ±2.7 (113)
Takahashi	3.01 ±0.29 (120)	11.8 ±1.6 (116)	4.24 ±0.37 (120)	11.3 ±1.5 (118)
Dual Threshold	15.1 ±2.6 (119)	14.2 ±1.9 (115)	13.7 ±2.0 (119)	11.9 ±1.9 (115)

Table 7.2: Vertical EOG REM:NREM density ratios for EM classification methods under preprocessing configurations: none, wavelet denoising, adaptive filtering of EEG (F-4 position), and both wavelet denoising and adaptive filtering. Results are shown as the mean, plus or minus the SEM, and number of subjects in parenthesis.

	None	Wavelet denoising	Adaptive filtering of F-4 EEG	Wavelet denoising with adaptive filtering
Tan	1.61 ±0.09 (119)	1.03 ±0.01 (120)	1.21 ±0.05 (119)	1.03 ±0.01 (120)
Simple Slope + Amplitude	7.38 ±1.01 (119)	7.60 ±1.20 (117)	4.96 ±0.72 (119)	4.48 ±1.19 (114)
Gopal	8.59 ±0.97 (119)	9.34 ±1.22 (117)	7.00 ±1.32 (118)	6.38 ±1.36 (116)
Takahashi	2.05 ±0.18 (119)	4.90 ±0.60 (119)	5.19 ±0.44 (119)	4.45 ±0.53 (119)
Dual Threshold	6.36 ±0.72 (118)	6.72 ±0.81 (118)	4.48 ±0.85 (119)	4.37 ±1.57 (118)

7.4 QUADRANT DETECTION

The *quadrant detector* determines ocular position directly from the vertical and horizontal EOG channels, which allows for separate EM detections from each channels. Figure 7.2 shows estimates of pupil position and corresponding HEOG and VEOG signal segments. Placing HEOG samples on the x-axis and VEOG values on the y-axis, we can map the pupil's position for each instance of time $n \in [0, N-1]$ where N is the total

229

number of samples. Using the HEOG and VEOG channel values as x and y coordinates of a Cartesian system we estimate the pupil's position directly for each digital sample in time. We can use Euler's formula

$$re^{j\theta} = r\cos(\theta) + jr\sin(\theta)$$

to describe the pupil's position in Polar coordinates where $j = \sqrt{-1}$, radius r defined as

$$r = \sqrt{x^2 + y^2} \,,$$

and the angle θ,

$$\theta = \arctan\left(\frac{y}{x}\right),$$

which is obtained by the trigonometry relations

$$x = r\cos(\theta) \text{ and } y = r\sin(\theta).$$

Tracking the pupil in Polar coordinates, while enticing, is only a remapping of the data taken directly from Cartesian coordinates and does not reduce the data. There is still too much data to examine given the number of subjects with multiple studies each. Furthermore, there is no foundation yet, from other research, on how to interpret angular patterns of EMs in sleep. To this point, evaluating EMs in sleep has primarily focused on identifying their occurrence and evaluating derived metrics (e.g. count, activity, density, etc.). As a compromise to simply detecting eye movement occurrence and flooding myself with every pupil position and orientation through the study, I introduce the quadrant detector.

The quadrant detector classifies EM based on magnitude and duration of ocular deflection. The Pythagorean theorem defines the deflection magnitude r at sample index n as

$$r_n = \sqrt{x_n{}^2 + y_n{}^2} \ ,$$

which is combined with the EM classification rule

$$r_{n,n+\Delta} \geq T, \Delta \in \{D_{\min},..,D_{\max}\}$$

for amplitude threshold T and duration threshold Δ. The classification rule sets T to 30 μV, which follows amplitude thresholds of previously published EM detectors, and $D_{\min}$ to one sample and $D_{\max}$ to N. The quadrant detector quantizes and stores the pupil angle into one of four quadrants

$$Q_1 = x > 0 \cap y > 0$$
$$Q_2 = x < 0 \cap y > 0$$
$$Q_3 = x < 0 \cap y < 0$$
$$Q_4 = x > 0 \cap y < 0$$

when an EM is classified. The quadrant detector can be used simply to classify any radial EM, as opposed to horizontal and vertical EMs, by determining when $r \geq T$, while simultaneously providing flexibility for state space analysis of detected quadrants.

7.4.1 QUADRANT DETECTOR OPTIMIZATION

I obtain the quadrant detector's optimal preprocessing configuration using the REM:NREM density ratio described in Section 7.3.6. Table 7.3 presents this ratio for the quadrant detector after both EOG channels are preprocessed with nothing, wavelet denoising, adaptive filtering of the EEG (F-4 electrode position), or adaptively filtering the EEG and then wavelet denoising.

Adaptively removing F-4 EEG interference from both HEOG and VEOG channels and then applying wavelet denoising gives the highest REM:NREM density. I also examined the performance of preprocessing the HEOG and VEOG channels differently, but none were better than the combined adaptive filtering and wavelet denoising result. For

example, wavelet denoising the HEOG channel and both adaptively filtering and wavelet denoising the VEOG channel gives the next highest REM:NREM density mean of 3.28 ±0.26 for 120 subjects.

Table 7.3: Quadrant detector REM:NREM density ratios without preprocessing (none), and with HEOG and VEOG channel preprocessing. The preprocessing methods shown are applied to both channels and the mean results given with SEM and number of subjects tested in parenthesis.

	None	Wavelet denoising	Adaptive filtering of EEG	Wavelet Denoising+Adaptive filtering of EEG
Quadrant detector	1.14 ±0.06 (120)	2.00 ±0.14 (120)	2.66 ±0.17 (120)	3.88 ±0.32 (120)

7.5   EXPLORING EYE MOVEMENTS IN SLEEP OF PTSD

This section first introduces three ways EM metrics can be clustered or grouped across a sleep study, and then describes a new visual analytics software program for exploring phasic events and complements the SEV's batch mode processing.

7.5.1 EYE MOVEMENT CLUSTERING IN SLEEP

Sleep research has seen limited analysis of phasic events in context of macro level sleep architecture changes. Clustering EMs in sleep by stage, cycle, or elapsed time gives greater insight into physiology around EMs and differences between combat veterans with and without PTSD. These measures, described next, are included as meta-data with each EM detected as output by the SEV.

7.5.1.1 Sleep stages

Human sleep is characterized in terms of REM and NREM sleep (with NREM sleep broken further into Stage 1, Stage 2, and Stages 3 and 4).   The CVC was manually scored for sleep stage in 30 sec epochs according to R&K guidelines.

7.5.1.2 NREM/REM cycles

Healthy, nocturnal sleep cycles from NREM sleep to REMS and back again with NREM sleep occurring in greater length at the beginning of the night and lessening toward the morning, while REMS occurs in shorter duration at first and increases with ensuing cycles. Currently, there is no standard method for measuring NREM/REM sleep cycles, so I developed an algorithm to do so. The sleep cycle classification algorithm takes a vector of consecutive sleep stage scores, by epoch, as input and returns a same size vector with the NREM/REM cycle corresponding to each. The algorithm models the transition between NREM/REM cycles by assigning different *weights* for wake, NREM and REM epochs and determines the point at which enough NREM and REM *pressure* has been alleviated in one cycle and the next cycle begins (see Figure 7.5 for MATLAB code).

7.5.1.3 Time elapsed

Diurnal and circadian effects are often observed in sleep. Clustering biomarker observations into hourly blocks allows evaluation of the changes directly by elapsed sleep, apart from explicit sleep cycle changes. This can be done easily with the SEV, which attaches the sleep stage, epoch, and elapsed time to each detected event.

```
REM_weight = 0;
lastREM_index = [];
curCycle = 0;
start_index = find(stages~=7&stages~=0,1);
stop_index = numStages;
if(start_index>1)
    sleepCycles(1:start_index)=curCycle;
end
curCycle=1;
wake_delta = 10;
nrem_delta = 25;
rem_delta = 300;
cycle_start_index = start_index;
for s=start_index:stop_index
    if(stages(s) == 5)
        lastREM_index = s;
        REM_weight = rem_delta;
    else
        if(~isempty(lastREM_index))
            if(stages(s)==0||stages(s)==7)
                REM_weight = REM_weight - wake_delta;
            else
                REM_weight = REM_weight - nrem_delta;
            end
            if(REM_weight < 0)
                sleepCycles(cycle_start_index:lastREM_index) …
                                = curCycle;
                cycle_start_index = lastREM_index+1;
                lastREM_index = [];
                REM_weight = 0;
                curCycle = curCycle+1;
            end
        end
    end
end

if(sleepCycles(s)~=curCycle)
    sleepCycles(cycle_start_index:s) = curCycle;
end
```

Figure 7.5: NREM/REM sleep cycle classification algorithm in MATLAB. The algorithm takes a vector of consecutive sleep epochs scored by stage and returns a same length vector representing the NREM/REM sleep cycle of the corresponding epoch index.

7.5.1 BUILDING A DATA MOUNTAIN FOR MINING

The large amount of data and equally large number of ways it can be clustered make analyzing and exploring it a difficult task by conventional means such as direct spreadsheet viewing in a program like Excel or repetitive scripting (where slight changes are used to generate new plots) in R or MATLAB. The CVC consists of four consecutive sleep studies from approximately 100 combat veterans who are diagnosed with PTSD, PTSD and MDD, or with neither PTSD or MDD. EMs are detected on two separate EOG channels and represent vertical and horizontal activity, and also by the quadrant detector which detects EMs and also quadrants from the combined EOG channels. Furthermore, all six detectors have been evaluated with four preprocessing configurations.

Classifying EMs for each patient, visit, detector, EOG channel, and preprocessing option is straightforward in the SEV. However, the work is not done. The data must still be evaluated. This situation does not fall cleanly into the domain of data mining. Processing the EOGs and classifying EMs by various detectors is more akin to building the mountain than mining it. Because I am the one building the mountain of data, I have flexibility in its design, which I use to make the mining easier. Specifically, I create a MySQL database to store the SEV's output and the CVC's demographic information. This gives a flexible framework for investigating the data in parts (e.g. EMs from quadrant detector by elapsed hour), alleviates the tedious task of managing output data storage directly (e.g. creating folders of specifically named flat files), and reduces load time required to place everything into working memory.

7.5.2 VISIT VIEW – VISUAL ANALYTICS SOFTWARE FOR CONSECUTIVE SLEEP STUDIES

Visit View is the visual analytics software program I developed to CVC biomarkers as output by the SEV. Initially, I parsed the SEV's EM classifications output and associated metadata (e.g. elapsed time of occurrence, sleep stage, etc.) to produce EM metrics (e.g. activity, density, mean duration, etc.) and store as one large, consolidated file. I used several smaller scripts to parse this file for metrics of interest (e.g. NREM EM density)

and then plot the results for each visit of each patient group (e.g. with and without PTSD) by sleep cycle or elapsed time and categorized by NREM, REM, or all sleep. The number of EM configurations and metrics to examine made this a tedious, time consuming task, and I found it easier and more productive to write a graphical user interface (GUI) to explore the data. Instead of updating and rerunning new scripts, I tied the possible query options to various control widgets (e.g. drop down menus) that the program populates. At first, the software parsed the single flat file with stored EM metrics for the data and to populate the GUI's controls, but this required the file to be rebuilt from the database whenever the SEV was used to produce new biomarkers. The single file option was removed and Visit View's design changed to query the database directly for data and produce metrics on the fly and keep solutions in working memory so that different metrics derived from the same classification method selected (e.g. quadrant detector) could be selected quickly without repeated queries of the source biomarker events (e.g. mean EM activity and mean EM duration can be produced using the same database query for EM events).

A screenshot of Visit View is shown in Figure 7.6. The duration of REMS in minutes is shown by elapsed hour for each visit in the lower window of Visit View, and the total duration of REMS (in minutes) for each visit is shown in the upper window. Each EM detector may be selected in combination with the preprocessing options available (i.e. none, adaptive noise cancelling of EEG, wavelet denoising, or adaptive noise cancelling and wavelet denoising) for either the horizontal or vertical EOG channels using the pull down menus. Per visit metrics are shown in the legend on the lower left corner with the number of studies evaluated per demographic group in parenthesis.
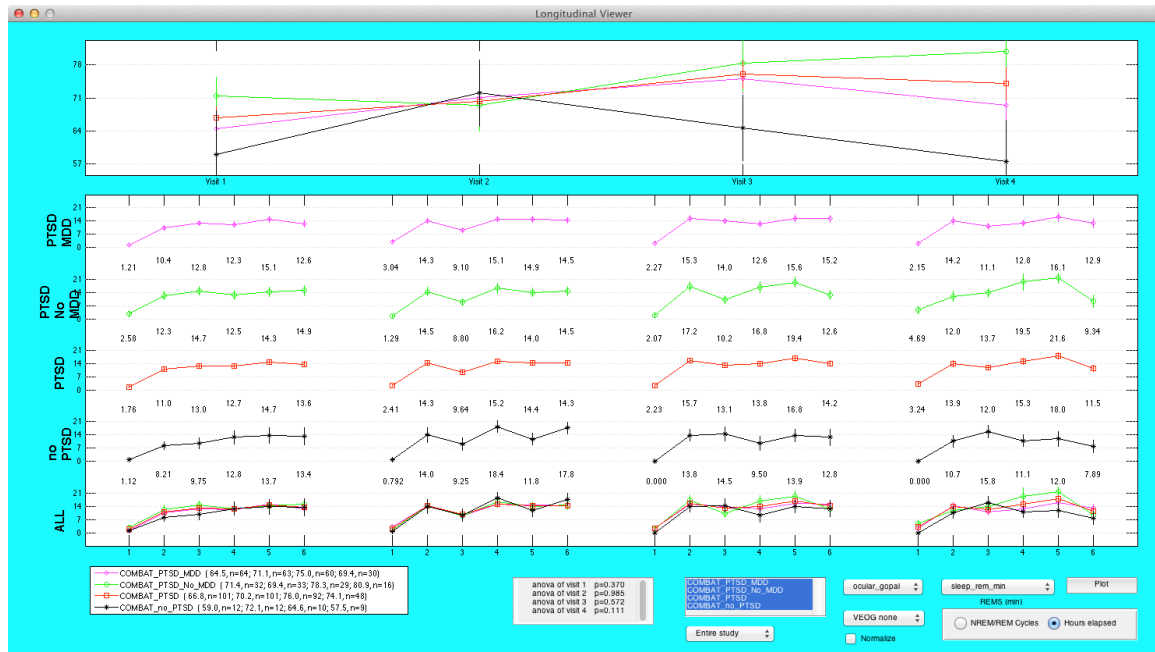
Figure 7.6: Visit View screenshot. Visit View is a visual analytics tool for exploring Combat Veteran Cohort biomarkers by visit as output from SEV's batch processing mode and stored in a MySQL database. Total REMS duration is shown in the upper window per visit, while REMS duration is shown by elapsed hour for each visit in the lower window. Each stratified group (e.g. PTSD with MDD) is identified by a unique color, and placed at a different offset in the lower window with the group name directly to the left. The legend in the lower left further identifies the groups and provides per visit measurement values and sample sizes for each. The top window reveals differences between the PTSD and non PTSD groups, which converge on the second visit, but otherwise drift in opposite directions with PTSD showing increased REMS overall.

A major benefit of visually exploring data is the ease in which it is to find problems because outlier values are so easy to see. I added context menus to Visit View's elapsed time plots (lower window) so that users can select and reveal the patient database key and associated metric values of a selected, problematic point of interest. For example, Figure 7.7 reveals a major shift in non PTSD group measures (REM activity by quadrant detector) during the third visit, which are significantly elevated due to an outlier measurement in the fourth hour. Right clicking on this point open the context menu shown in the bottom window (Figure 7.7). Selecting "Show patient values" causes Visit View to display the patient keys and values in the MATLAB console and identifies key

237

433 to be the problem with 15,600 events (30-50 times higher than the other studies at this same point in time).
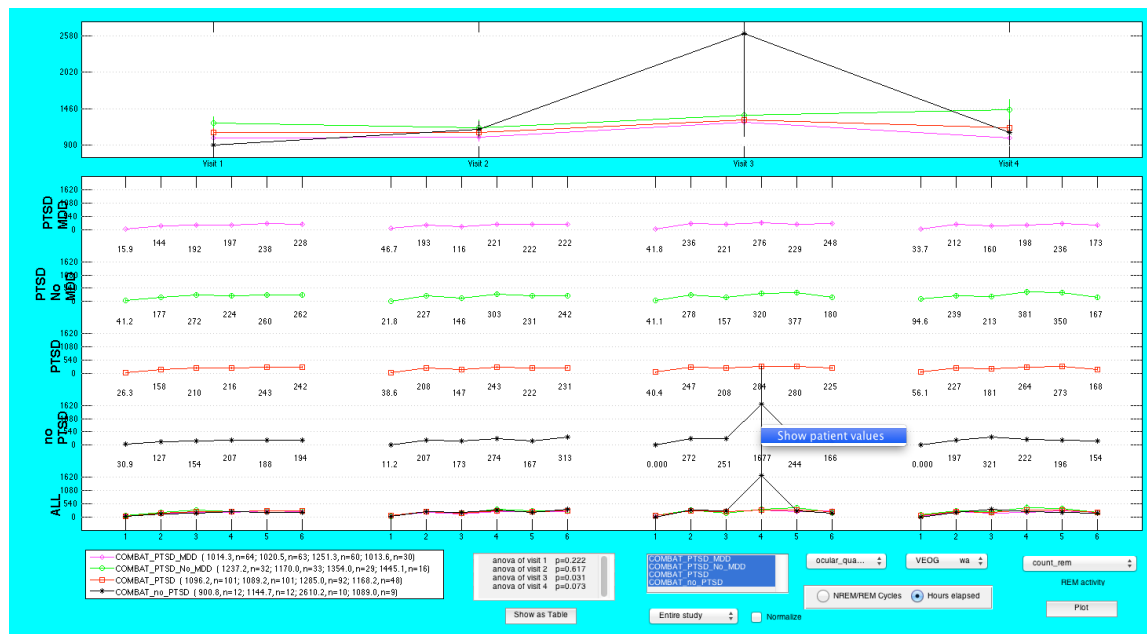
Figure 7.7 Identifying outliers with Visit View. The selected context menu above allows a user to see the non PTSD patient database keys and associated values for the fourth hour of the third visit to isolate the outlier study for investigation with the SEV.

The SEV shows significant interference at 25 Hz at the time of question for the outlier study (i.e. fourth elapsed hour of study ID 433).   The artifact is relatively short, approximately 12 minutes (likely identified and corrected quickly during the study), but must be addressed and removed considering its strong impact on the overall results. Automatically detecting and removing this type of artifact is an issue that can be addressed using the SEV's power frequency artifact detector. However, this is issue is representative of the general problem of handling outlier data, regardless of the source, and a more general solution should be available within Visit View itself to handle unforeseen outlier problems directly, and immediately.
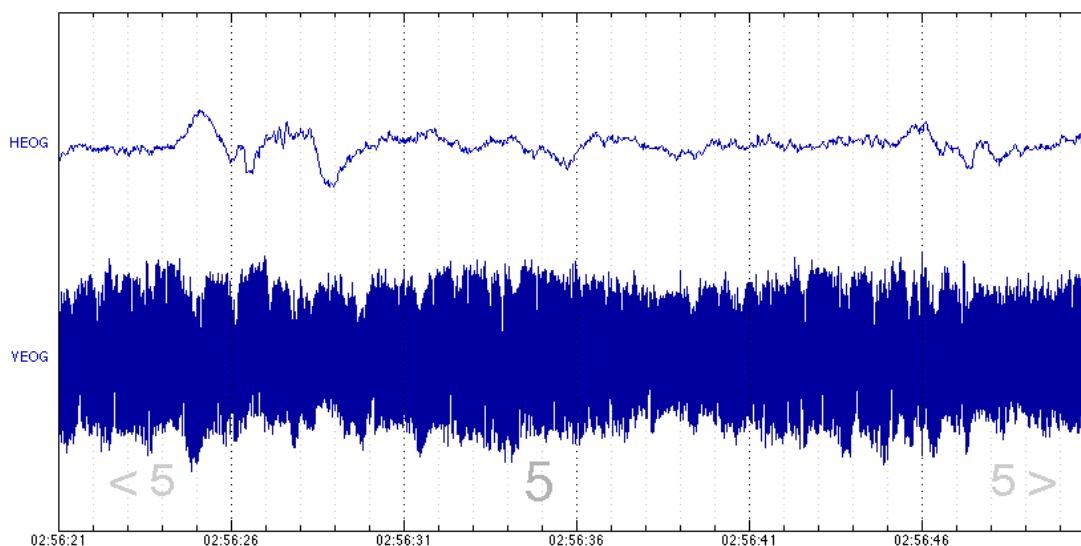
Figure 7.8: SEV display of VEOG and HEOG channels during REM period of a patient study identified with Visit View. The abrupt, strong, 25 Hz interference in the VEOG channel during REMS here causes the outlying EM activity seen in Figure 7.6.

Figure 7.9: Visit View's outlier selection dialog. Visit View removes any studies selected from this dialog and recalculates the current detectors metrics by visit and elapsed time or NREM/REM sleep cycle.

I created a simple dialog to present the sleep study database keys and associated measure currently being evaluated in Visit View to the user and which is also activated through a context menu selection. Figure 7.9 shows the dialog for our current example. In this case, I selected study 433 and its current REM activity of 15,600 REMs (for visit three,

elapsed hour four). Once the dialog is closed, Visit View removes the selected study from the group (i.e. visit three combat veterans without PTSD diagnosis) and re-queries and recalculates the ocular detector REM metrics for visit three and updates the display. Figure 7.10 shows Visit View with the outlier removed. Tabular results are provided in the following section (i.e. Section 7.5.3).
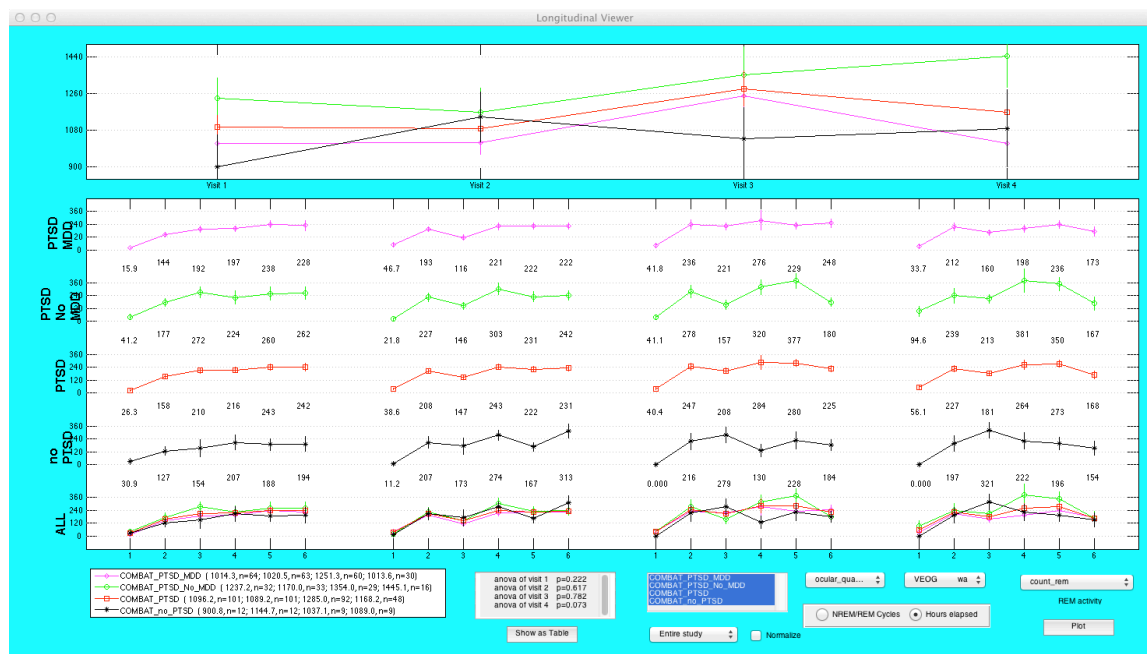
Figure 7.10: Visit View after using the selection dialog in Figure 7.9 to remove the outlier study that caused the erroneous results seen in Figure 7.8.

Visit View includes a unique visualization built specifically for EM quadrant detector events, which in addition to time, stage, epoch, and duration of occurrence also include the quadrant of occurrence. These metrics are shown using a circle divided into four quadrants with the radius and color encoding the normalized mean value of each quadrant relative to the other three. Figure 7.11 shows the Visit View with the quadrant detector selected for analysis of ocular distributions elapsed by hour (i.e. EM duration per quadrant) for each visit in the lower window and the average distribution by visit in the upper window. Circle radii are normalized in the upper and lower axes by the greatest

mean metric observed in the respective axes of each. This visualization provides a global observation of each measure relative to others (e.g. between visits or PTSD categories). The symmetry of a measures four quadrants is calculated as mean quadrant value (i.e. sum of all quadrant values divided by four) normalized by the maximum quadrant value for that observation. An observation whose quadrant values are all very similar will have a symmetry score close to one, while an observation with a single high quadrant value and three other, relatively smaller quadrant values will have a score close to 0.25.
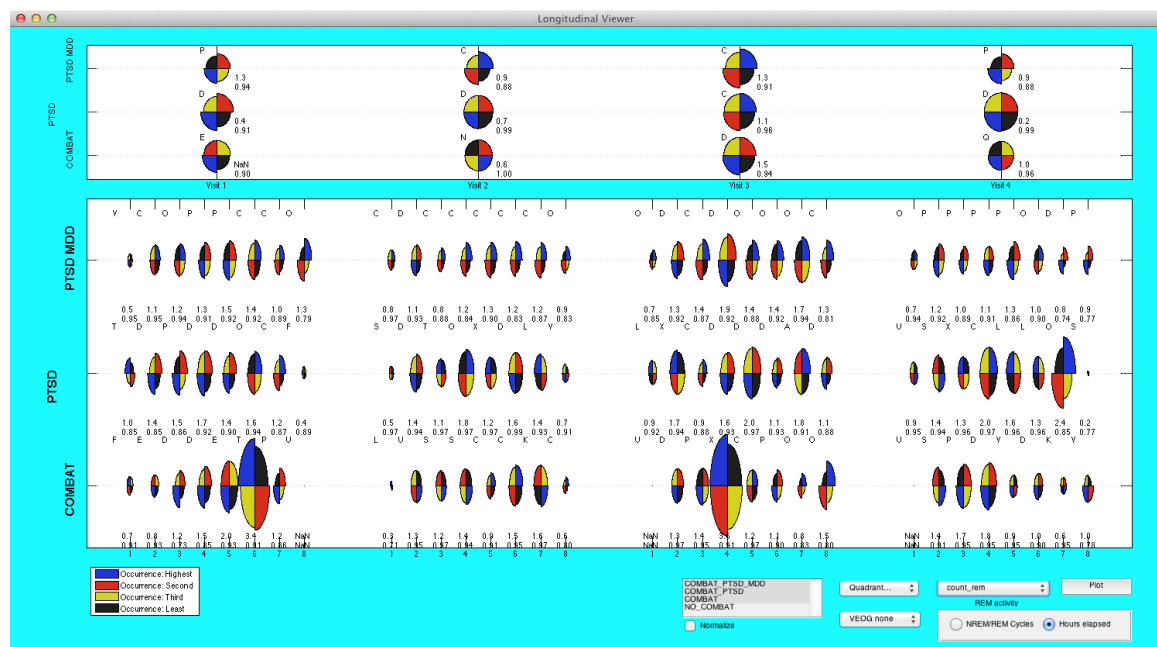
Figure 7.11: Quadrant detector ocular patterns. The Visit View program offers a unique visual display of the quadrant detector parameters that lets users see each EM metric's distribution by quadrant across the night by visit. Combat, combat with PTSD, and combat with PTSD and MDD groups are shown here.

Locally, per observation, the relative time spent in each quadrant is color-coded. Blue represents the greatest occupancy duration, red the second most, yellow third most, and black for the quadrant with the least time occupied. There are 24 possible combinations for relative quadrant duration, which are arbitrarily identified by letters A through X (ties are not allowed).

7.5.3 RESULTS

PSG hypnogram measures are shown in Table 7.4 for the PTSD+ and controls. This includes total sleep time (TST), duration of NREMS, REMS, and WASO, and the total number of NREM/REM cycles determined by the cycle detection algorithm discussed in 7.5.1.2. REMS duration converges to similar levels on the second visit, but is otherwise much higher in PTSD (see also Figure 7.6). Often several consecutive sleep studies are conducted to alleviate *first night effects* related to sleeping in a new setting; however the second night data here is least representative by comparison to the three other visits. Single trial studies may be sufficient for research not directly investigating night-to-night changes in sleep.

Table 7.4 Hypnogram based measures of the CVC stratified by PTSD+ (patients diagnosed with PTSD) and controls (combat veterans without PTSD diagnosis). Total sleep time (TST) and duration of REMS, NREMS, and WASO are given in minutes. *Cycles* refers to the total number of NREM/REM cycles in a study as determined by the cycle detection algorithm.

		Visit 1	Visit 2	Visit 3	Visit 4
TST	PTSD+	295.6±4.5 (101)	311.0±4.9 (101)	306.5±6.0 (92)	309.1±6.3 (48)
	Controls	280.2±13.1 (12)	322.8±5.2 (12)	303.0±13.6 (10)	297.3±14.3 (9)
	p	0.265	0.100	0.817	0.454
NREMS	PTSD+	228.8±3.9 (101)	240.7±4.0 (101)	230.5±5.0 (92)	235.0±6.0 (48)
	Controls	221.2±11.7 (12)	250.7±4.9 (12)	238.4±12.3 (10)	239.8±10.3 (9)
	p	0.533	0.117	0.552	0.685
REMS	PTSD+	66.8±2.3 (101)	70.2±2.5 (101)	76.0±2.8 (92)	74.1±3.5 (48)
	Controls	59.0±9.3 (12)	72.1±7.0 (12)	64.6±7.0 (10)	57.5±8.9 (9)
	p	0.416	0.805	0.133	0.086
WASO	PTSD+	27.9±2.1 (101)	21.8±2.0 (101)	18.7±1.9 (92)	16.1±2.1 (48)
	Controls	38.8±10.9 (12)	23.7±4.2 (12)	18.3±6.9 (10)	29.2±6.4 (9)
	p	0.328	0.674	0.953	0.055
Cycles	PTSD+	4.22±0.14 (101)	4.50±0.13 (101)	4.24±0.14 (92)	4.46±0.15 (48)
	Controls	4.00±0.46 (12)	5.00±0.33 (12)	4.40±0.31 (10)	4.33±0.58 (9)
	p	0.651	0.151	0.633	0.835

Table 7.5 and Table 7.6 give REM activity (i.e. total number of REMs in REMS) determined by Takahashi, Gopal, dualthreshold, and SHW detectors for wavelet denoised EOG channels (VEOG and HEOG respectively). Wavelet denoising was selected for preprocessing to keep comparisons consistent and because in some cases adaptive filtering increased the number of REM detections because of uncertain, corresponding EEG activity.

The Takahshi method detects the most REM activity for VEOG and HEOG channels, followed next by the dual threshold method, and finally SHW and Gopal methods, which perform similarly with conservative REM activity detection levels by each. PTSD without MDD (i.e. PTSD MDD-) shows the highest vertical (i.e. VEOG channel) REM activity. PTSD with MDD (i.e. PTSD MDD+) shows the next highest REM activity levels, and combat veterans without PTSD or MDD (i.e. PTSD-) predominantly show the lowest activity levels. While PTSD MDD- remains relatively high across all four visits (Takahashi visit 3 is a small exception), PTSD MDD+ remains second highest only through visits one through three and then shows similar to lower activity levels than PTSD- in visit four. It should be noted that the fourth night consisted of a much smaller, distinct subset of the cohort. PTSD+ shows significantly more vertical REM activity than PTSD- for visits one and two with p-values less than 0.01 for t-test comparisons. Takahashi's method also reveals significant differences between PTSD+ and PTSD- for visit three, but this is the only method that reaches significance less than 0.01.

Horizontal EOG REM activity levels (Table 7.6) do not reach statistical significant (p<0.01) for the various groups, but PTSD MDD- shows a similar trend with the highest activity for all visits. PTSD MDD- however alternates each visit with PTSD- for the second most horizontal REM activity (i.e. PTSD with MDD is higher on visit one, PTSD- is higher on visit two, PTSD with MDD is higher on visit three, and PTSD- is higher on visit four).

Table 7.5 VEOG REM activity for Takahashi, Gopal, dual threshold, and SHW algorithms in wavelet denoised VEOG channel. Mean results are shown with SEM and number of samples (in parenthesis) for PTSD with MDD (PTSD MDD+), PTSD without MDD (PTSD MDD-), PTSD regardless of MDD (PTSD+), and combat veterans without PTSD (PTSD-). Statistically significant differences (p<0.01) between PTSD+ and PTSD- are shown in bold.

		Visit 1	Visit 2	Visit 3	Visit 4
Takahashi	PTSD MDD+	224.8±28.38 (64)	226.5±30.49 (63)	277.2±36.48 (59)	177.1±29.35 (30)
	PTSD MDD-	244.5±28.27 (31)	239.9±33.47 (33)	262.9±57.62 (28)	327.9±65.15 (16)
	PTSD+	**236.3±20.9 (100)**	**237.7±22.4 (101)**	**274.8±30.03 (90)**	228.9±29.85 (48)
	PTSD-	**119.0±31.19 (9)**	**131.7±19.44 (11)**	**151.4±38.8 (9)**	221.3±79.68 (9)
Gopal	PTSD MDD+	66.3±8.65 (64)	61.9±7.17 (63)	80.6±9.78 (59)	62.9±9.04 (30)
	PTSD MDD-	71.2±10.41 (31)	80.8±13.46 (33)	94.1±18.64 (28)	89.6±19.93 (16)
	PTSD+	**69.2±6.57 (100)**	71.6±6.78 (101)	86.4±8.93 (90)	71.3±8.80 (48)
	PTSD-	**36.7±6.95 (9)**	47.9±13.67 (11)	66.2±18.59 (9)	63.6±23.97 (9)
Dual threshold	PTSD MDD+	120.0±13.64 (64)	112.7±14.48 (63)	145.3±18.50 (59)	108.0±15.70 (30)
	PTSD MDD-	137.4±20.18 (31)	149.8±24.16 (33)	167.3±29.89 (28)	165.4±32.34 (16)
	PTSD+	**130.1±11.5 (100)**	**130.3±12.6 (101)**	154.9±15.78 (90)	127.1±14.90 (48)
	PTSD-	**61.8±11.88 (9)**	**69.4±15.71 (11)**	106.1±29.94 (9)	113.9±41.27 (9)
SHW	PTSD MDD+	66.7±10.60 (64)	61.5±9.14 (63)	74.9±9.77 (59)	59.8±9.93 (30)
	PTSD MDD-	71.1±10.52 (31)	82.8±14.28 (33)	88.5±17.20 (28)	84.7±18.78 (16)
	PTSD+	**70.6±7.90 (100)**	**72.5±7.93 (101)**	82.1±9.08 (90)	67.8±8.87 (48)
	PTSD-	**29.4±6.48 (9)**	**34.1±8.36 (11)**	58.8±19.02 (9)	62.7±26.49 (9)

Table 7.6 HEOG REM activity for Takahashi, Gopal, dual threshold, and SHW algorithms in wavelet denoised HEOG channel. Mean results are shown with SEM and number of samples (in parenthesis) for PTSD MDD+, PTSD MDD-, PTSD+, and PTSD-.

		Visit 1	Visit 2	Visit 3	Visit 4
Takahashi	PTSD MDD+	297.4±40.16 (64)	294.0±35.40 (63)	377.3±51.73 (59)	330.9±61.8 (30)
Takahashi	PTSD MDD-	321.4±48.22 (31)	328.8±47.78 (33)	412.4±62.39 (28)	474.2±113.5 (16)
Takahashi	PTSD+	317.8±30.8 (100)	320.2±30.1 (101)	394.2±39.40 (90)	377.5±54.34 (48)
Takahashi	PTSD-	187.8±43.87 (9)	320.5±85.20 (11)	292.1±105.97 (9)	402.3±115.03 (9)
Gopal	PTSD MDD+	102.0±14.04 (64)	94.3±12.35 (63)	118.8±16.57 (59)	99.7±19.25 (30)
Gopal	PTSD MDD-	106.3±14.59 (31)	106.7±17.06 (33)	152.9±27.06 (28)	157.2±36.85 (16)
Gopal	PTSD+	106.2±10.3 (100)	102.2±9.86 (101)	131.6±14.01 (90)	118.6±17.42 (48)
Gopal	PTSD-	62.7±18.50 (9)	98.9±25.59 (11)	104.7±32.83 (9)	126.9±37.01 (9)
Dual threshold	PTSD MDD+	208.0±27.51 (64)	190.6±22.48 (63)	245.9±32.55 (59)	211.7±36.52 (30)
Dual threshold	PTSD MDD-	230.5±33.17 (31)	236.5±37.05 (33)	320.7±52.13 (28)	339.9±75.54 (16)
Dual threshold	PTSD+	222.8±21.2 (100)	214.4±19.4 (101)	273.6±27.40 (90)	255.1±34.65 (48)
Dual threshold	PTSD-	138.0±38.15 (9)	211.0±49.09 (11)	224.7±66.66 (9)	275.9±73.67 (9)
SHW	PTSD MDD+	108.6±18.51 (64)	94.8±16.05 (63)	128.0±22.10 (59)	107.6±27.95 (30)
SHW	PTSD MDD-	107.9±17.94 (31)	105.6±18.90 (33)	159.4±34.19 (28)	179.4±49.30 (16)
SHW	PTSD+	112.0±13.4 (100)	103.4±12.4 (101)	140.8±18.33 (90)	130.7±24.20 (48)
SHW	PTSD-	65.4±22.02 (9)	90.1±28.12 (11)	112.4±41.17 (9)	145.2±49.95 (9)

The quadrant detector provides another perspective on eye movement activity during REMS. The quadrant detector is not optimized to detect REM, but rather determines and tracks the pupil's estimated quadrant when the voltage magnitude of both VEOG and HEOG channels exceeds the set threshold (i.e. 30 μV). We limit our analysis of quadrant detector EMs to REMS, and again use wavelet denoising for both EOG channels to remain consistent with the previous REM detector configurations. Figure 7.12 shows the ocular quadrant patterns for each visit in PTSD+ and PTSD-. Interestingly, despite the large number of possibilities (i.e. 24) only two patterns are observed in each group. Furthermore both PTSD+ and PTSD- change or adjust their pattern on the second visit from the first, and then return to the initial ocular pattern for visits three and four. The circles in Figure 7.12 can be thought of as the average amount of time patients spent in

REMS *looking* in the respective quadrant relative to other quadrants; locally based on quadrant color, and globally based on quadrant size (i.e. radius). The highest symmetry score of 0.97 occurs in visit two for those without PTSD. This represents an interesting change compared to the otherwise more elliptical, cross quadrant patterns seen by both groups elsewhere.
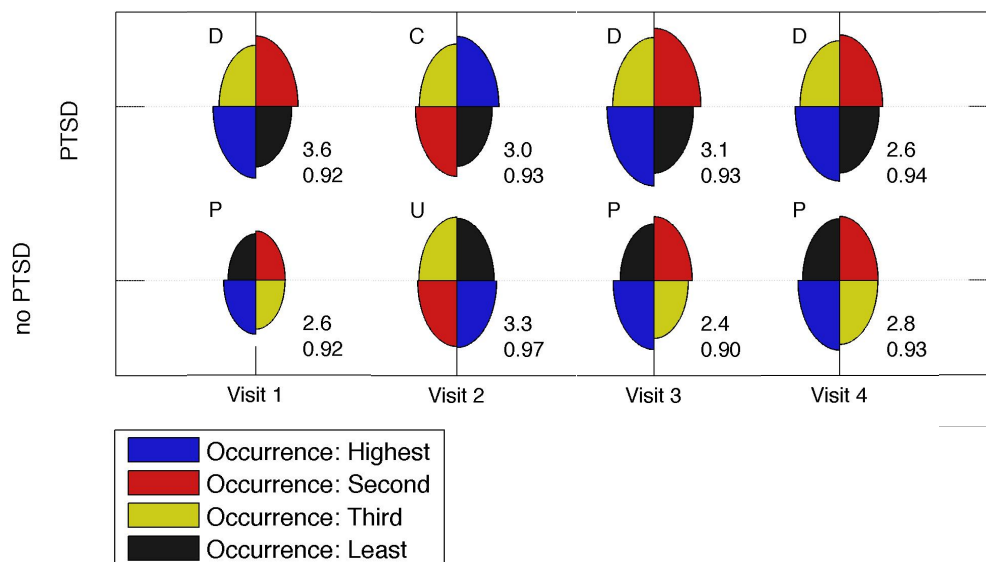
Figure 7.12 Ocular quadrant patterns, per visit, in PTSD (i.e. PTSD+) and combat veterans without PTSD. Both groups show a change in pattern on visit two, however the non PTSD shift in pattern is more unusual and highly symmetric (0.97 symmetry score).

In visits one, three, and four patients with PTSD follow pattern D, looking primarily to the bottom right, then the upper left, then upper right and lastly bottom left. In visit two, the first two quadrant's orders reverse, now going from upper left to lower right first. The last two quadrant's ordering remains the same. Patients without PTSD (PTSD-) follow pattern P for visits one, three, and four looking lower right most, then upper left, then lower left, and lastly upper right. Pattern P is like pattern D but with the last two quadrants' ordering reversed. Like the PTSD group, the PTSD- ocular pattern changes on the second visit, but the change is more unusual. Ocular pattern U is not elliptically

shaped with cross quadrant activity, but rather PTSD- patients look predominantly down first (bottom left, then bottom right) and then up (right, then left). If there is added pressure to *perform* well on the second visit for some reason, this break in pattern may reflect some neurological or cognitive adjustment. It may also reflect a *catch-up* mechanism given the particularly low level of REM activity seen by this group on visit one. Further research is needed.

7.6 REDUCED INTER-HEMISPHERIC COHERENCE OF SLEEP EEG IN COMBAT VETERANS DIAGNOSED WITH PTSD

This section examines the magnitude squared coherence between the F-3 and F-4 EEG channels as a measure of inter-hemispheric synchrony during sleep. This analysis is driven by a hypothesis drawn from ethological work on animal sleep under the threat of predation[139]. In brief, a number of species appear to employ unihemispheric sleep in order to achieve sleep while maintaining vigilance. A residuum of this functionality in humans could be manifested as reduced inter-hemispheric coherence during sleep.

Reduced EEG synchrony has been observed in MDD [140] and Alzheimer's disease (AD) [141, 142] compared to controls. EEG synchrony has been explored in the context of epilepsy and schizophrenia [143] as well. Dauwels *et al*. provides an excellent comparative study of several approaches they used to measure EEG synchrony in an effort to predict early stages of AD [144]. The magnitude-squared coherence will be used instead to measure hemispheric symmetry in the CVC and is described next.

7.6.1 MAGNITUDE SQUARED COHERENCE

The following derivations are taken from Therrien [145]. The cross-power density spectrum or *cross-spectrum* is a helpful measure of correlation between to random signal at a particular frequency ω. It is the Fourier transform of the cross-correlation function R_{xy} between two random signals x and y,

$$S_{xy}(e^{j\omega}) = \sum_{l=-\infty}^{\infty} R_{xy}[l]e^{-j\omega}$$

If x and y are real values (i.e. not complex) then R_{xy} is real and

$$S_{xy}(e^{j\omega}) = S_{yx}^*(e^{j\omega}).$$

Thus the cross-spectrum magnitude is even and its phase odd for real random signals.

The inverse relation

$$R_{xy}[l] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xy}(e^{j\omega})e^{j\omega l} d\omega$$

allows the cross-correlation R_{xy} to be obtained from the cross spectrum $S_{xy}(e^{j\omega})$.

The *coherence function* is the normalized cross-spectrum

$$C_{xy}(e^{j\omega}) \equiv \frac{S_{xy}(e^{j\omega})}{\sqrt{S_x(e^{j\omega})}\sqrt{S_y(e^{j\omega})}}$$

The magnitude squared coherence (MSC)

$$\left|C_{xy}(e^{j\omega})\right|^2 \equiv \frac{\left|S_{xy}(e^{j\omega})\right|^2}{S_x(e^{j\omega})S_y(e^{j\omega})}$$

has the useful property of ranging from 0 to 1

$$0 \le \left|C_{xy}(e^{j\omega})\right|^2 \le 1.$$

Furthermore, the band-averaged MSC can be calculated across a range of frequencies as

$$\left| C_{xy}(e^{j\bar{\omega}}) \right|^2 \equiv \frac{\left| \sum_{\omega=\omega_o}^{\omega_1} S_{xy}(e^{j\omega}) \right|^2}{\sum_{\omega=\omega_o}^{\omega_1} S_x(e^{j\omega}) \cdot \sum_{\omega=\omega_o}^{\omega_1} S_{xy}(e^{j\omega})} .$$

I have taken some liberty in the above notation to simply define the band of interest as discrete steps from $\omega_0 \le \omega \le \omega_1$ which are defined by the sampling rate, the bands start and stop frequencies, and the number of samples used.

The SEV's *detection_mscohere* function calculate the average MSC between the F-3 and F-4 EEGs on 30-sec intervals or epochs. The function calculates the MSC, using MATLAB's *mscohere*[2] function, for consecutive, overlapping, user-specified length blocks of the two input signals for each epoch. I specified a block length of 6-sec and an overlap interval of 3-sec overlap, which results in nine MSC's calculated and averaged together each epoch for a single mean MSC with frequencies from 0 to 50 Hz (f_s= 100 Hz) at 1/6 Hz resolution (ie. $\frac{f_s}{6s \cdot f_s}$).

In addition to the standard information included with each SEV event (e.g. onset, offset, stage and epoch of occurrence) the band-average MSC is calculated for the following frequency bands

- Delta –   [0, 4) Hz
- Theta –   [4, 8) Hz
- Alpha –   [8, 12) Hz
- Sigma –   [12, 16) Hz
- Beta –   [16, 30) Hz
- Gamma – [30, 50) Hz

---

[2] *mscohere* uses modified periodogram Welch averaging to calculate both the individual spectrums and the cross spectrum in obtaining the MSC

The SEV outputs these 30-sec interval events and parameters to the MySQL database I setup for the CVC.

Fein *et al*. show that EEG spectral coherency data is confounded by power and phase effects when calculated from channels that use a common reference whose observed power is relatively large [146]. CVC uses a "linked ears" montage that uses the mastoids to obtain a common reference. Fein suggests using digitally linking the ears for appropriate EEG coherency studies.

7.6.2 EXPLORING MEAN MAGNITUDE SQUARED COHERENCY WITH VISIT VIEW

Visit View, previously described for exploring eye movements, can be adjusted to consider the cross EEG channel MMSC in different bands in a manner similar to the eye movement detections (i.e. by visit number and by elapsed duration, or cycle) with some modification. Specifically, we want to examine the frequency band parameters. It would suffice to directly call these parameters by name (since we have defined them) from within our query and then present them to the user. A more general solution however is to query the parameter field of one detector event, use MATLAB's *fieldnames* function to determine the names of the available parameters, and present these names as selections for the user. This allows further frequency bands to be defined and modified in the SEV as well as completely new parameters, without requiring further modification to Visit View. The only changes needed to be made to the viewer are which detection methods to allow the user to query, and this can also be handled in a similar manner by querying the MySQL detector information table which lists all detectors with events available in the CVC database. I also added a control button to dump on-screen results to a tab delimited table for easy *cut-and-paste* transfer to a document or spreadsheet editor.
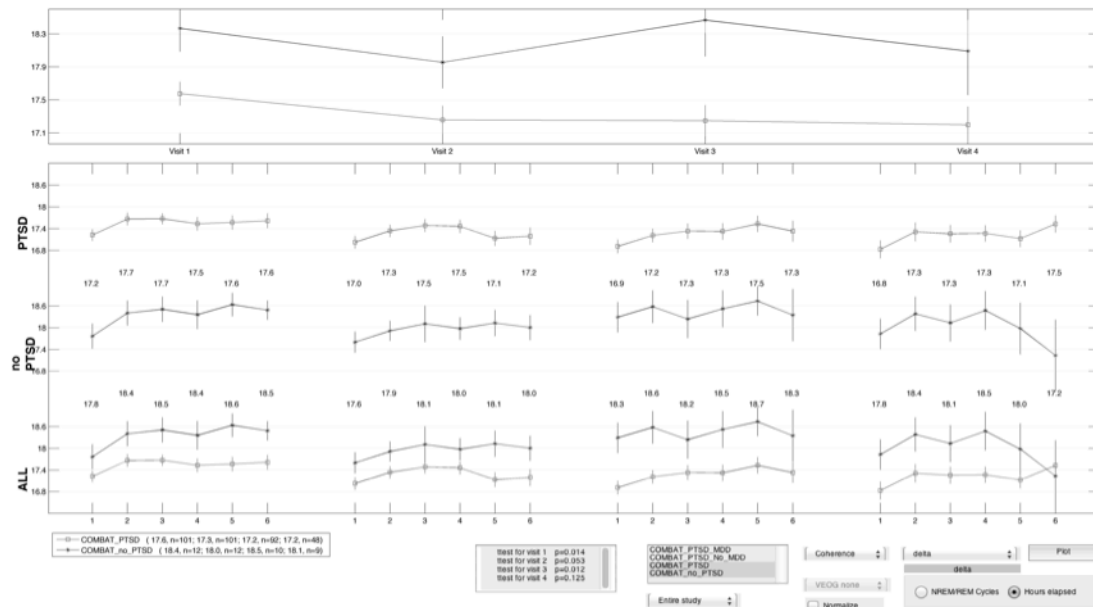
Figure 7.13 PTSD View screenshot showing magnitude squared coherence as a delta band sum (0, 4 Hz) by visit (top portion) and by elapsed hour per visit (bottom portion). Visit 1 and Visit 3 show statistically significant differences ($p < 0.05$) between combat veterans diagnosed with posttraumatic stress disorder and those without.

7.6.3 RESULTS

SEV processed and Visit View interpreted band-average MSC values (F-3 and F-4 EEGs) for all sleep (NREMS and REMS) are shown in

Table 7.7 for the CVC's PTSD+ and control groups. PTSD+ shows lower band-average MSC values than controls in every band, in every visit. Significant differences (i.e. student t-test with p<0.05) are observed in delta, theta, and alpha frequency bands, compatible with a residuum of unihemispheric sleep in PTSD.

Table 7.7 Magnitude squared coherency between F-3 and F-4 EEG leads in sleep (REMS and NREM sleep) of combat veterans diagnosed with posttraumatic stress disorder (PTSD+) or without (Controls). Coherency is shown by frequency band for each group by visit, as are the p-values for t-test comparisons between groups by visit.

		Visit 1	Visit 2	Visit 3	Visit 4
Delta	PTSD+	0.739±0.01 (101)	0.725±0.01 (101)	0.726±0.01 (92)	0.721±0.01 (48)
	Controls	0.775±0.01 (12)	0.754±0.01 (12)	0.774±0.02 (10)	0.760±0.02 (9)
	p	**0.005**	**0.042**	**0.013**	0.077
Theta	PTSD+	0.735±0.01 (101)	0.719±0.01 (101)	0.722±0.01 (92)	0.718±0.01 (48)
	Controls	0.777±0.01 (12)	0.758±0.01 (12)	0.775±0.02 (10)	0.763±0.02 (9)
	p	**0.003**	**0.019**	**0.006**	**0.031**
Alpha	PTSD+	0.709±0.01 (101)	0.695±0.01 (101)	0.697±0.01 (92)	0.692±0.01 (48)
	Controls	0.735±0.02 (12)	0.717±0.02 (12)	0.747±0.02 (10)	0.724±0.02 (9)
	p	0.168	0.297	**0.012**	0.216
Sigma	PTSD+	0.636±0.01 (101)	0.620±0.01 (101)	0.626±0.01 (92)	0.621±0.01 (48)
	Controls	0.658±0.02 (12)	0.641±0.03 (12)	0.668±0.03 (10)	0.649±0.03 (9)
	p	0.376	0.508	0.144	0.441
Beta	PTSD+	0.576±0.01 (101)	0.569±0.01 (101)	0.570±0.01 (92)	0.574±0.01 (48)
	Controls	0.613±0.03 (12)	0.587±0.03 (12)	0.621±0.03 (10)	0.597±0.04 (9)
	p	0.168	0.613	0.141	0.577
Gamma	PTSD+	0.502±0.01 (101)	0.500±0.01 (101)	0.499±0.01 (92)	0.504±0.01 (48)
	Controls	0.535±0.02 (12)	0.507±0.03 (12)	0.513±0.03 (10)	0.523±0.03 (9)
	p	0.118	0.803	0.651	0.57

Table 7.8 and

Table 7.9 show the band-average MSC during REMS and NREMS respectively. Inner frequency bands theta, alpha, sigma, and beta show higher coherency levels in REMS, while the outer delta and gamma bands show higher coherency levels in NREMS for both PTSD+ and controls. Higher coherency levels in these bands are likely representative of the increased ocular activity during REMS. Statistical difference of theta is lost in REMS of visit-4 as is delta for visit-2 and visit-3. The statistical difference found in alpha-band MSC during visit-3 sleep – NREMS and REMS together and separately – is likely coincidental given since similar results are not seen during the other three visits.

Table 7.8 Magnitude squared coherency between F-3 and F-4 EEG leads in REMS of combat veterans diagnosed with posttraumatic stress disorder (PTSD+) or without (Controls). Coherency is shown by frequency band for each group by visit, as are the p-values for t-test comparisons between groups by visit.

		Visit 1	Visit 2	Visit 3	Visit 4
Delta	PTSD+	0.720±0.01 (99)	0.700±0.01 (100)	0.700±0.01 (92)	0.699±0.01 (48)
	Controls	0.760±0.02 (11)	0.723±0.01 (12)	0.741±0.03 (10)	0.725±0.03 (9)
	p	**0.028**	0.174	0.138	0.435
Theta	PTSD+	0.782±0.01 (99)	0.767±0.01 (100)	0.773±0.01 (92)	0.776±0.01 (48)
	Controls	0.820±0.01 (11)	0.814±0.01 (12)	0.836±0.02 (10)	0.805±0.02 (9)
	p	**0.009**	**0.006**	**0.001**	0.173
Alpha	PTSD+	0.749±0.01 (99)	0.735±0.01 (100)	0.742±0.01 (92)	0.746±0.01 (48)
	Controls	0.779±0.02 (11)	0.772±0.02 (12)	0.788±0.02 (10)	0.763±0.03 (9)
	p	0.191	0.119	**0.036**	0.527
Sigma	PTSD+	0.690±0.01 (99)	0.672±0.01 (100)	0.680±0.01 (92)	0.679±0.01 (48)
	Controls	0.698±0.03 (11)	0.702±0.03 (12)	0.734±0.03 (10)	0.691±0.04 (9)
	p	0.781	0.339	0.057	0.752
Beta	PTSD+	0.590±0.01 (99)	0.575±0.01 (100)	0.578±0.01 (92)	0.574±0.01 (48)
	Controls	0.633±0.03 (11)	0.601±0.04 (12)	0.639±0.03 (10)	0.584±0.04 (9)
	p	0.187	0.485	0.081	0.807
Gamma	PTSD+	0.485±0.01 (99)	0.475±0.01 (100)	0.479±0.01 (92)	0.470±0.01 (48)
	Controls	0.526±0.03 (11)	0.484±0.03 (12)	0.507±0.03 (10)	0.470±0.03 (9)
	p	0.167	0.743	0.358	0.982

Table 7.9 Magnitude squared coherency between F-3 and F-4 EEG leads in NREMS of combat veterans diagnosed with posttraumatic stress disorder (PTSD+) or without (Controls). Coherency is shown by frequency band for each group by visit, as are the p-values for t-test comparisons between groups by visit.

		Visit 1	Visit 2	Visit 3	Visit 4
Delta	PTSD+	0.745±0.01 (101)	0.733±0.01 (101)	0.735±0.01 (92)	0.729±0.01 (48)
	Controls	0.779±0.01 (12)	0.763±0.01 (12)	0.780±0.02 (10)	0.768±0.02 (9)
	p	**0.007**	**0.039**	**0.017**	0.063
Theta	PTSD+	0.721±0.01 (101)	0.706±0.01 (101)	0.706±0.01 (92)	0.701±0.01 (48)
	Controls	0.765±0.01 (12)	0.741±0.01 (12)	0.758±0.02 (10)	0.751±0.02 (9)
	p	**0.006**	**0.03**	**0.008**	**0.016**
Alpha	PTSD+	0.697±0.01 (101)	0.684±0.01 (101)	0.683±0.01 (92)	0.676±0.01 (48)
	Controls	0.721±0.02 (12)	0.700±0.02 (12)	0.735±0.02 (10)	0.712±0.02 (9)
	p	0.231	0.452	**0.012**	0.2
Sigma	PTSD+	0.620±0.01 (101)	0.604±0.01 (101)	0.609±0.01 (92)	0.603±0.01 (48)
	Controls	0.649±0.02 (12)	0.624±0.03 (12)	0.651±0.03 (10)	0.639±0.03 (9)
	p	0.247	0.531	0.151	0.321
Beta	PTSD+	0.572±0.01 (101)	0.567±0.01 (101)	0.569±0.01 (92)	0.573±0.01 (48)
	Controls	0.608±0.03 (12)	0.581±0.03 (12)	0.615±0.03 (10)	0.597±0.04 (9)
	p	0.184	0.684	0.182	0.567
Gamma	PTSD+	0.506±0.01 (101)	0.506±0.01 (101)	0.507±0.01 (92)	0.514±0.01 (48)
	Controls	0.538±0.02 (12)	0.512±0.03 (12)	0.511±0.03 (10)	0.531±0.03 (9)
	p	0.132	0.847	0.873	0.635

7.7 SUMMARY

This chapter investigates automatically detected eye movements and EEG coherence from CVC sleep studies and introduces *Visit View* to analyze these measures across each study visits as well as between study visits. Several different eye movement detectors were examined and optimized for increased REM:NREM detection density by applying wavelet denoising, adaptive noise cancellation of the EEG, and the combination of adaptive noise cancellation and wavelet denoising. The spectral classification method for classifying REM density performed poorly in all cases and has limited utility for eye movement analysis as is. A new method, *quadrant detector*, was introduced to take

advantage of the CVC's EOG montage by combining the vertical and horizontal EOG channels to obtain a single set of eye movement detections which include the ocular quadrant that each EM occurs in. The *Visit View* software was updated to provide specific visualizations for investigating the possible ocular transition states for future state space analysis work.

A method for obtaining the band-average MSC for two EEG channels was implemented in the SEV and used to investigate inter-hemisheric coherence in the CVC using *Visit View*. A statistically significant decrease in this coherency measure is seen in combat veterans with PTSD compared to healthy combat veterans for the delta, theta, and alpha bands. These differences were particularly prominent in the theta band, which consistently showed decreased MSC for the PTSD group across all four studies.

The SEV and Visit View can easily be updated to investigate other measures of coherency in the data set of interest by following an approach similar to the one described in this chapter.

CHAPTER 8   CONCLUSION

In this dissertation, I have shown how signal estimation and classification techniques, combined with visual interaction and the medical community accepted receiver operating characteristics (ROC) can be used to investigate PSG based sleep studies (and measures) from large, diverse populations for genetic, medical, and clinically relevant purposes. I did this by considering six problems currently faced by the sleep research community and developed the signal processing, classification, optimization, and visualization measures needed for each. This was specifically accomplished by: (1) extending ROC theory to improve diagnostic criteria for narcolepsy using clinical and PSG measures from 1,000 narcoleptics and 1,000 controls in Chapter 2; (2) developing visualization tools for identifying electroencephalogram (EEG) power spectral density (PSD) phenotypes from 1,836 studies for genome wide association (GWAS) in Chapter 4; (3) developing, validating, and implementing a robust periodic leg movement (PLM) detector with improved criteria for handling respiratory related events in Chapter 5; (4) identifying single nucleotide polymorphisms (SNP) with susceptibility for PLM using 2,396 PSG sleep studies in Chapter 6; (5) applying wavelet denoising and least mean square adaptive filtering and also developing a new eye movement detection method to improve rapid eye movement recognition and characterization in 457 PSG studies of combat veterans diagnosed with PTSD in Chapter 7; and (6) used magnitude squared coherence of EEG channels to show reduced inter-hemispheric coherence in the sleep EEG in PTSD in Chapter 7.

Much of the research that began or was developed during the course of this dissertation has continued into the research of others. For example, the PLM classification algorithm introduced in Chapter 4 is being used to investigate the relationship to PLM and excessive daytime sleepiness by Eileen Leery, the relationship to iron deficiency and PLM by Jason Li, and as a feature for analysis for narcolepsy by Emil Munk. Laurel

Finn, at the University of Wisconsin Madison, is also investigating PLM within the WSC. Furthermore, the PLM detector is currently being implemented for use at Stanford's Sleep Clinic in Redwood City, CA. The softROC program discussed in Chapter 2 continues to find use in analyzing narcolepsy as recently seen acknowledged in a manuscript submitted to the journal *Science*. And the SEV, introduced in Chapter 3, continues to fill a niche in ongoing sleep research. Ear nose and throat surgeons at Stanford are hoping to find predictive links, using the SEV, between PSG quantifiable obstructive breathing patterns and what type and degree of surgery will be necessary for patients. At the University of Wisconsin Madison, researchers are now using the SEV to investigate changes in EEG delta power during sleep in women before and after menopause.

The sleeping brain provides a window to the brain apart from sensory input noise. Understanding the brain's behavior during this relatively isolated state can reveal new things. Searching for and understanding new biomarkers of brain health will lead to greater insight, both individually and collectively as part of societies and culture. We are making progress, but a long journey remains ahead.

Polysomnography (PSG) based sleep studies are data-rich yet often just visually interpreted by technicians and doctors for basic parameters to define sleep architecture and highlight sleep disordered breathing events. Human sleep is a complex process that plays a significant role in our health and wellbeing that PSG can give insight to. Unfortunately, while the number of devices used to monitor sleep continues to grow and allow digital collection of this data to reach proportions necessary for genetic and epidemiological level analysis, the growth in algorithms and analytic frameworks necessary to examine it has not.

Fortunately, it is not too late to catch up, particularly with the continued advancement in more general, collaborative aspects of software design and development. Open source software projects present a huge opportunity for sharing with and helping others with

their sleep related inquiries. Publicly available software repositories widen the gates of collaborative candidates. The SEV, PhenoFinder, and softROC programs, all software developed over the course of this thesis, are now available online (e.g. http://www.github.com/informaton) for the research community to use, modify, and enhance for their own inquiries (sleep related or not). These repositories provide built in support for tracking changes to software, forums for discussing problems or features, and wikis to easily update documentation of the software in step with changes. This type of openness and flexibility is necessary for projects like the SEV to continue to grow and find new applications under the guidance and motivation of new graduate students, research laboratories, and industry supported efforts. Many of the ideas that began and flourished over the course of this dissertation, now have the ability continue living on.

There are additional challenges faced by the medical research community where, fortunately treatment and safeguarding of human life is of the highest priority. This brings necessary oversight to ensure compliance with current measures for protecting patients and their privacy. However, the placement and existence of such protective measures does not necessitate a slow-down in algorithm development and software design for robust large scale processing of human collected data. The process must be done smartly so that robust development, design, and testing can be done within applicable data constraints (e.g. de-identification and proper safeguarding of human subject data) in parallel with efforts to bring compliance to larger data storage frameworks. For example, at the start of my thesis I was provided a very small number of sleep studies to work with (i.e. less than five) and develop specific algorithms on a small scale (e.g. high frequency artifact detection). However, using just a few studies it was possible to begin developing an automation framework for the forthcoming studies that would continue to arrive from our collaborators at the University of Wisconsin Madison, and hopefully from new opportunities (e.g. the PTSD cohort). Typically, these NPSG-based sleep studies were transferred to encrypted hard drives that required special care and handling to exchange. Once these studies are obtained they must be

259

preprocessed to a form suitable for automation. PSG montage configurations and collection technology is continually evolving and these changes must be addressed and accounted for in order to utilize the software processing pipelines that are in place. Adjusting the pipeline to handle a variety of cases adds complexity and the likelihood of processing errors. Still, *normalizing* different PSG configurations so that as many as possible can be analyzed together remains somewhat of an art.

I expect and hope that as time continues, a more flexible, open standard for collection and storage of PSG-based sleep studies will arrive to remove this current bottleneck in researching sleep data on a large scale. Another bottleneck is the exchange of this data. Fortunately, commercial vendors have recognized the need for safeguarding data like this. Companies like *Amazon* already provide Health Insurance Portability and Accountability Act (HIPAA) compliant services for the storage and exchange of patient data. Soon, it will no longer be necessary to mail hard drives between collaborators to exchange information, or develop local software solutions that must then be specially configured at other sites. Instead, PSG-based sleep cohorts will exist in the *cloud*, in online servers, which will simultaneously provide privacy protection and accessibility for continued research development. Database frameworks will be utilized to optimize access to these databases and software will be developed within this context. Web based tools will continue to grow in value and the statistical tools of the future will be those that are able to keep up with and even lead this trend toward decentralized storage and automated analysis.

The number of sleep studies performed each year, already in the millions, will continue to rise. And for the sake of cost, more and more of these studies will be conducted at home using ambulatory devices, which the primary care physician or caretaker must understand (or trust the software or trained technician's interpretation of) to effectively care for his or her patient. The field of medicine is changing. With so much digital information available, automatic, assistive scoring and recognition of health biomarkers will play a

more, commonly accepted role in medical practice. There is a huge need for electrical engineers and medical doctors to continue working together so that we can better understand the data stored on computers without losing sight of the patient in front of us.

Today, when so much continues to remain unknown about our health and sleep, it is imperative to engage medical experts and provide useful tools, methods, and theory to move forward together. The integration of signal processing methods with visual analytics has the potential of accelerating this partnership.

LIST OF ABBREVIATIONS

AASM	American Association Of Sleep Medicine
ACC	Accuracy
AHI	Apnea Hypopnea Index
ANC	Adaptive Noise Cancellation
API	Application Programming Interface
ASDA	American Sleep Disorders Association
AUC	Area Under The Curve
BMI	Body Mass Index
C-3	Central EEG Position 3
CAPS	Clinician-Administered PTSD Scale
CPAP	Continuous Positive Air Pressure
CSF	Cerebral Spinal Fluid
CVC	Combat Veteran Cohort
DNA	Deoxyribonucleic Acid
DPS	Delayed Phase Syndrome
DSM	Diagnostics And Statistics Manual
ECG	Electrocardiography
EDF	European Data Format
EEG	Electroencephalography
EM	Eye Movement
EMG	Electromyography
EOG	Electrooculagraphy
F-3	Frontal EEG Position 3
F-4	Frontal EEG Position 4
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
GEE	Generalize Estimating Equations
GUI	Graphical User Interface
GWA	Genome Wide Association
GWAS	Genome Wide Association Study
HEOG	Horizontal Electroencephalography
LAT	Left Anterior Tibialis
LM	Leg Movement
LMS	Least mean square
MDD	Major Depressive Disorder

MMSC	Mean Magnitude Squared Coherence
MSC	Magnitude Squared Coherence
MSE	Mean-Squared-Error
MSL	Mean Sleep Latency
MSLT	Multiple Sleep Latency Test
MUSIC	Multiple Signal Components
NPSG	Nocturnal Polysomnography
NPV	Negative Predictive Value
NREM	Non-Rapid Eye Movement Sleep
NREMS	Non-Rapid Eye Movement Sleep
O-1	Occipital EEG Position 1
OR	Odds Ratio
PLM	Periodic Leg Movement
PLMD	Periodic Limb Movement Disorder
PLMI	Periodic Leg Movement Index
PLMS	Periodic Leg Movements In Sleep
PLMW	Periodic Leg Movements In Wake
PPV	Positive Predictive Value
PSD	Power Spectral Density
PSG	Polysomnography
PTSD	Posttraumatic Stress Disorder
qROC	Quality Receiver Operating Characteristics
R&K	Rechtschaffen and Kales
RAT	Right Anterior Tibialis
REM	Rapid Eye Movement
REMBD	Rapid Eye Movement Behavior Disorder
REML	Rapid Eye Movement Latency
REMS	Rapid Eye Movement Sleep
RIP	Respiratory Inductance Plethysmography
RLS	Restless Legs Syndrome
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
SCID	Structured Clinical Interview For The DSM-III-R
SDB	Sleep Disordered Breathing
SE	Sensitivity
SEI	Sleep Efficiency
SLAT	Sleep Latency
SNP	Single Nucleotide Polymorphism
SNR	Signal-To-Noise-Ratio

SNR+	Signal-To-Noise-Ratio Enhancement
SOREMP	Sudden Onset Rapid Eye Movement Period
SP	Specificity
Spo2	Oxygen Saturation
SSC	Stanford Sleep Cohort
SWS	Slow Wave Sleep
TN	True Negative
TP	True Positive
tROC	Test Receiving Operating Characteristics
TRT	Total Recording Time
TRT	Total Recording Time
TST	Total Sleep Time
VA	Veteran Affairs (Department Of Veteran Affairs)
VAT	Variable Amplitude Thresholding
VEOG	Vertical Electroencephalography
WASM	World Association Of Sleep Medicine
WASO	Wake After Sleep Onset
WASO	Wake After Sleep Onset
WSC	Wisconsin Sleep Cohort

LIST OF REFERENCES

[1]   K. A. Rechtschaffen A, editors. Los Angeles: Brain Information Service/Brain Research Institute, "A manual of standardized terminology, techniques and scoring system of sleep stages in human subjects," ed: University of California, 1968.

[2]   A.-I. S. Iber C, Chesson A, Quan SF, "The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications," American Academy of Sleep Medicine, Westchester, Ill2007.

[3]   S. a. B. Noachtar, C and Ebersole, J and Mauguiere, F and Sakamoto, A and Westmoreland, B, "A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings. The International Federation of Clinical Neurophysiology," *Electroencephalography and clinical neurophysiology. Supplement,* vol. 52, p. 21, 1999.

[4]   W. R. Ruehland, P. D. Rochford, F. J. O'Donoghue, R. J. Pierce, P. Singh, and A. T. Thornton, "The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index," *Sleep,* vol. 32, pp. 150-7, Feb 2009.

[5]   E. Mignot, G. J. Lammers, B. Ripley, M. Okun, S. Nevsimalova, S. Overeem, J. Vankova, J. Black, J. Harsh, C. Bassetti, H. Schrader, and S. Nishino, "The role of cerebrospinal fluid hypocretin measurement in the diagnosis of narcolepsy and other hypersomnias," *Archives of neurology,* vol. 59, pp. 1553-62, Oct 2002.

[6]   M. A. Carskadon, W. C. Dement, M. M. Mitler, T. Roth, P. R. Westbrook, and S. Keenan, "Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness," *Sleep,* vol. 9, pp. 519-24, Dec 1986.

[7]   R. Ferri, M. Zucconi, M. Manconi, G. Plazzi, O. Bruni, and L. Ferini-Strambi, "New approaches to the study of periodic leg movements during sleep in restless legs syndrome," *Sleep,* vol. 29, pp. 759-69, Jun 2006.

[8]   H. Scofield, T. Roth, and C. Drake, "Periodic limb movements during sleep: population prevalence, clinical correlates, and racial differences," *Sleep,* vol. 31, pp. 1221-7, Sep 2008.

[9]   A. S. Walters and D. B. Rye, "Review of the relationship of restless legs syndrome and periodic limb movements in sleep to hypertension, heart disease, and stroke," *Sleep,* vol. 32, pp. 589-97, May 2009.

[10]   Y. Li, W. Wang, J. W. Winkelman, A. Malhotra, J. Ma, and X. Gao, "Prospective study of restless legs syndrome and mortality among men," *Neurology,* vol. 81, pp. 52-9, Jul 2 2013.

[11]   D. B. Boivin, J. Montplaisir, and G. Poirier, "The effects of L-dopa on periodic leg movements and sleep organization in narcolepsy," *Clinical neuropharmacology,* vol. 12, pp. 339-45, Aug 1989.

[12] O. Lapierre and J. Montplaisir, "Polysomnographic features of REM sleep behavior disorder: development of a scoring method," *Neurology,* vol. 42, pp. 1371-4, Jul 1992.

[13] S. Ancoli-Israel, D. F. Kripke, W. Mason, and O. J. Kaplan, "Sleep apnea and periodic movements in an aging sample," *Journal of gerontology,* vol. 40, pp. 419-25, Jul 1985.

[14] T. C. Wetter, V. Collado-Seidel, T. Pollmacher, A. Yassouridis, and C. Trenkwalder, "Sleep and periodic leg movement patterns in drug-free patients with Parkinson's disease and multiple system atrophy," *Sleep,* vol. 23, pp. 361-7, May 1 2000.

[15] J. Winkelmann, D. Czamara, B. Schormair, F. Knauf, E. C. Schulte, C. Trenkwalder, Y. Dauvilliers, O. Polo, B. Hogl, K. Berger, A. Fuhs, N. Gross, K. Stiasny-Kolster, W. Oertel, C. G. Bachmann, W. Paulus, L. Xiong, J. Montplaisir, G. A. Rouleau, I. Fietze, J. Vavrova, D. Kemlink, K. Sonka, S. Nevsimalova, S. C. Lin, Z. Wszolek, C. Vilarino-Guell, M. J. Farrer, V. Gschliesser, B. Frauscher, T. Falkenstetter, W. Poewe, R. P. Allen, C. J. Earley, W. G. Ondo, W. D. Le, D. Spieler, M. Kaffe, A. Zimprich, J. Kettunen, M. Perola, K. Silander, I. Cournu-Rebeix, M. Francavilla, C. Fontenille, B. Fontaine, P. Vodicka, H. Prokisch, P. Lichtner, P. Peppard, J. Faraco, E. Mignot, C. Gieger, T. Illig, H. E. Wichmann, B. Muller-Myhsok, and T. Meitinger, "Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1," *PLoS genetics,* vol. 7, p. e1002171, Jul 2011.

[16] A. Y. Samson, S. Bensen, A. Beck, D. Price, and C. Nimmer, "Posttraumatic stress disorder in primary care," *The Journal of family practice,* vol. 48, pp. 222-7, Mar 1999.

[17] R. C. Kessler, A. Sonnega, E. Bromet, M. Hughes, and C. B. Nelson, "Posttraumatic stress disorder in the National Comorbidity Survey," *Archives of general psychiatry,* vol. 52, pp. 1048-60, Dec 1995.

[18] W. E. Copeland, G. Keeler, A. Angold, and E. J. Costello, "Traumatic events and posttraumatic stress in childhood," *Archives of general psychiatry,* vol. 64, pp. 577-84, May 2007.

[19] H. S. Resnick, D. G. Kilpatrick, B. S. Dansky, B. E. Saunders, and C. L. Best, "Prevalence of civilian trauma and posttraumatic stress disorder in a representative national sample of women," *Journal of consulting and clinical psychology,* vol. 61, pp. 984-91, Dec 1993.

[20] F. H. Norris, "Epidemiology of trauma: frequency and impact of different potentially traumatic events on different demographic groups," *Journal of consulting and clinical psychology,* vol. 60, pp. 409-18, Jun 1992.

[21] N. Breslau, R. C. Kessler, H. D. Chilcoat, L. R. Schultz, G. C. Davis, and P. Andreski, "Trauma and posttraumatic stress disorder in the community: the 1996 Detroit Area Survey of Trauma," *Archives of general psychiatry,* vol. 55, pp. 626-32, Jul 1998.

[22] A. Germain, M. K. Shear, M. Hall, and D. J. Buysse, "Effects of a brief behavioral treatment for PTSD-related sleep disturbances: a pilot study," *Behaviour research and therapy,* vol. 45, pp. 627-32, Mar 2007.

[23] H. Kang. (2006). *VA health care utilization among Operation Iraqi Freedom / Operation Enduring Freedom Veterans* [electronic]. Available: http://www.iom.edu

[24] N. Breslau, G. C. Davis, P. Andreski, and E. Peterson, "Traumatic events and posttraumatic stress disorder in an urban population of young adults," *Archives of general psychiatry,* vol. 48, pp. 216-22, Mar 1991.

[25] N. Breslau, "The epidemiology of posttraumatic stress disorder: what is the extent of the problem?," *The Journal of clinical psychiatry,* vol. 62 Suppl 17, pp. 16-22, 2001.

[26] D. F. Zatzick, C. R. Marmar, D. S. Weiss, W. S. Browner, T. J. Metzler, J. M. Golding, A. Stewart, W. E. Schlenger, and K. B. Wells, "Posttraumatic stress disorder and functioning and quality of life outcomes in a nationally representative sample of male Vietnam veterans," *The American journal of psychiatry,* vol. 154, pp. 1690-5, Dec 1997.

[27] L. Amaya-Jackson, J. R. Davidson, D. C. Hughes, M. Swartz, V. Reynolds, L. K. George, and D. G. Blazer, "Functional impairment and utilization of services associated with posttraumatic stress in the community," *Journal of traumatic stress,* vol. 12, pp. 709-24, Oct 1999.

[28] S. D. Solomon and J. R. Davidson, "Trauma: prevalence, impairment, service use, and cost," *The Journal of clinical psychiatry,* vol. 58 Suppl 9, pp. 5-11, 1997.

[29] K. A. Babson and M. T. Feldner, "Temporal relations between sleep problems and both traumatic event exposure and PTSD: a critical review of the empirical literature," *Journal of anxiety disorders,* vol. 24, pp. 1-15, Jan 2010.

[30] A. Hefez, L. Metz, and P. Lavie, "Long-term effects of extreme situational stress on sleep and dreaming," *The American journal of psychiatry,* vol. 144, pp. 344-7, Mar 1987.

[31] V. I. Spoormaker and P. Montgomery, "Disturbed sleep in post-traumatic stress disorder: secondary symptom or core feature?," *Sleep medicine reviews,* vol. 12, pp. 169-84, Jun 2008.

[32] A. P. Association, "Diagnostic and statistical manual of mental disorders (DSM-IV)," ed. Washington, DC: American Psychiatric Association, 1994.

[33] G. Pillar, A. Malhotra, and P. Lavie, "Post-traumatic stress disorder and sleep-what a nightmare!," *Sleep medicine reviews,* vol. 4, pp. 183-200, Apr 2000.

[34] M. A. Schuster, B. D. Stein, L. Jaycox, R. L. Collins, G. N. Marshall, M. N. Elliott, A. J. Zhou, D. E. Kanouse, J. L. Morrison, and S. H. Berry, "A national survey of stress reactions after the September 11, 2001, terrorist attacks," *The New England journal of medicine,* vol. 345, pp. 1507-12, Nov 15 2001.

[35] M. M. Ohayon and C. M. Shapiro, "Sleep disturbances and psychiatric disorders associated with posttraumatic stress disorder in the general population," *Comprehensive psychiatry,* vol. 41, pp. 469-78, Nov-Dec 2000.

[36]   A. Germain, M. Hall, B. Krakow, M. Katherine Shear, and D. J. Buysse, "A brief sleep scale for Posttraumatic Stress Disorder: Pittsburgh Sleep Quality Index Addendum for PTSD," *Journal of anxiety disorders,* vol. 19, pp. 233-44, 2005.

[37]   B. Krakow, D. Melendrez, B. Pedersen, L. Johnston, M. Hollifield, A. Germain, M. Koss, T. D. Warner, and R. Schrader, "Complex insomnia: insomnia and sleep-disordered breathing in a consecutive series of crime victims with nightmares and PTSD," *Biological psychiatry,* vol. 49, pp. 948-53, Jun 1 2001.

[38]   B. Krakow, A. Artar, T. D. Warner, D. Melendrez, L. Johnston, M. Hollifield, A. Germain, and M. Koss, "Sleep disorder, depression, and suicidality in female sexual assault survivors," *Crisis,* vol. 21, pp. 163-170, 2000.

[39]   B. Krakow, D. Melendrez, L. Johnston, T. D. Warner, J. O. Clark, M. Pacheco, B. Pedersen, M. Koss, M. Hollifield, and R. Schrader, "Sleep-disordered breathing, psychiatric distress, and quality of life impairment in sexual assault survivors," *The Journal of nervous and mental disease,* vol. 190, pp. 442-52, Jul 2002.

[40]   G. A. Clum, P. Nishith, and P. A. Resick, "Trauma-related sleep disturbance and self-reported physical health symptoms in treatment-seeking female rape victims," *The Journal of nervous and mental disease,* vol. 189, pp. 618-22, Sep 2001.

[41]   M. E. Saladin, K. T. Brady, B. S. Dansky, and D. G. Kilpatrick, "Understanding comorbidity between PTSD and substance use disorders: two preliminary investigations," *Addictive behaviors,* vol. 20, pp. 643-55, Sep-Oct 1995.

[42]   P. Nishith, P. A. Resick, and K. T. Mueser, "Sleep difficulties and alcohol use motives in female rape victims with posttraumatic stress disorder," *Journal of traumatic stress,* vol. 14, pp. 469-79, Jul 2001.

[43]   T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *The New England journal of medicine,* vol. 328, pp. 1230-5, Apr 29 1993.

[44]   O. M. Andlauer, H.; Jouhier,L.; Drake, C.; Peppard,P.; Han, F.; Hong, S.; Poli, F.; Plazzi, G.; O'Hara, R.; Haffen, E.; Roth, T.; Young, T.; Mignot, E. , "Nocturnal REM sleep latency for identifying patients with narcolepsy/hypocretin deficiency.," *JAMA Neurology,* 2013.

[45]   in *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*, H. R. Colten and B. M. Altevogt, Eds., ed Washington (DC), 2006.

[46]   H. C. Kraemer, *Evaluating Medical Tests*. Newbury Park: Sage Publications, 1992.

[47]   C. Stephan, S. Wesseling, T. Schink, and K. Jung, "Comparison of eight computer programs for receiver-operating characteristic analysis," *Clinical chemistry,* vol. 49, pp. 433-9, Mar 2003.

[48]   J. Savage, "ROC5," ed: Department of Veterean's Affairs and the National Institute of Aging of the United States, 2008, pp. Decision based receiver-operating characteristics software program.

[49]   D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Machine Learning,* vol. 45, pp. 171-186, 2001.

[50]   O. Andlauer, H. t. Moore, S. C. Hong, Y. Dauvilliers, T. Kanbayashi, S. Nishino, F. Han, M. H. Silber, T. Rico, M. Einen, B. R. Kornum, P. Jennum, S. Knudsen, S. Nevsimalova, F. Poli, G. Plazzi, and E. Mignot, "Predictors of hypocretin (orexin) deficiency in narcolepsy without cataplexy," *Sleep,* vol. 35, pp. 1247-55F, Sep 2012.

[51]   K. K. Dobbin and R. M. Simon, "Optimally splitting cases for training and testing high dimensional classifiers," *BMC medical genomics,* vol. 4, p. 31, 2011.

[52]   R. Simon, M. D. Radmacher, K. Dobbin, and L. M. McShane, "Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification," *Journal of the National Cancer Institute,* vol. 95, pp. 14-18, Jan 1 2003.

[53]   G. F. Cooper, C. F. Aliferis, R. Ambrosino, J. Aronis, B. G. Buchanan, R. Caruana, M. J. Fine, C. Glymour, G. Gordon, B. H. Hanusa, J. E. Janosky, C. Meek, T. Mitchell, T. Richardson, and P. Spirtes, "An evaluation of machine-learning methods for predicting pneumonia mortality," *Artificial intelligence in medicine,* vol. 9, pp. 107-38, Feb 1997.

[54]   S. Sonnenburg, M. L. Braun, C. S. Ong, S. Bengio, L. Bottou, G. Holmes, Y. LeCun, K. R. Muller, F. Pereira, C. E. Rasmussen, G. Ratsch, B. Scholkopf, A. Smola, P. Vincent, J. Weston, and R. C. Williamson, "The need for open source software in machine learning," *Journal of Machine Learning Research,* vol. 8, pp. 2443-2466, Oct 2007.

[55]   T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters,* vol. 27, pp. 861-874, Jun 2006.

[56]   R. Oostenveld, P. Fries, E. Maris, and J. M. Schoffelen, "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Computational intelligence and neuroscience,* vol. 2011, p. 156869, 2011.

[57]   A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods,* vol. 134, pp. 9-21, Mar 15 2004.

[58]   V. D. V. Gerla, L. Lhotska, V. Krajca, "PSGLab Matlab Toolbox for Polysomnographic Data Processing: Development and Practical Application," in *10th Internation Conference on Information Technology and Applications in Biomedicine*, 2010.

[59]   I. The MathWorks. (2011). *MATLAB -- The Language of Technical Computing*.

[60]   S. H. Woodward, M. J. Friedman, and D. L. Bliwise, "Sleep and depression in combat-related PTSD inpatients," *Biological psychiatry,* vol. 39, pp. 182-92, Feb 1 1996.

[61]   B. Kemp, "European data format (EDF): current availability and additional applications," *Journal Sleep Research,* vol. 11, p. 120, 2002.

[62]   H. Moore, E. Leery, S. Lee, O. Carrillo, R. Stubbs, P. Peppard, T. Young, and E. Mignot, "Design and Validation of a Periodic Leg Movement detector " *SLEEP,* 2013 (submitted).

[63]   P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *Audio and Electroacoustics, IEEE Transactions on,* vol. 15, pp. 70-73, 1967.

[64]   R. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on,* vol. 34, pp. 276-280, 1986.

[65]   D. P. Brunner, R. C. Vasko, C. S. Detka, J. P. Monahan, C. F. Reynolds, 3rd, and D. J. Kupfer, "Muscle artifacts in the sleep EEG: automated detection and effect on all-night EEG power spectra," *Journal of sleep research,* vol. 5, pp. 155-64, Sep 1996.

[66]   R. Ferri, M. Zucconi, M. Manconi, O. Bruni, S. Miano, G. Plazzi, and L. Ferini-Strambi, "Computer-assisted detection of nocturnal leg motor activity in patients with restless legs syndrome and periodic leg movements during sleep," *Sleep,* vol. 28, pp. 998-1004, Aug 1 2005.

[67]   N. P. T. Tauchmann, "Automatic Detection of Periodic Leg Movements," *Journal of sleep research,* vol. 5, pp. 273-275, 1996.

[68]   T. C. Wetter, G. Dirlich, J. Streit, C. Trenkwalder, A. Schuld, and T. Pollmacher, "An automatic method for scoring leg movements in polygraphic sleep recordings and its validity in comparison to visual scoring," *Sleep,* vol. 27, pp. 324-8, Mar 15 2004.

[69]   I. S. Gopal and G. G. Haddad, "Automatic detection of eye movements in REM sleep using the electrooculogram," *The American journal of physiology,* vol. 241, pp. R217-21, Sep 1981.

[70]   R. J. McPartland and D. J. Kupfer, "Computerised measures of electro-oculographic activity during sleep," *International journal of bio-medical computing,* vol. 9, pp. 409-19, Nov 1978.

[71]   K. Takahashi and Y. Atsumi, "Precise measurement of individual rapid eye movements in REM sleep of humans," *Sleep,* vol. 20, pp. 743-52, Sep 1997.

[72]   X. Tan, I. G. Campbell, and I. Feinberg, "A simple method for computer quantification of stage REM eye movement potentials," *Psychophysiology,* vol. 38, pp. 512-6, May 2001.

[73]   E. Magosso, F. Provini, P. Montagna, and M. Ursino, "A wavelet based method for automatic detection of slow eye movements: a pilot study," *Medical engineering & physics,* vol. 28, pp. 860-75, Nov 2006.

[74]   F. Ferrarelli, R. Huber, M. J. Peterson, M. Massimini, M. Murphy, B. A. Riedner, A. Watson, P. Bria, and G. Tononi, "Reduced sleep spindle activity in schizophrenia patients," *The American journal of psychiatry,* vol. 164, pp. 483-92, Mar 2007.

[75]   J. Carrier, I. Viens, G. Poirier, R. Robillard, M. Lafortune, G. Vandewalle, N. Martin, M. Barakat, J. Paquet, and D. Filipini, "Sleep slow wave changes during the middle years of life," *The European journal of neuroscience,* vol. 33, pp. 758-66, Feb 2011.

[76]   P. He, G. Wilson, and C. Russell, "Removal of ocular artifacts from electro-encephalogram by adaptive filtering," *Medical & biological engineering & computing,* vol. 42, pp. 407-12, May 2004.

[77]   H. Moore, "Visualization of EEG activity for stimulating sleep research," *Computer Methods in Biomechanics and Biomedical Engineering: Imagine & Visualization,* 2013.

[78]   S. M. Warby, HE; Carrillo, O; Faraco, J; Lin, L; Peppard, PE; Young, T; Mignot, E, "Genome Wide Association Study and Confounders of Sigma Power and Sleep Spindles," *Sleep,* vol. 35, 2012.

[79]   H. W. Moore, S. ; Mignot E., "High resolution detection of polysomnography based phasic events of REM sleep in postraumatic stress disorder," *Sleep,* vol. 35, 2012.

[80]   H. W. Moore, S; Woodward, S; Peppard, P; Young, T; Mignot, E, "Leg movement detection software and period leg movement index calculator," *Sleep,* vol. 35, 2012.

[81]   E. W. Agustsson, S; Welinder, P; Carrillo, O; Moore, HE; Mignot, E; Perona, I, "Reliability of Sleep Spindle Identification by Experts non-Experts and Automated Methods," *Sleep,* vol. 34, 2011.

[82]   S. C. Warby, O; Moore, HE; Kawashima, M; Apple, R; Faraco, J; Lin, L; Peppard, PE; Young, T; Mignot E, "Genetic Analysis of Sleep Duration," *Sleep,* vol. 34, 2011.

[83]   L. De Gennaro, C. Marzano, F. Fratello, F. Moroni, M. C. Pellicciari, F. Ferlazzo, S. Costa, A. Couyoumdjian, G. Curcio, E. Sforza, A. Malafosse, L. A. Finelli, P. Pasqualetti, M. Ferrara, M. Bertini, and P. M. Rossini, "The Electroencephalographic Fingerprint of Sleep Is Genetically Determined: A Twin Study," *Annals of neurology,* vol. 64, pp. 455-460, Oct 2008.

[84]   J. Buckelmuller, H. P. Landolt, H. H. Stassen, and P. Achermann, "Trait-like individual differences in the human sleep electroencephalogram," *Neuroscience,* vol. 138, pp. 351-6, 2006.

[85]   L. De Gennaro, M. Ferrara, F. Vecchio, G. Curcio, and M. Bertini, "An electroencephalographic fingerprint of human sleep," *NeuroImage,* vol. 26, pp. 114-22, May 15 2005.

[86]   H. W. Moore, S; Mignot E, "The Stanford EEG Viewer: A High-Throughput Platform for the Visualization and Analysis of Sleep Data," presented at the Biomedical Computation at Stanford, 2010, Stanford, CA, 2010.

[87]   F. J. Langheim, M. Murphy, B. A. Riedner, and G. Tononi, "Functional connectivity in slow-wave sleep: identification of synchronous cortical activity during wakefulness and sleep using time series analysis of electroencephalographic data," *Journal of sleep research,* vol. 20, pp. 496-505, Dec 2011.

[88]   G. M. a. F. Mang, Paul, "Sleep and EEG phenotyping in mice," *Current Protocols in Mouse Biology,* pp. 55-74, 2012.

[89]   M. S. Scher and K. A. Loparo, "Neonatal EEG/sleep state analyses: a complex phenotype of developmental neural plasticity," *Developmental neuroscience,* vol. 31, pp. 259-75, 2009.

[90]   C. J. Davis, J. M. Clinton, K. A. Jewett, M. R. Zielinski, and J. M. Krueger, "Delta wave power: an independent sleep phenotype or epiphenomenon?," *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine,* vol. 7, pp. S16-8, Oct 15 2011.

[91]   I. Feinberg, G. Fein, and T. C. Floyd, "Period and amplitude analysis of NREM EEG in sleep: repeatability of results in young adults," *Electroencephalography and clinical neurophysiology,* vol. 48, pp. 212-21, Feb 1980.

[92]   X. Tan, I. G. Campbell, L. Palagini, and I. Feinberg, "High internight reliability of computer-measured NREM delta, sigma, and beta: biological implications," *Biological psychiatry,* vol. 48, pp. 1010-9, Nov 15 2000.

[93]   U. Ambrosius, S. Lietzenmaier, R. Wehrle, A. Wichniak, S. Kalus, J. Winkelmann, T. Bettecken, F. Holsboer, A. Yassouridis, and E. Friess, "Heritability of sleep electroencephalogram," *Biological psychiatry,* vol. 64, pp. 344-8, Aug 15 2008.

[94]   R. Godbout, C. Bergeron, E. Limoges, E. Stip, and L. Mottron, "A laboratory study of sleep in Asperger's syndrome," *Neuroreport,* vol. 11, pp. 127-30, Jan 17 2000.

[95]   O. Bruni, R. Ferri, L. Novelli, M. Terribili, M. Troianiello, E. Finotti, V. Leuzzi, and P. Curatolo, "Sleep Spindle Activity Is Correlated With Reading Abilities in Developmental Dyslexia," *Sleep,* vol. 32, pp. 1333-1340, Oct 1 2009.

[96]   W. C., "Information Visualization: Perception for Design," 2nd ed: Morgan Kaufmann/Academic Press, 2004.

[97]   B. E. a. T. Rogowitz, L.A. and Bryson, S., "How not to lie with visualization," *Computers in Physics*

, vol. 10, pp. 268--273, 1996.

[98]   M. D. Plumlee and C. Ware, "Zooming Versus Multiple Window Interfaces: Cognitive Costs of Visual Comparisons," *ACM Transactions on Computer-Human Interaction (TOCHI),* vol. 13, pp. 179-209, 2006.

[99]   T. Munzner, "Process and pitfalls in writing information visualization research papers

," *Information visualization,* pp. 134--153, 2008.

[100]   L. T. Fernholz, S. Morgenthaler, J. W. Tukey, and E. Tukey, "A conversation with John W. Tukey and Elizabeth Tukey," *Statistical Science,* vol. 15, pp. 79-94, Feb 2000.

[101]   M. Tory, A. E. Kirkpatrick, M. S. Atkins, and T. Moller, "Visualization task performance with 2D, 3D, and combination displays," *Ieee Transactions on Visualization and Computer Graphics,* vol. 12, pp. 2-13, Jan-Feb 2006.

[102] AtlasTaskForce, "Recording and scoring leg movements.," *Sleep,* vol. 16, pp. 748-59, Dec 1993.

[103] J. W. Winkelman, L. Finn, and T. Young, "Prevalence and correlates of restless legs syndrome symptoms in the Wisconsin Sleep Cohort," *Sleep medicine,* vol. 7, pp. 545-52, Oct 2006.

[104] R. P. Allen, D. Picchietti, W. A. Hening, C. Trenkwalder, A. S. Walters, and J. Montplaisi, "Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health," *Sleep medicine,* vol. 4, pp. 101-19, Mar 2003.

[105] U. o. Wisconsin-Madison, "Policies & Procedures Sleep Cohort Study: Scoring Guidelines," April, 1995 1995.

[106] M. Kubat, R. C. Holte, and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," *Machine Learning,* vol. 30, pp. 195-215, Feb-Mar 1998.

[107] B. Widrow, J. R. Glover, J. M. Mccool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin, "Adaptive Noise Cancelling - Principles and Applications," *Proceedings of the Ieee,* vol. 63, pp. 1692-1716, 1975.

[108] M. Zucconi, R. Ferri, R. Allen, P. C. Baier, O. Bruni, S. Chokroverty, L. Ferini-Strambi, S. Fulda, D. Garcia-Borreguero, W. A. Hening, M. Hirshkowitz, B. Hogl, M. Hornyak, M. King, P. Montagna, L. Parrino, G. Plazzi, and M. G. Terzano, "The official World Association of Sleep Medicine (WASM) standards for recording and scoring periodic leg movements in sleep (PLMS) and wakefulness (PLMW) developed in collaboration with a task force from the International Restless Legs Syndrome Study Group (IRLSSG)," *Sleep medicine,* vol. 7, pp. 175-83, Mar 2006.

[109] C. Trenkwalder, W. A. Hening, A. S. Walters, S. S. Campbell, K. Rahman, and S. Chokroverty, "Circadian rhythm of periodic limb movements and sensory symptoms of restless legs syndrome," *Movement disorders : official journal of the Movement Disorder Society,* vol. 14, pp. 102-10, Jan 1999.

[110] K. P. Parker and D. B. Rye, "Restless legs syndrome and periodic limb movement disorder," *The Nursing clinics of North America,* vol. 37, pp. 655-73, Dec 2002.

[111] R. Ferri, M. Zucconi, F. Rundo, K. Spruyt, M. Manconi, and L. Ferini-Strambi, "Heart rate and spectral EEG changes accompanying periodic and non-periodic leg movements during sleep," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology,* vol. 118, pp. 438-48, Feb 2007.

[112] J. W. Winkelman, "The evoked heart rate response to periodic leg movements of sleep," *Sleep,* vol. 22, pp. 575-80, Aug 1 1999.

[113] C. Medigue, C. Vermeiren, P. Bourgin, C. Debouzy, and P. Escourrou, "Cardiovascular perturbations involved in periodic leg movements during sleep," *Computers in Cardiology 1995,* pp. 477-480, 1995.

[114] B. Schormair, J. Plag, M. Kaffe, N. Gross, D. Czamara, W. Samtleben, P. Lichtner, A. Strohle, I. Stefanidis, A. Vainas, E. Dardiotis, G. K. Sakkas, C. Gieger, B. Muller-Myhsok, T. Meitinger, U. Heemann, G. M. Hadjigeorgiou, K. Oexle, and J. Winkelmann, "MEIS1 and BTBD9: genetic association with restless leg syndrome in end stage renal disease," *Journal of medical genetics*, vol. 48, pp. 462-6, Jul 2011.

[115] H. Stefansson, D. B. Rye, A. Hicks, H. Petursson, A. Ingason, T. E. Thorgeirsson, S. Palsson, T. Sigmundsson, A. P. Sigurdsson, I. Eiriksdottir, E. Soebech, D. Bliwise, J. M. Beck, A. Rosen, S. Waddy, L. M. Trotti, A. Iranzo, M. Thambisetty, G. A. Hardarson, K. Kristjansson, L. J. Gudmundsson, U. Thorsteinsdottir, A. Kong, J. R. Gulcher, D. Gudbjartsson, and K. Stefansson, "A genetic risk factor for periodic limb movements in sleep," *The New England journal of medicine*, vol. 357, pp. 639-47, Aug 16 2007.

[116] A. S. Walters, "Toward a better definition of the restless legs syndrome. The International Restless Legs Syndrome Study Group," *Movement disorders : official journal of the Movement Disorder Society*, vol. 10, pp. 634-42, Sep 1995.

[117] E. L. Hyatt Moore IV, Seo-Young Lee, Oscar Carrillo, Robin Stubbs, Paul Peppard, Terry Young, Bernard Widrow, Emmanuel Mignot, "Design and Validation of a Periodic Leg Movement detector," *Sleep*, 2013.

[118] J. S. Sarah J. Ratcliffe, "GEEQBOX: A MATLAB Toolbox for Generalized Estimating Equations and Quasi-Least Squares," *Journal of statistical software*, vol. 25, 2008.

[119] R. K. Singareddy and R. Balon, "Sleep in posttraumatic stress disorder," *Annals of clinical psychiatry : official journal of the American Academy of Clinical Psychiatrists*, vol. 14, pp. 183-90, Sep 2002.

[120] T. A. Mellman, V. Bustamante, A. I. Fins, W. R. Pigeon, and B. Nolan, "REM sleep and the early development of posttraumatic stress disorder," *The American journal of psychiatry*, vol. 159, pp. 1696-701, Oct 2002.

[121] R. J. Ross, W. A. Ball, K. A. Sullivan, and S. N. Caroff, "Sleep disturbance as the hallmark of posttraumatic stress disorder," *The American journal of psychiatry*, vol. 146, pp. 697-707, Jun 1989.

[122] T. A. Mellman, A. Kumar, R. Kulick-Bell, M. Kumar, and B. Nolan, "Nocturnal/daytime urine noradrenergic measures and sleep in combat-related PTSD," *Biological psychiatry*, vol. 38, pp. 174-9, Aug 1 1995.

[123] R. Stickgold, "EMDR: a putative neurobiological mechanism of action," *Journal of clinical psychology*, vol. 58, pp. 61-75, Jan 2002.

[124] L. Wittmann, M. Schredl, and M. Kramer, "Dreaming in posttraumatic stress disorder: A critical review of phenomenology, psychophysiology and treatment," *Psychotherapy and psychosomatics*, vol. 76, pp. 25-39, 2007.

[125] E. Fukuma and T. Okuma, "[Dream and mental function--psychological development and dreams in children]," *Saishin igaku. Modern medicine*, vol. 26, pp. 35-40, Jan 1971.

[126] T. A. Mellman and M. M. Hipolito, "Sleep disturbances in the aftermath of trauma and posttraumatic stress disorder," *CNS spectrums,* vol. 11, pp. 611-5, Aug 2006.

[127] T. A. Mellman, R. Kulick-Bell, L. E. Ashlock, and B. Nolan, "Sleep events among veterans with combat-related posttraumatic stress disorder," *The American journal of psychiatry,* vol. 152, pp. 110-5, Jan 1995.

[128] T. A. Mellman, W. R. Pigeon, P. D. Nowell, and B. Nolan, "Relationships between REM sleep findings and PTSD symptoms during the early aftermath of trauma," *Journal of traumatic stress,* vol. 20, pp. 893-901, Oct 2007.

[129] E. Aserinsky and N. Kleitman, "Regularly occurring periods of eye motility, and concomitant phenomena, during sleep," *Science,* vol. 118, pp. 273-4, Sep 4 1953.

[130] E. Aserinsky and N. Kleitman, "Two types of ocular motility occurring in sleep," *Journal of applied physiology,* vol. 8, pp. 1-10, Jul 1955.

[131] K. Uenoyama, N. Uenoyama, and I. Iinuma, "Vector-Electro-Oculography and Its Clinical Application. Two-Dimensional Recording of Eye Movements," *The British journal of ophthalmology,* vol. 48, pp. 318-29, Jun 1964.

[132] D. Schneider, "Spatiotemporal properties of rapid eye movements in human REM sleep 1. Qualitative analysis," *Waking & Sleeping,* vol. 2, pp. 63-67, 1978.

[133] D. Schneider, "Spatiotemporal properties of rapid eye movements in human REM sleep. 2. Quantitative analysis," *Waking & Sleeping,* vol. 2, pp. 63-67, 1978.

[134] V. Gabersek and H. Ghiloni, "[Electrooculographic study of eye movements during the paradoxical phase of sleep]," *Revue neurologique,* vol. 123, pp. 251-5, Oct 1970.

[135] V. Gabersek, "[Eye movements during sleep]," *Revue d'oto-neuro-ophtalmologie,* vol. 44, pp. 69-78, Jan-Feb 1972.

[136] D. L. Donoho, "De-Noising by Soft-Thresholding," *Ieee Transactions on Information Theory,* vol. 41, pp. 613-627, May 1995.

[137] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage - Asymptopia," *Journal of the Royal Statistical Society Series B-Methodological,* vol. 57, pp. 301-337, 1995.

[138] W. S. Agustsson EP, Welinder P, Carrillo O, Moore HE, Mignot E, Perona I, "Reliability of sleep spindle identification by experts, non-experts and automated methods," *Sleep,* 2011.

[139] S. L. Lima, N. C. Rattenborg, J. A. Lesku, and C. J. Amlaner, "Sleeping under the risk of predation," *Animal Behaviour,* vol. 70, pp. 723-736, Oct 2005.

[140] R. Armitage and R. F. Hoffmann, "Sleep EEG, depression and gender," *Sleep medicine reviews,* vol. 5, pp. 237-246, Jun 2001.

[141] J. S. Jeong, "EEG dynamics in patients with Alzheimer's disease," *Clinical Neurophysiology,* vol. 115, pp. 1490-1505, Jul 2004.

[142] P. J. Uhlhaas and W. Singer, "Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology," *Neuron,* vol. 52, pp. 155-68, Oct 5 2006.

[143] V. Sakkalis, T. Oikonomou, E. Pachou, I. Tollis, S. Micheloyannis, and M. Zervakis, "Time-significant wavelet coherence for the evaluation of schizophrenic brain activity using a graph theory approach," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference,* vol. 1, pp. 4265-8, 2006.

[144] J. Dauwels, F. Vialatte, T. Musha, and A. Cichocki, "A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG," *NeuroImage,* vol. 49, pp. 668-93, Jan 1 2010.

[145] C. W. Therrien, *Discrete random signals and statistical signal processing.* Englewood Cliffs, NJ: Prentice Hall, 1992.

[146] G. Fein, J. Raz, F. F. Brown, and E. L. Merrin, "Common Reference Coherence Data Are Confounded by Power and Phase Effects," *Electroencephalography and clinical neurophysiology,* vol. 69, pp. 581-584, Jun 1988.