University of Kentucky

UKnowledge

2015

EXPERIMENTAL-COMPUTATIONAL ANALYSIS OF VIGILANCE DYNAMICS FOR APPLICATIONS IN SLEEP AND EPILEPSY

Farid Yaghouby
University of Kentucky, farid.yaghouby@gmail.com

Right click to open a feedback form in a new tab to let us know how this document benefits you.

EXPERIMENTAL-COMPUTATIONAL ANALYSIS OF VIGILANCE DYNAMICS
FOR APPLICATIONS IN SLEEP AND EPILEPSY

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the College of Engineering at the University of Kentucky

By

Farid Yaghouby

Lexington, KY

Director: Dr. Sridhar Sunderam, Assistant Professor of Biomedical Engineering

Lexington, Kentucky

2015

ABSTRCAT OF DISSERTATION

EXPERIMENTAL-COMPUTATIONAL ANALYSIS OF VIGILANCE DYNAMICS FOR APPLICATIONS IN SLEEP AND EPILEPSY

Epilepsy is a neurological disorder characterized by recurrent seizures. Sleep problems can cooccur with epilepsy, and adversely affect seizure diagnosis and treatment. In fact, the relationship between sleep and seizures in individuals with epilepsy is a complex one. Seizures disturb sleep and sleep deprivation aggravates seizures. Antiepileptic drugs may also impair sleep quality at the cost of controlling seizures. In general, particular vigilance states may inhibit or facilitate seizure generation, and changes in vigilance state can affect the predictability of seizures.

A clear understanding of sleep-seizure interactions will therefore benefit epilepsy care providers and improve quality of life in patients. Notable progress in neuroscience research—and particularly sleep and epilepsy—has been achieved through experimentation on animals. Experimental models of epilepsy provide us with the opportunity to explore or even manipulate the sleep-seizure relationship in order to decipher different aspects of their interactions. Important in this process is the development of techniques for modeling and tracking sleep dynamics using electrophysiological measurements.

In this dissertation experimental and computational approaches are proposed for modeling vigilance dynamics and their utility demonstrated in nonepileptic control mice. The general framework of hidden Markov models is used to automatically model and track sleep state and dynamics from electrophysiological as well as novel motion measurements. In addition, a closed-loop sensory stimulation technique is proposed that, in conjunction with this model, provides the means to concurrently track and modulate vigilance dynamics in animals.

The feasibility of the proposed techniques for modeling and altering sleep are demonstrated for experimental applications related to epilepsy. Finally, preliminary data from a mouse model of temporal lobe epilepsy are employed to suggest applications of these techniques and directions for future research. The methodologies developed here have clear implications the design of intelligent neuromodulation strategies for clinical epilepsy therapy.

KEYWORDS: Sleep, Epilepsy, Vigilance dynamics, Hidden Markov model, EEG

Farid Yaghouby
_____

August 3, 2015
_____

EXPERIMENTAL-COMPUTATIONAL ANALYSIS OF VIGILANCE DYNAMICS

FOR APPLICATIONS IN SLEEP AND EPILEPSY

By

Farid Yaghouby

Dr. Sridhar Sunderam
_____

Director of Dissertation

Dr. Abhijit Patwardhan
_____

Director of Graduate Studies

August 3, 2015
_____

 Date

DEDICATION

I dedicate my dissertation work to my family. A special feeling of gratitude to my lovely wife, Reihaneh and our dear parents whose words of encouragement and push for tenacity ring in my ears. I could not have achieved this without their endless support and love.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. CHAPTER I INTRODUCTION

1.1. Significance

Epilepsy is a common neurological disorder that affects more than 65 million people worldwide. Each year, 200,000 people are diagnosed with epilepsy; the costs in the United States alone total $15.5 billion (Holland 2014). The indirect costs associated with uncontrolled seizures are seven times higher than that of the average for all other chronic diseases (Holland 2014). Even with today's medication, epilepsy cannot be cured completely and treatment does not work for everyone. Although in 70% of cases seizures can be controlled with the help of available treatments (specialized diet, medication, surgery, etc.), a large number of patients (almost one million Americans) have to live with seizures that resist such treatments. There is an urgent need for more research to discover better treatments with fewer side effects for the first group, as well as novel techniques to predict or control seizures in the second group of patients.

The National Institute of Neural Disorders and Stroke (NINDS), the major sponsor of epilepsy research in the U.S., identified as one of the important goals in epilepsy research, the need to: "Prevent, limit, and reverse the co-morbidities associated with epilepsy and its treatment" (Dingledine et al. 2007). Along with many other neurological conditions accompanying epilepsy such as depression, anxiety, cognitive impairment etc., "sleep disturbances" was specifically identified by NINDS as an area requiring closer attention. A deep investigation of interactions between sleep disorders and epilepsy could shed light on "underlying mechanisms" and "optimal treatments", and lead to discoveries of new ways to "prevent the adverse consequences". Hence, it seems essential to conduct research on realistic animal models of epilepsy and determine novel diagnostic and treatment options. For example, electrical stimulation is known to control seizures in many patients but it is not clear what the underlying mechanism is and how to design an optimal stimulation protocol. In addition to electrical stimulation, determination of effective or optimal dosage and timing of antiepileptic drugs (AEDs) requires a thorough understanding of how sleep and seizures interact, which can be obtained through investigations on animals.

The sleep-seizure interplay constitutes a vicious cycle that accentuates the burden on individuals with epilepsy (Figure. 1.1). Seizures disrupt normal sleep and sleep quality in turn influences seizure generation and likelihood. The effects of epilepsy on sleep structure have been shown in clinical trials and animal models (Matos et al. 2010b). In general, epilepsy can impair sleep at two levels: the acute effect of seizures that disturbs sleep continuity and the chronic effect of seizures that alters gross sleep architecture (Crespel et al. 2000). For instance, complex partial seizures (particularly in temporal lobe epilepsy) that happen during sleep significantly reduce the amount of rapid eye movement (REM) sleep (Bazil et al. 2000). In contrast, the effect of sleep on seizures can be protective or precipitating (Baldy and Moulinier, 1984). The precipitating role of sleep on seizures has been linked to the generation of interictal epileptiform discharges (IEDs) (Niedermeyer, 1982). IEDs are the electrophysiological markers most commonly associated with epilepsy and their generation and spread can be affected by stages of sleep in different types of epilepsy. In most cases of epilepsy, and particularly in temporal lobe epilepsy, IEDs present during wakefulness, spread in synchronized non-REM sleep and diminish sharply during REM sleep (Badawy et al. 2012). Distribution of partial seizures during the sleep-wake cycle was summarized in a comprehensive clinical study (Herman et al. 2001). It was concluded that in patients with temporal lobe epilepsy, after correcting for the fraction of time spent in each stage of sleep, sleep-onset seizures are much more likely to occur during lighter stages of NREM sleep (N2). It was shown that synchronization in neural activity during NREM facilitates initiation and propagation of partial seizures which is consistent with observations of IED generation and spread. On the other hand, desynchronized REM sleep prevents generalized discharges and prevents seizure incidence (Crespel et al. 2000 and Badawy et al. 2012).
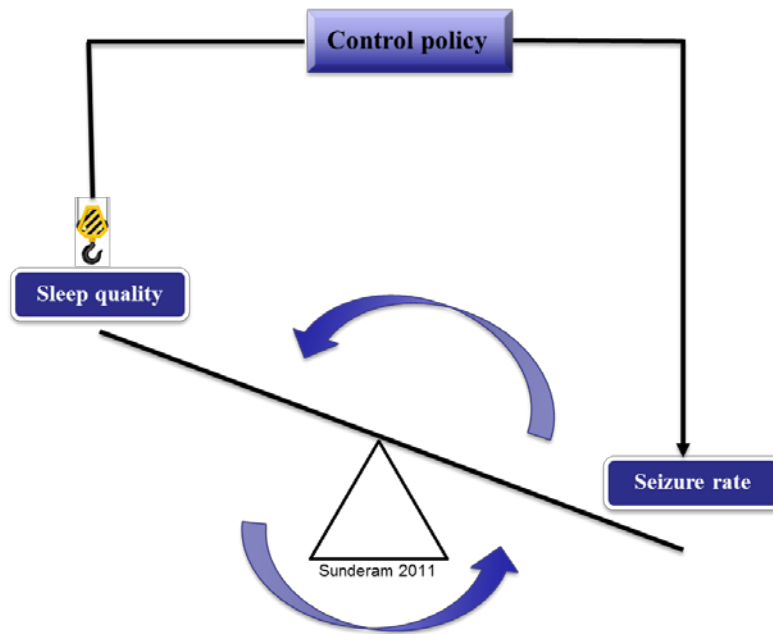
Figure 1.1 A simple model for sleep-seizure interactions in epilepsy. Seizures impair sleep quality and sleep states bias seizure generation and likelihood. This interplay can aggravate epilepsy in patients (Image courtesy of S. Sunderam, 2011)

Thus, experimental investigations on established animal models of epilepsy would help scientists develop ways to control seizures while sustaining sleep quality. The same general framework can be utilized for many different methods of neuromodulation being considered for automated seizure control like electrical stimulation, brain cooling, and drug infusion. Understanding the sleep-seizure relationship could also help doctors adjust the timing and dose of antiseizure and sleep medication so that seizure freedom and normal sleep are both achieved. The first step to decipher this reciprocal interaction is to design and develop appropriate experiments and computational techniques for analyzing sleep in animal models.

Sleep quality is a critical determinant of public health. Sleep disorders are not only problematic diseases by themselves but also aggravate the management of a wide range of neurological syndromes such as epilepsy, Parkinson's and Alzheimer's disease. Clinical evaluation of sleep involves overnight polysomnography (PSG) and explicit guidelines for scoring. A PSG recording essentially includes electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG) and electrocardiogram (EKG).

The polysomnogram is inspected by a human expert typically in 30s epochs to score sleep into five distinct states of vigilance: Wakefulness (*W*), rapid eye movement or REM sleep (*R*), and non-REM sleep (*N*) with stages *N1*, *N2*, and *N3* that reflect increasing sleep depth. Scoring sleep in PSG recordings, in spite of detailed guidelines, remains a tedious and subjective task for sleep physicians. Automated algorithms for clinical or experimental analysis of sleep are in great demand.

Despite increasing knowledge about sleep and its regulation, its basic functions are still unclear. Since homeostatic modulation of sleep and circadian rhythms is similar across different mammalian species, available knowledge about sleep mechanisms in humans has been widely improved through appropriate animal models. The application of animal models, particularly rodents such as mice and rats, provides scientists with the opportunity to discover the neurobiological alterations underlying sleep abnormalities. Mice are of great value in sleep and behavior research since they are readily available and electrophysiological determination of sleep state is quite similar to humans. Mice are also widely used to study the contribution of genes in sleep and behavior. Hence, progress in experimental investigation of sleep is always welcome.

This dissertation aims to cover different experimental and computational aspects related to sleep investigation in animal models. We first propose novel techniques ranging from electrophysiological recording and modeling to manipulation of sleep and apply them to wild type control mice; then we discuss their application in an epilepsy model. The specific goals of this dissertation are as follows.

1.2. Specific aims

Aim I: Unsupervised analysis of mouse sleep using EEG/EMG measurements

Sleep scoring is the prerequisite step for analysis of vigilance dynamics and studying the correlation with seizures. However, sleep scoring usually requires tedious visual review of prolonged EEG/EMG recordings by a human expert. Since data-driven segmentation of sleep could give reproducible output while reducing the need for human scoring, computerized analysis of sleep is highly desirable. To this end, much effort has gone into using computer algorithms to imitate human scoring. The majority of proposed classification algorithms are *supervised*, meaning that they require scored data sets for

both training and validation. In contrast, unsupervised models seek out inherent partitions or patterns in the data to score vigilance states without prior training, which a supervised classifier cannot do.

As this specific aim, a completely unsupervised classifier based on hidden Markov models (HMMs) was developed to accurately track instantaneous sleep-wake states in mice from continuous EEG/EMG measurements. The proposed model was tested on several animals and performance compared with true manual scores. The outcome of this study demonstrated the HMM's efficacy in prediction of vigilance state and can be very useful in chronic epilepsy models.

Aim II: Automated noninvasive determination of vigilance state in mice using a piezoelectric motion sensor

Although EEG/EMG measurements are the gold standard for scoring sleep in mice, the need for surgical implantation of electrodes limits the scope and rate of experimentation. In this aim, we applied the computational modeling procedure proposed in Aim I to a motion measurement based on a piezoelectric pressure sensor to assess the potential for noninvasive scoring of sleep in mice. Successful implementation of such a system not only benefits high-throughput sleep screening, but also provides several advantages for screening animal models of epilepsy.

Aim III: Computational tracking of sleep dynamics following brain injury

The ability to track sleep quality and microstructure following brain injury could potentially provide clues for diagnosis and treatment of neural disorders such as epilepsy. Also, continuous tracking of vigilance dynamics during intervention (e.g. electrical stimulation or drug injection) would help to dynamically optimize treatment dose and timing. Here, we proposed metrics of sleep quality and dynamics derived from HMM parameters estimated from different physiological measurements and showed how sequential reestimation of these metrics enables us to track progressive changes in sleep quality in long-term recordings. This technique will be useful for extracting dynamic information associated with epileptogenesis and epilepsy without the need for visual scoring.

Aim IV: Selective restriction of sleep states using sensory stimulation

Experimental manipulation of sleep is necessary to understand the underlying

mechanisms of it. It is also a useful tool to study the relationship between seizures and vigilance state in experimental investigations. One approach is to deprive sleep partially or totally in pre-clinical animal models of epilepsy to investigate the consequences on seizures and fill the gaps in our knowledge of sleep-epilepsy interactions. In this specific aim, a closed-loop sensory stimulation technique was proposed and the functionality for selective REM sleep restriction validated in mice. This system can be used to modulate sleep quality in epilepsy models and study the effects on seizures (Figure. 1.1).

In specific aims 1-4, we tried to develop experimental and computational tools to facilitate sleep research in animal models. These tools, each presented in a separate chapter, are expected to be useful for addressing problems related to the analysis of sleep-seizure interactions in epilepsy. In a separate chapter, we discuss the implementation of a chronic model of epilepsy in mice and the application of tools proposed in the four specific aims for the analysis of sleep and sleep-seizure interactions. This discussion is intended to point the way forward for further investigation in this area. The final chapter wraps up the main findings of this dissertation. Three manuscripts published after peer review are presented in the Appendix as evidence of dissemination of the findings, mainly for clinically relevant applications.

2. CHAPTER II UNSUPERVISED ANALYSIS OF MOUSE SLEEP USING EEG/EMG MEASURMENTS

2.1. Rationale

To overcome the difficulties associated with manual scoring of sleep and behavior, several statistical classifiers have been developed. The main objective of such models is to automate this process and replicate human performance. These computational tools either look for inherent patterns to define dominant vigilance states, referred to as unsupervised classification, or require supervision in the form of expert heuristics or expert-scored training data to stage sleep, referred to as supervised classification. Both techniques are mutually exclusive and implemented in essentially the same manner: a model is fitted to data from a set of subjects and validated on out-of-sample data from another set. This guarantees that the model will work reliably on new subjects.

The need for scored training data and subjectivity and variability in human scoring are the main constraints for supervised classifiers. Thus, a method to generate a reasonable first-pass hypnogram (i.e., a time sequence of vigilance state scores) from a sleep recording without supervision or previous training is highly desirable. Unsupervised scoring of sleep has been attempted almost since the time digital EEG was first available. The earliest algorithms translated heuristics used by sleep experts to analyze different features of EEG/EMG signals and divide them into vigilance states (Frost 1970). With advances in machine learning algorithms, various unsupervised classifiers were developed to extract natural partitions in features possibly corresponding to sleep states. Unsupervised sleep classifiers can be further divided into static and dynamic models. In general, static models do not consider time or context when classifying state and label each observation only based on its location with respect to boundaries that separate clusters of data in the feature space (Figure. 2.1). Nevertheless, subsequent steps such as minimum duration criterion or median filtering are often applied to refine the static classifier output. On the other hand, investigations of sleep dynamics suggest that human sleep can be fairly well represented by a Markov chain model (Zung et al. 1965), which

incorporates dynamics. The hidden Markov model (HMM) maps continuous-valued observations onto discrete hidden states (Rabiner 1989). The notion of dynamics can be implicitly captured by HMMs as follows: each sleep state follows a trajectory whose likelihood depends (only) on the previous state at any instant (Figure. 2.1). Markov chain models were first used to extract sleep dynamics from human-scored hypnograms by representing probabilistic transitions between different sleep stages (Zung et al. 1965). Optimized versions of Markov chain models were later applied to characterize sleep disorders and medication effects on sleep (Kim et al. 2009 and Bizzotto et al. 2010). Based on Markov chain theory, the HMM assumes that sleep data contain observations generated in various discrete sleep states that are hidden from view. HMM training and parameter estimation are often done in an unsupervised manner, so that the model's prediction of state is not biased by human opinion and will depend on the features selected to represent the data and how much they vary between states. This assumption has made the HMM a popular model in automatic sleep scoring (Flexer et al. 2005; Doroshenkov et al. 2006; Pan et al. 2012 and Langrock et al. 2013).
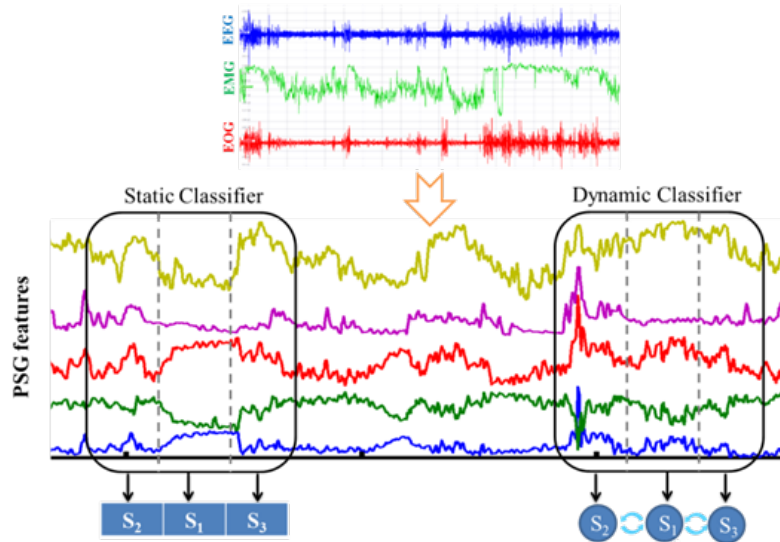
Figure 2.1 A comparison between static and dynamic classifiers in the context of sleep. A static (left) classifier labels features estimated from continuous PSG time series solely based on distributions on feature space. A dynamic classifier (right) includes context (e.g.

time dependencies between states) in scoring.

The advantages of dynamic classifiers (e.g. HMMs) over other unsupervised but static classifiers have been demonstrated on a clinical sleep dataset (Yaghouby et al. 2014a). To justify the computational burden imposed by Markov dynamics on sleep scoring performance, HMMs were shown to outperform other static classifiers (e.g. Gaussian mixture models, k-means clustering and linkage trees) in a naïve scenario i.e. there was no prior knowledge of number of vigilance states and labels.

Automated detection of sleep in rodents has widely benefited from progresses in machine learning. Although the majority of reported techniques in this field are rule-based algorithms (Louis et al. 2004) which are indeed supervised, there are a few attempts to develop completely unsupervised classifiers of rodent's sleep. These unsupervised classifiers are mainly static and model the distribution of features using parametric (Libourel et al. 2015) and nonparametric (Bastianini et al. 2014 and Sunagawa et al. 2013) techniques to label sleep states. Dynamic classifiers have been used, e.g. HMM (McShane et al. 2013) and naïve Bayes classifier (Rytkönen et al. 2011), but have been used in a supervised manner: i.e., the model was estimated using a labeled dataset (scored manually by human rater) and validated on an out-of-sample set.

In this chapter, an HMM-based sleep classifier using EEG/EMG signals is proposed and validated for automated scoring of vigilance state in mice without the assistance of expert opinion. To our knowledge, this is the first attempt to implement a dynamic-unsuppressed sleep classifier using EEG/EMG measurements which are required in epilepsy research as well. However, similar applications of the proposed method in human sleep scoring are presented in Appendix section at the end of this dissertation.

2.2. Animal species, care and protocols

All experimental procedures in this dissertation were conducted with the approval of the Institutional Animal Care and Use Committee (IACUC) at the University of Kentucky. All experiments were performed on adult male wild type mouse (C57BL/6J, henceforth

"BL6"; 6-12 weeks old; 24-29 g). BL6 mice are the most widely used inbred strain which are identical in genotype and provided by the Jackson Laboratory (Bar Harbor, ME). UK's Division of Laboratory Animal Resources (DLAR) is a fully equipped animal facility that provides stable environmental conditions to minimize variability in research results and interruptions in animal's normal behavior. Mice were housed independently upon arrival in animal holding rooms with ambient temperature $23°C\pm2°C$, $50\%\pm10\%$ humidity, 14h light/10h dark cycle (lights on at 7 am) and free access to food and water as well as daily animal care.

2.3. Electrode implantation and data acquisition

The gold standard for sleep staging in mice is invasive EEG/EMG recordings. After a 2-3 day acclimatization period following arrival, EEG/EMG electrodes for brain and muscle activity recordings were implanted in the mice under 2.5% isoflurane anesthesia. The surgery procedure is as follows: the mouse's head is positioned in the surgical apparatus using ear bars and under isoflurane anesthesia (Figure. 2.2a). Using a surgical blade and scissors a longitudinal cut is made in the scalp and the skull surface is gently cleaned by sterile gauze. A small headmount chip (Part# 8201, Pinnacle Technology, Inc, Lawrence, KS) is attached with super glue to the skull so that the upper edge is located directly over bregma (Figure. 2.2b). A 23 gauge needle is used to bore four fine holes through the skull in the frontal and parietal regions corresponding to holes in the headmount. Then, four silver screws are passed through the holes to serve as EEG electrodes (0.1" anterior and 0.12" posterior) and to keep the headmount fixed to the skull (Parts# 8209 and 8212). Silver epoxy (Part# 8226) is also applied to the screws to enhance the conductivity of the EEG electrodes (Figure. 2.2c). The EMG electrodes, a pair of Teflon-coated leads, are tunneled bilaterally into the dorsal neck muscle posterior to the skull. Two or three sutures on the posterior side of the incision help to fix the loose skin and heal the wound. Finally, dental acrylic is used to coat the headmount and EMG electrodes (Figure. 2.2d). The animal is then allowed to recover and adapt to the headmount for at least one week before starting any experiment.
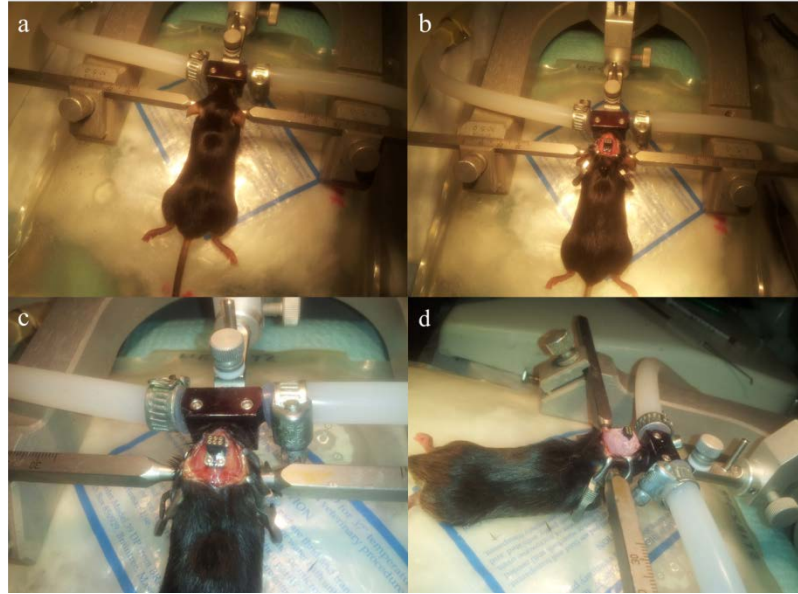
Figure 2.2 Surgical implantation of EEG/EMG electrodes in a wild type mouse. a) Animal's head is fixed using ear bars under anesthesia. b) A headmount chip is attached to the skull using super glue. c) Four screws with silver epoxy act as EEG electrodes. d) Dental acrylic covers the headmount and electrodes

Once the animal has recovered from surgery, continuous recording of EEG/EMG is initiated. A pre-amplifier (Part# 8202) is plugged into the mouse headmount that provides initial amplification of 100X for EEG and EMG signals. A low-torque commutator (Part# 8204) connects the preamplifier to the acquisition/conditioning system (Part# 8206). The 8206 system provides an additional 50x amplification as well as highpass and antialiasing filtering (0.5 Hz for EEG and 10Hz for EMG). The acquisition part digitizes input signals using a 14-bit A/D converter with a 400Hz sampling rate and directs them to computer-based acquisition software (Sirenia, Pinnacle Technology) via a USB connection (Figure. 2.3). Along with EEG/EMG recording, a video monitoring system was implemented to capture the animal's behavior. A small USB camera (Microsoft LifeCam VX-6000) is attached to the animal's cage so that it is always in the field of view. The digital video is recorded by a custom-written LabVIEW^{TM} program (National Instruments, Austin, TX) and timestamped to be comparable to the EEG/EMG recording. An infrared (IR) LED

illuminator source is placed in close proximity to the animal's cage and enables uninterrupted video recording through the light and dark period.

Figure 2.3 Data acquisition system. The preamplifier (#8202) is attached to the animal's headmount and provide initial 50x amplification. A commutator (#8204) connects the preamp to the data acquisition system (#8206). After a final amplification and conditioning by 8206, digital data are recorded using Sirenia^{TM}.

2.4. Manual determination of vigilance state

Unsupervised automated sleep scoring can produce very reasonable outcomes without prior training, but must ultimately satisfy the gold standard of human expert assessment. EEG and EMG recordings are the gold standard for sleep analysis in all mammals. Three vigilance states are typically defined in mice based on EEG/EMG: 1) Wakefulness or Wake, characterized by low amplitude desynchronized EEG and high amplitude variable EMG; 2) Paradoxical or rapid eye movement (REM) sleep, characterized by a prominent theta rhythm in EEG (θ: 6-9 Hz) similar to activity during Wake and suppression of tonic EMG, but with phasic muscle twitches ;3) Non-REM ("NREM") or slow wave sleep stage has increasingly prominent delta (δ: 0.5-4 Hz) in EEG and low tonic EMG (Figure. 2.4). The vigilance state was manually scored based on well-established criteria using an EEG viewer (SireniaTM, Pinnacle Tech.) and video in consecutive 4s epochs as Wake, REM and NREM. Then states of sleep are verified as: NREM, with suppressed EMG and slow oscillation (delta wave: 0.5-4Hz) in EEG signals and REM, with further decrement

in EMG amplitude and faster frequencies (theta waves: 6-9Hz) in EEG signals. Sleep is further confirmed from video when the mouse is motionless with eyes closed. Wake state could also be verified as high EMG activity and various range of frequencies in EEG.
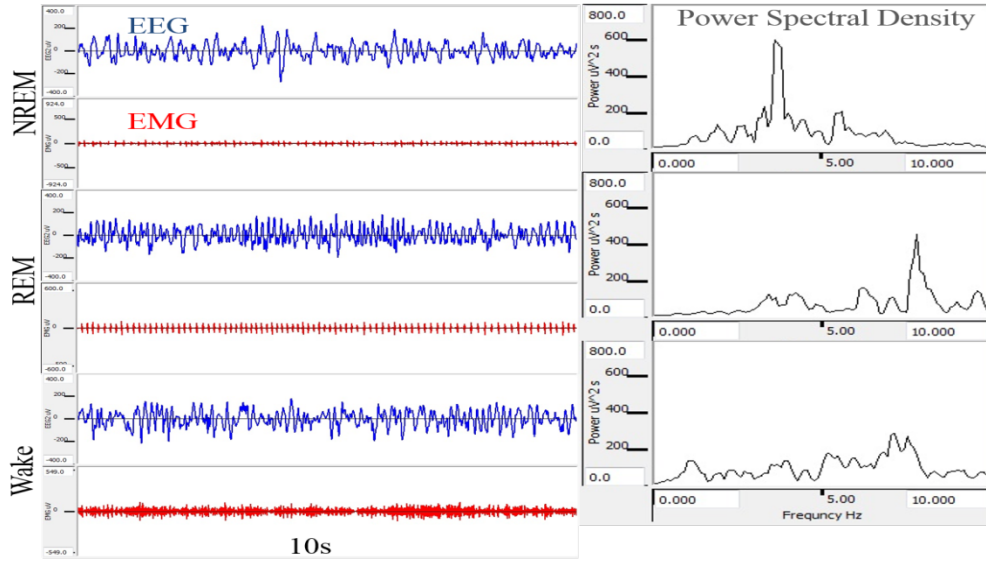
Figure 2.4 Samples of EEG/EMG signals and EEG power spectral density in different vigilance states. Wake state is distinct from high muscle activity and distribution of EEG power separates REM from NREM in sleep.

2.5. Signal processing and feature extraction

Automated signal analysis and classification was performed based on continuous recordings from six mice. A 24-hour EEG/EMG segment from each animal (starting at 7 AM) was first exported to European Data Format (EDF) file and then divided into 4s epochs for processing in MATLAB^{TM}. All epochs were scored manually by a human expert according to the aforementioned guidelines. Based on these scoring criteria and EEG/EMG in different vigilance states, spectral band power features were chosen to construct an automated sleep staging algorithm. Each 4s epoch was bandpass-filtered into seven different frequency bands: δ_L (0.5-2 Hz), δ_H (2-4Hz), θ (6-9Hz), α (9-13Hz), β (13-30Hz) and γ (30-45Hz) using 3rd order Butterworth IIR filters. The mean signal power in each band was estimated as an initial signal feature in each epoch. To reduce

13

dimensionality of the feature space, the $(\delta_{L+}\delta_H)/\boldsymbol{\theta}$ band power ratio was selected as a discriminative feature to separate NREM and REM in sleep. This ratio is high when animal is in NREM sleep and delta is prominent, but declines during REM as theta becomes prominent (Figure. 2.5). The mean power of the bandpass-filtered EMG (80-100 Hz) was also used to discriminate sleep from wake based on muscle tone.
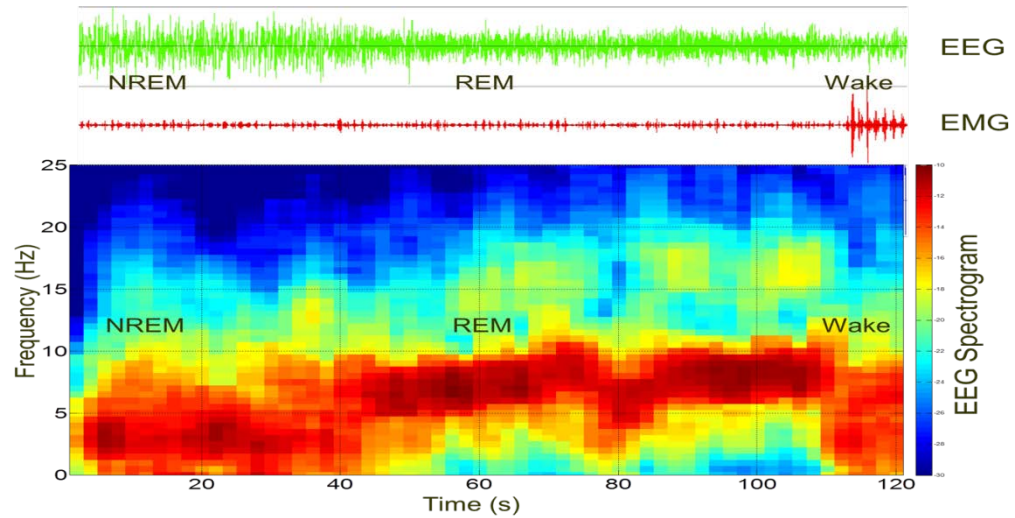
Figure 2.5 A sample EEG/EMG recording and EEG spectrogram during different vigilance states. EEG power has shifted from delta to theta region following NREM to REM transition in sleep.

2.6. Unsupervised modeling and classification

Here we utilized HMMs for automated scoring of sleep and wake states. This section starts with a general introduction to the statistical classification following by a derivation of Bayesian models, particularly HMMs.

In general, a statistical classifier assigns sample observations $\boldsymbol{X}$ to one of N distinct classes $S \in \{1, \cdots, N\}$ by assuming a statistical model that maps $\boldsymbol{X} \to S$. A supervised model needs class-labeled training samples to estimate the parameters; while unsupervised models can be fitted to unlabeled training data and predict the class of upcoming observations. Unsupervised classifiers typically look for intrinsic clusters in the data that coincide with the classes of interest. The prediction of an unsupervised classifier may or may not correspond absolutely with an expert rater's opinion of the

class. But these classifiers can still be beneficial particularly when no prior class definitions are available. Bayesian models, specifically Gaussian mixture models (GMMs) and hidden Markov models (HMMs), are introduced here to demonstrate the feasibility of unsupervised classification of sleep.

In the context of mouse sleep, we can assume that the vector of M features $\boldsymbol{X} = [x_1, \cdots, x_M]^T, \boldsymbol{X} \in \mathbb{R}^M$ (T = transpose) is extracted from EEG/EMG samples (e.g., 4 s epochs) of an animal that is always in one of N vigilance states $S \in \{1, \cdots, N\}$. So, a set of observations $\boldsymbol{X}_{1:T} = \{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_T\}$ can be made in states $S_{1:T} = \{S_1, \cdots, S_T\}$. In fact S includes model states that may correspond directly to true vigilance state i.e. NREM, REM and Wake. The state of the subject at any given time S_t is quantified by the observation $\boldsymbol{X}_t$. The basic role of a classifier is to infer S_t from $\boldsymbol{X}_t$ with adequate accuracy. However, estimation of $\boldsymbol{X}$ can be affected by variability and noise and this is modeled using a probability density function $f(\boldsymbol{X})$. The probability measure $P(\boldsymbol{X})$ is obtained by integrating $f(\boldsymbol{X})$ over a region of $\boldsymbol{X}$. Since the N states are mutually exclusive, the probability of observation $\boldsymbol{X}$ integrates the probability that $\boldsymbol{X}$ is observed in any state: i.e.,

$$P(\boldsymbol{X}) = \sum_S P(\boldsymbol{X} \cap S) \qquad (1)$$

Similarly, the conditional probability is the probability that $\boldsymbol{X}$ is observed and the state is known to be S:

$$P(\boldsymbol{X}|S) = P(\boldsymbol{X} \cap S)/P(S) \qquad (2)$$

$P(S)$ Indicates the prior probability of state S when there is no information about $\boldsymbol{X}$. Eq. 2 is known as Bayes rule. By combining equations 1 and 2, an expression for the probability distribution of $\boldsymbol{X}$ in terms of the conditional and prior probabilities is obtained:

$$P(\boldsymbol{X}) = \sum_S P(\boldsymbol{X}|S)P(S) \qquad (3)$$

Posterior probability of state S can be computed knowing probability of observation $\boldsymbol{X}$:

$$P(S|\boldsymbol{X}) = P(S \cap \boldsymbol{X})/P(\boldsymbol{X}) = P(\boldsymbol{X}|S)P(S)/P(\boldsymbol{X}) \qquad (4)$$

And a reasonable prediction of state is the one that maximizes the posterior probability:

$$\hat{S} = \underset{S}{\operatorname{argmax}} P(S|\boldsymbol{X}) \qquad (5)$$

To make any prediction using a Bayesian model, a prior knowledge of the conditional probability $P(X|S)$, usually in a form of standard parameters is required. For example in a GMM, $P(X|S)$ is modeled as a Gaussian distribution parameterized by a state-dependent mean vector $\boldsymbol{\mu}_S \in \mathbb{R}^{N \times M}$ and covariance matrix $\Sigma_S \in \mathbb{R}^{M \times M}$. A linear coefficient α_S replaces the state prior $P(S)$ in Eq. 3 to represent $P(X)$ as a mixture of Gaussian components. In the other words, a GMM fitted to sleep data assumes that observation X can be modeled as a mixture of Gaussian components and each component corresponds to one of the vigilance states.

HMMs have been used to model sleep dynamics and track the evolution of a process over time. "Dynamics" in HMMs can be added as a layer of complexity to Eq. 3 by linking the model states to each other. In fact, rather than independent observations, an HMM models the distribution $P(X_{1:T})$ of the time series of observations generated by a hidden state sequence $S_{1:T}$ (Figure. 2.6A-B). The Markov property is the first assumption for HMM in which the current state exclusively determines the distribution of future states:

$$P(S_t|S_{1:t-1}, X_{1:t}) = P(S_t|S_{t-1}) \qquad (6)$$

Eq. 6 represents state transition probabilities that form an $N \times N$ state transition matrix γ when all possible combination of S_{t-1} and S_t get assembled. In addition to Markov transitions, the second assumption is that given the current state, observation is conditionally independent of previous observations and states (Figure. 2.6C):

$$P(X_t|S_{1:t}, X_{1:t-1}) = P(X_t|S_t) \qquad (7)$$

This conditional probability $P(X|S)$ along with a matrix of transition probability γ and a set of state prior probabilities $\pi = P(S)$ characterize an HMM parameters. To simplify the calculations, we assume these properties are independent of time t (stationary). In such manner, the dynamics of the process underlying observations $X_{1:T}$ can be interpreted by applying recursive Bayes's rules.

Once HMM parameters are fixed, we can decode a sequence of states $S_{1:T}$ most likely to have generated a sequence of observations $X_{1:T}$ using the Viterbi algorithm (Rabiner 1989). The probability of first observation X_1 is first defined by Viterbi algorithm as $\delta_1(S) = P(X_1|S)$, for $S \in \{1, \cdots, N\}$. Then, the preceding state is sequentially computed as the one that maximizes the probability of successive observations $\delta_t(S') = \max_S[\delta_{t-1}(S)\gamma(S, S')]P(X_t|S')$. At the end, the optimal probability of state sequence is

$P^*(S) = \max_S \delta_T(S)$; i.e. terminal state is the one that maximizes $P^*(S)$. Once sequence δ_t is determined, the most likely sequence of states $S_{1:T}$ is optimized by backtracking δ_t (Figure. 2.6D).

Estimation of GMM and HMM parameters is based on maximum likelihood (ML) algorithm on training data (Rabiner 1989). For a GMM with parameter set $\Theta = \{\alpha_S, \boldsymbol{\mu}_S, \Sigma_S\}$, likelihood function L is defined as the joint probability density of a set $\boldsymbol{X}_{1:T}$ of independent and identically distributed observations:

$$L(\Theta|\boldsymbol{X}_{1:T}) = P(\boldsymbol{X}_{1:T}|\Theta) = \prod_{t=1}^{T} P(\boldsymbol{X}_t|\Theta) \qquad (8)$$

The product is converted to a sum over data using log operator:

$$\log L = \sum_{t=1}^{T} P(\boldsymbol{X}_t|\Theta) \qquad (9)$$

The likelihood function L defines the model parameters as a function of observations. Hence, by taking the partial derivative of $\log L$ with respect to each parameter and equating it to zero, we can obtain parameters Θ that maximize $\log L$.

In supervised case, where labeled training data is available, the estimation of model parameters is relatively easy and straightforward. For example, mean of state S is estimated as the arithmetic average of all training samples labeled as S by a human scorer. However, the main application of GMMs and HMMs is in unsupervised classification where no labeled training data are available. In this case, observations become dependent on the parameters through hidden states $S_{1:T}$ apart from the unknown Θ:

$$\log L = \sum_{t=1}^{T} P(\boldsymbol{X}_t, S_t|\Theta) \qquad (10)$$

Hence, in Eq. 10 we have unknowns on both side of the conditional: Θ and S_t. The Expectation-Maximization (E-M) algorithm becomes a solution to this problem (Rabiner 1989). E-M starts with an initial guess for the model parameters and converges to a local minimum through an iterative process. Several initial guesses are tested and the parameters set with greatest likelihood is selected. In this manner, we avoid to get trapped in local optima. A widely known E-M algorithm for HMMs is the Baum-Welch algorithm (Rabiner 1989).
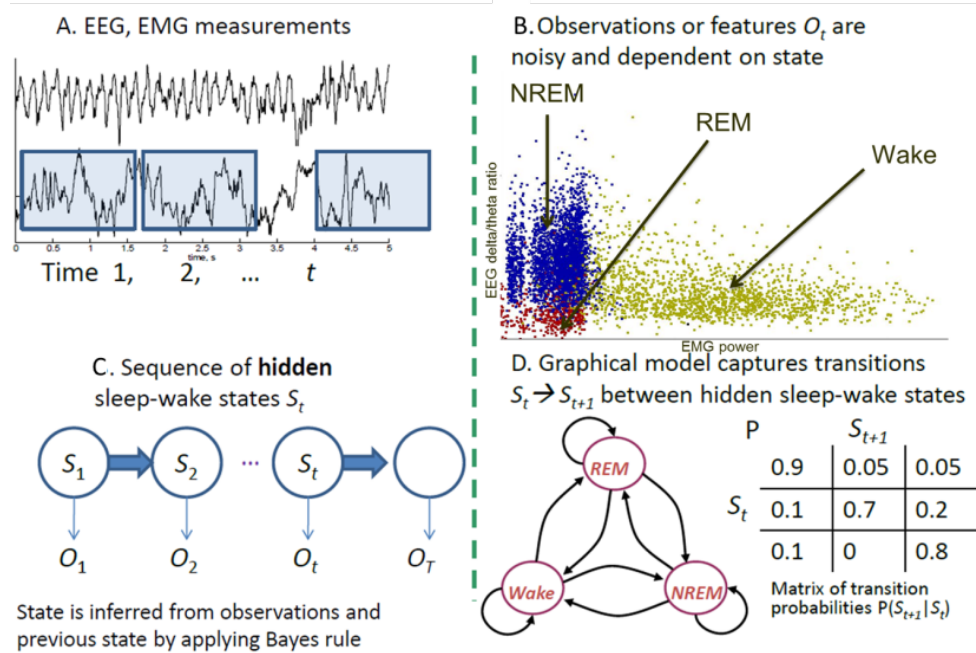
Figure 2.6 Sequential steps for HMM modeling of sleep-wake in mice. A) EEG/EMG features are extracted from continuous signals in 4s epochs. B) These observations are noisy but the distributions in feature space depend on state. C) A hidden state can be inferred by applying sequential Bayes' rule. D) HMM gives a graphical representation of sleep dynamics along with a matrix of transition probabilities.

2.7. Results and discussion

The feasibility of HMMs for automated sleep scoring in mice (n=6) was investigated. One complete 24-hour day of EEG/EMG recording from each mouse was manually scored in 4-s epochs and two features were calculated: EEG δ/θ power ratio which is low in REM sleep and EMG power which is high during wake. Figure. 2.7 demonstrates a 6-hour recording sample of EEG/EMG features (starting at 7 AM). Model performance can be assessed by comparing HMM-predicted labels against true hypnogram labels from visual scoring. Contingency tables in Figure. 2.7 show strong agreement between the classifier and independent scorers in every state of vigilance.

The overall performance of the proposed classifier on six mice is shown in Figure. 2.8a. Two conventional metrics from the contingency table are used to assess detection accuracy: 1. The sensitivity (expected true positive rate) of a specific vigilance state reflects the proportion of actual sample epochs of that state correctly identified by the classifier; and 2. The specificity (expected true negative rate) for a particular state is the proportion of other states not wrongly classified as the state of interest. HMMs predict manual sleep scores with >90% sensitivity and specificity and within the bounds of inter-rater variability.
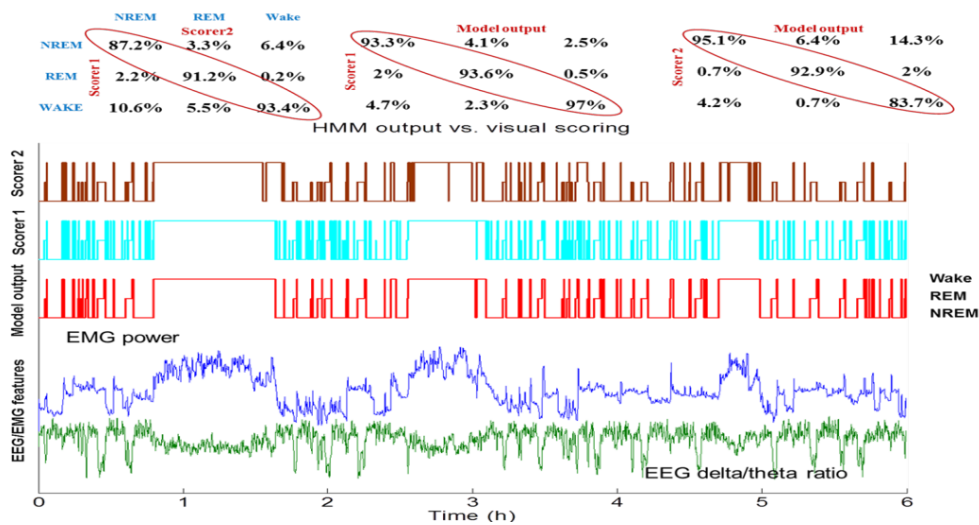
Figure 2.7 Comparison of HMM output with two human scorers in a 6-hour sample recording. EMG power (blue) reflects muscle tone and separates sleep from wake. EEG delta/theta (green) is used to distinguish REM from NREM. An HMM fitted to these features predicts instantaneous vigilance state with high accuracy. Comparison between HMM output and human raters shows over 90% agreement.

The % time spent in NREM, REM and Wake is a popular metric that can be derived from manual or automated scores. The prior probability of each HMM state is analogous to this metric. A comparison with manual estimates in 24 h scored samples shows strong agreement (Figure. 2.8b). Hence, HMMs can extract sleep metrics without tedious visual scoring.
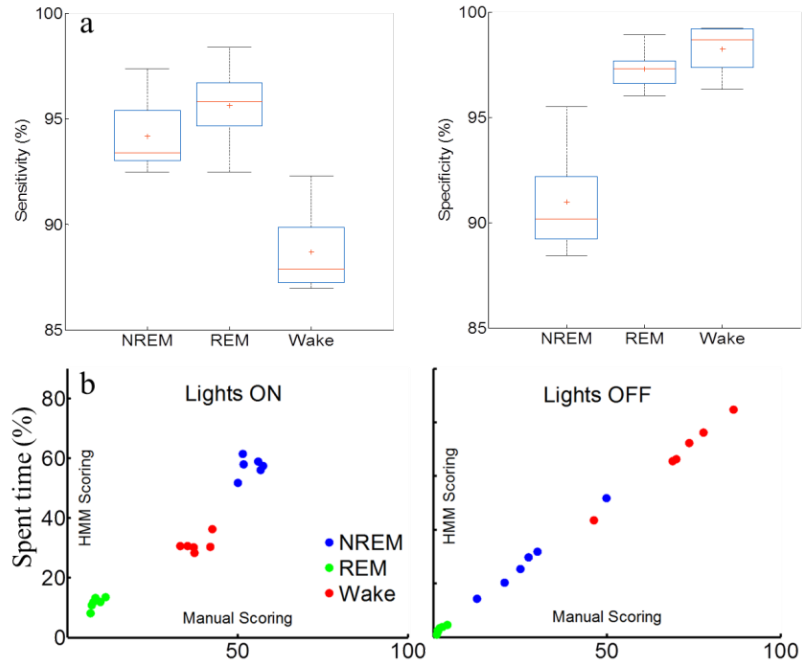
Figure 2.8 HMM performance for sleep classification in 6 mice. a) Sensitivity and specificity are high (>90%) for all states. b) Comparison of time spent in each state between model and manual estimations shows strong correlation (Yaghouby et al. 2013).

Unsupervised HMMs were found to track instantaneous sleep-wake state with high accuracy from continuous EEG/EMG measurements. Unsupervised HMM scoring of sleep-wake state can be used to estimate conventional metrics of sleep quality such as the % time spent in each state and mean bout duration. This modeling approach could be a useful tool for studying vigilance dynamics in mouse models of epilepsy as well. In Chapter 7, we will show how to take advantage of such a model and score vigilance states in epilepsy models.

The same computational modeling framework can be employed to resolve clinical sleep scoring problems. Clinical sleep scoring requires tedious visual analysis of overnight Polysomnograms (PSG) by a human expert. To automate this process, many efforts have been made using computer algorithms to imitate human scoring patterns. The majority of proposed classification algorithms are supervised, and are typically trained on scored data and then validated using out-of-sample data. The need for human scoring of training data and subjectivity and variability are the main constraints for supervised classifiers. Thus, a method to generate a reasonable first-pass hypnogram from a sleep recording without

supervision or previous training is in great demand. Strictly speaking, there is no initial model available in the "naïve" scenario. Also, all vigilance states may not occur during a recording that needs to be scored. On the other hand, evolution from clinical sleep monitoring systems to portable/wearable devices depends on simplifying and improving conventional systems. Detection of sleep stages using a single EEG signal, instead of whole PSG data, not only brings more comfort for patients (less sensor contact and cabling) but also reduces the computational load of signal analysis. Hence, an exclusively EEG-based sleep classification algorithm will make the design and implement of a portable sleep monitoring device more efficient. In Appendix A, we described a methodology based on HMMs for scoring overnight sleep recordings with two key benefits: 1) only EEG data were used not whole PSG set; and 2) there was no need for expert labels to train the classifier. The proposed model was tested on a 22-subject sleep database and performance compared with other unsupervised classifiers (Yaghouby et al. 2014a).

Another clinical application of the proposed technique is presented in Appendix B. As we discussed in this chapter, manual scoring of sleep is a tedious and subjective task and uncertainty and variability in assessments by expert raters are the major obstacles. Even expert raters can be unsure about the presentation of particular states and may disagree widely in their assessment of specific recordings. None of the available automated sleep scoring algorithms explicitly address rater uncertainty and disagreement issues. We proposed a quasi-supervised classifier that models observations in an unsupervised manner but mimics an expert's scoring patterns wherever training scores are available. The novel technique proposed addresses three problems related to human sleep scoring: 1. The rater is confident of scoring only some of the states; 2. The rater scores all states but is uncertain of some observations; and 3. Two raters score all states and observations but with some disagreement. To address these problems, PSG features were first extracted from overnight recordings. Then, unsupervised statistical models (Gaussian mixture and hidden Markov models) were estimated from training features incorporating partial scores. Finally, the fitted models were used to predict scores for complete recordings and performance is assessed statistically (Yaghouby et al. 2015a).

3. CHAPTER III NONINVASIVE ANALYSIS OF MOUSE SLEEP USING A PIEZOELECTRIC MOTION SENSOR

3.1. Rationale

Automated analysis of mouse sleep using EEG/EMG measurements was explored in the previous chapter. However, invasive implantation of electrodes and recovery from surgery remain as the main obstacles particularly in large-scale experiments (e.g. genetic screening) and affect the scope of such investigations drastically. Furthermore, EEG/EMG acquisition often requires that the animal be tethered, thus restricting behavior. This severely limits the ability to simultaneously screen a large number of animals. Non-tethered telemetric systems exist but require invasive implantation of electrodes, preamplifier, battery pack, and transmitter that again alter behavior and restrict movement, especially in the smaller mouse. In this chapter the feasibility of a motion sensor sensitive to movement (during wakefulness) and respiration patterns (during sleep) for automated sleep scoring in mice is investigated.

3.2. Experimental setup and data acquisition

A completely noninvasive monitoring system discriminating stages of sleep and behavior in mice, at a level comparable to that of EEG/EMG, would greatly facilitate high throughput screening of sleep in neuroscience research and gene discovery. It has previously demonstrated that a continuous signal derived from a pressure-sensitive piezoelectric sensor can be used to distinguish sleep from wake in mice with accuracy comparable to a human expert (Donohue et al. 2008). The piezoelectric sensor (piezo) is a thin polyvinilidine diflouride (PVDF) film that produces a voltage signal in response to changes in surface pressure. Hence, motion associated with mouse behavior produces characteristic signals that typify behavior. The PVDF sensor is a 110um thick dielectric sheet (Signal Solutions LLC, Lexington, KY). Silver ink sputtered on each side creates a conductive link from any position where pressure is applied. The capacitance of the PVDF is ~30nF, and when coupled to the input instrumentation differential amplifiers followed by a lowpass filter, effectively bandpasses the pressure signals from 1.35-20Hz (-3dB). The pass band of the instrumentation amplifier filter was designed to cover the

frequency range associated with sleep and breathing in mice (Friedman et al. 2004). The amplified signal is acquired using a multichannel data acquisition board (NI-DAQ 6211, National Instruments) and controlled with Sirenia™ software (Pinnacle technology, INC).

The "piezo" signal can discriminate sleep from wake states with over 90% accuracy (Donohue et al. 2008). Simultaneous measurement of respiratory effort using an impedance pneumogram (Flores et al. 2007) showed that changes in breathing can be detected when the mouse is relatively inactive.

It is widely accepted that breath rate fluctuates largely in REM sleep while is highly regular during NREM sleep (Freidman et al. 2004).Two major characteristics of REM sleep, muscle atonia and phasic events, play a key role in ventilation control. Atonia in voluntary muscles affects respiratory motor output in diaphragm and consequently causes breathing variability during REM sleep. Phasic events, originated in brainstem, could spread their effects peripherally and alter heart rate or blood pressure. Therefore, they could profoundly alter respiration by impressing neural systems of ventilation control. Phasic events, and particularly rapid eye movements, alter three major respiratory variables consistently. By increased eye movement density in REM sleep, respiration rate elevates while expiratory duration and rib cage motion suppress. Accordingly, the main source of breathing variability during REM sleep is alteration in these variables in relation to phasic events (Pack 1995). Respiration in REM sleep differs from NREM sleep and this contrast could be because of: 1) During REM sleep, contribution of abdominal motion in ventilation is significantly high compared with rib cage motion in NREM sleep; and 2) Phasic events in REM cause asynchrony in rib cage and abdomen movement which are synchronous during NREM. Although the mechanism is not clear yet, some studies have shown significant suppression in transmission of respiratory afferent signals during REM that affects respiration pattern as well (Pack 1995).

Pressure changes on the piezo sensor associated with respiratory patterns may have signatures characteristic of different stages of sleep as well. Variations in respiratory patterns captured by piezo signal in mice are shown in Figure. 3.1. In all five sample signals, regular piezo became erratic following transition to REM sleep and before animal is woken up.

23
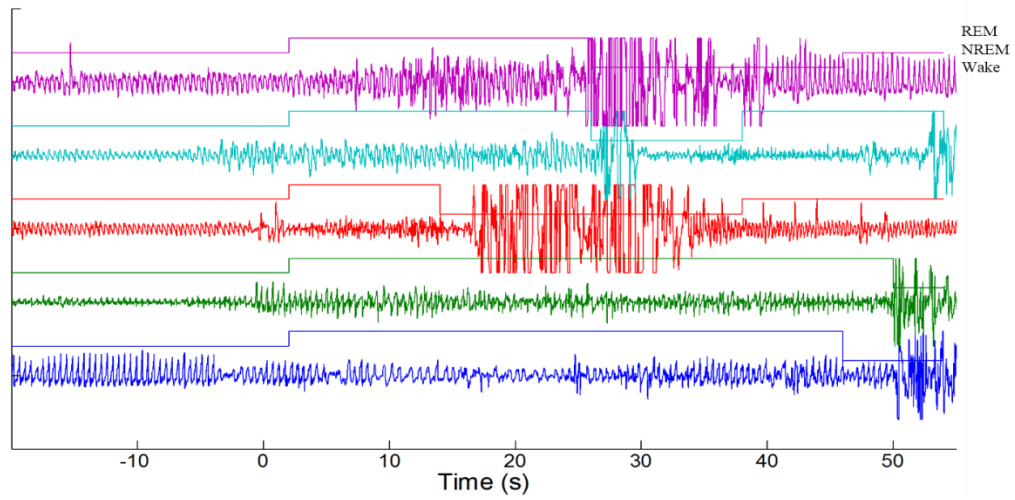
Figure 3.1 Samples of piezo signal changes time-locked to NREM-REM transitions. Regular breathing during NREM becomes erratic in both amplitude and frequency at transition to REM (at Time = 0s) (Yaghouby et al. 2011).

Figure. 3.2a shows how a regular pattern of breathing in NREM sleep is disturbed by transition to REM and then Wake states. The correlation between constant 3-4 Hz breath rate and EEG delta power along with correspondence between variable breath rate and EEG theta power is obvious in spectrograms (Figure. 3.2b).
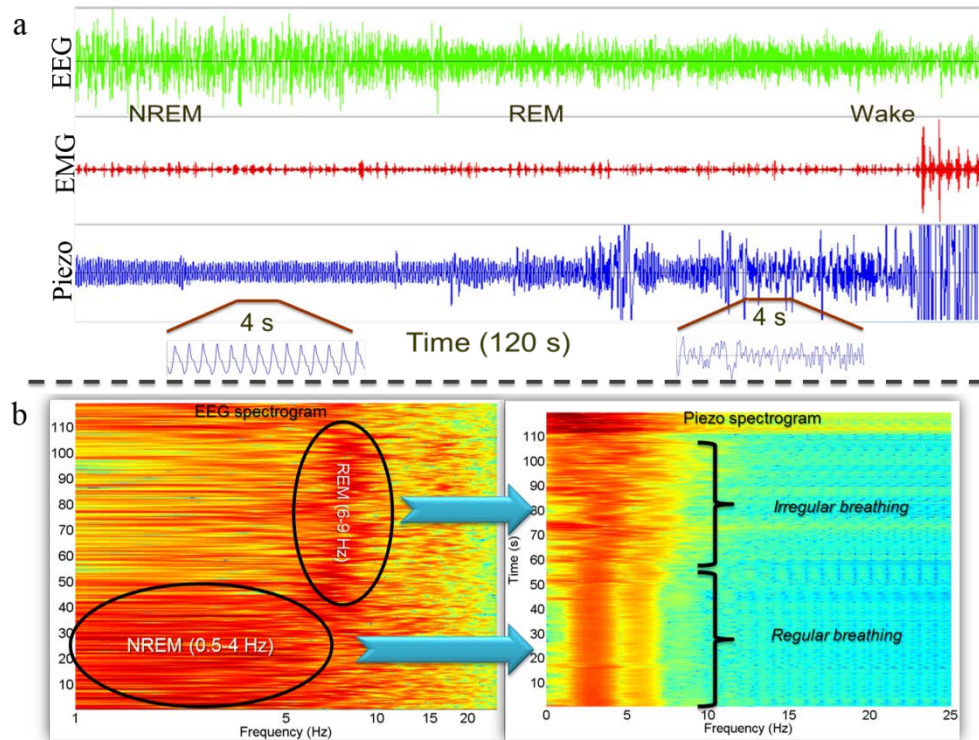
Figure 3.2 Simultaneous EEG/EMG/piezo recordings during different vigilance states. a) Regular breathing associated with NREM is interrupted at the middle of recording with transition to REM. b) EEG and piezo spectrograms show correlation between breathing patterns and sleep states. Regular breathing (3-4 Hz breathing rate) is associated with NREM and irregular breathing reflects REM sleep.

3.3. Signal processing and modeling

The introduced piezo system would be of much greater value if it could be proven to differentiate between stages of sleep i.e. REM and NREM as well. Here, we have further investigated the competence of piezo signal and features in tracking respiration during sleep. A computational framework for extracting piezo features that represent breathing regularity and levels of activity is proposed. Feature extraction is then followed by a HMM-based classification algorithm to segment 4-s epochs of the peizo signal first into sleep and wake states, as well as subdivisions of these that bear a striking statistical similarity to REM/NREM sleep and brief arousals. Similar to unsupervised classification using EEG/EMG features in previous chapter, we modeled piezo features using HMMs to correlate inherent breathing dynamics with vigilance states. Simultaneous recording of EEG/EMG signals along with piezo provides the capability of visual scoring that can be used to evaluate the performance of proposed algorithm.

Mouse sleep and behavior can be characterized using superficial pressure induced to the cage floor and captured by the piezo sensor. One can extract different piezo features to identify gross and fine movements of the animal. In general, two main feature categories from piezo signal were defined: motion and respiration features. The rest of this section is dedicated to introducing these features.

Motion features: This category consists of a feature that distinguishes gross motor movement (reflecting most likely wakefulness) from quite state (reflecting most likely sleep). It has been shown before that combination of a linear classifier and a range of piezo features could efficiently distinguish sleep and wake in mice (Donohue et al. 2008). Teager energy (Kaiser 1990) is a measurement of a signal broadband energy and chosen here to determine motion. Teager energy (TE) of a piezo time series $p(t)$ is estimated as:

$$\psi[p(t)] = p^2(t) - p(t+1)p(t-1) \quad (1)$$

As can be seen in Figure. 3.4a, piezo TE correlates strongly with EMG power and mirrors instantaneous power of muscle tone even in brief episodes of wake. So, it can be considered as a noninvasive surrogate for tracking muscle tone.
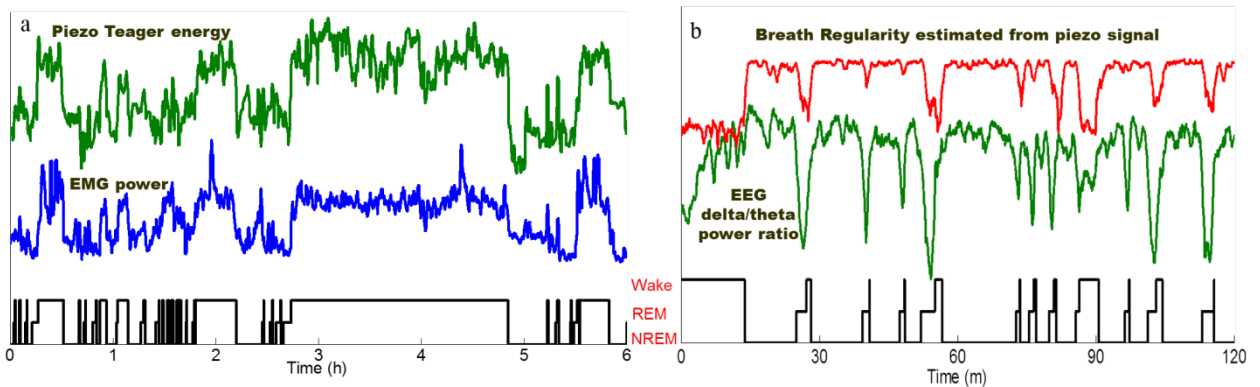
Figure 3.3 Correlation between piezo and EEG/EMG features in different vigilance states. a) Piezo TE correlates strongly with EMG power and separates sleep even from brief wake episodes. b) Piezo breath regularity suppresses when theta portion of EEG elevated in REM. So, breath regularity can be used noninvasively to track EEG delta/theta power.

Respiration features: Noninvasive tracking of sleep/wake using piezo TE is advantageous but discrimination of sleep stages (REM and NREM) is a higher-risk endeavor. Chest and abdominal wall movement becomes the predominant motion when

26

animal is inactive (i.e. during sleep). Hence, piezo signal can reflect respiratory patterns in sleep (Figures .3.1 and 3.2). Breathing characteristics (e.g. amplitude and frequency) are modulated by autonomic nervous system which is under sleep state influence itself. Generally, breathing in NREM and particularly in deeper states is more regular than during wakefulness and REM sleep (Pack 1995). The irregularity in breathing during wakefulness is because of body movements, conscious control, or other external factors. While, as explained before, muscle atonia is one of the major reasons for breathing variability during REM sleep (Pack 1995).

Figure. 3.2 shows that deviation form baseline breathing following REM transition could be in both amplitude and frequency. Therefore, features reflecting the regularity or variability of piezo signal during sleep can be used to separate NREM and REM episodes. First of all, piezo signal is low-pass filtered at 5 Hz to generate breathing signal (breath rate in mice is in 3-4 Hz range). Then, two groups of features were extracted from breathing signal to quantify variability in amplitude and regularity in time. Variation of breathing amplitude is one source of irregularity that can be estimated using "envelope" of the breathing signal. Here we used "analytic signal" to estimate envelope. Analytic signal is a complex time series in which the real part is original signal and imaginary part is Hilbert transform of the signal. If $p(t)$ is a time series of breathing signal, the signal envelope $e(t)$ is defined as the magnitude of the analytic signal $\hat{p}(t)$ (Figure. 3.5a):

$$\hat{p}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{p(s)}{t-s} ds \qquad (2)$$

$$e(t) = \sqrt{p^2(t) + \hat{p}^2(t)} \qquad (3)$$

Finally, the coefficient of variance of $e(t)$ is calculated as breathing amplitude variability (*BAV)* feature:

$$BAV = \frac{std(e(t))}{mean(e(t))} \qquad (4)$$

To assess regularity of breathing signal in time, an accurate estimation of instantaneous breath rate (*IBR*) is first required. *IBR* may be calculated as number of phase rotations in Hilbert transform of breathing signal. *IBR* doesn't change that much during NREM (3-4 Hz) but varies in REM and wakefulness (Figure. 3.2). Breathing time regularity (*BTR*) is another feature that can be estimated in three different ways. First approach is to calculate the phase coherence between foreground ($P(t)$) and background ($P(t-\tau); \tau = 1s$)

breathing signals (*BTR*1). When breathing is regular in baseline phase, coherence should constantly stay high. When breathing becomes irregular during REM, the coherence between piezo signal and its delayed version in time would drop and *BTR*1 becomes lower (Figure 3.4b).

The other alternative way for *BTR* estimation is based on the Rayleigh circular statistics (Fisher 1993) in which a breath, *b,* can be considered as a complex phasor, $r_b = \exp(-j\omega t_b)$ where $= \sqrt{-1}$, t_b is the time of breath *b* with respect to an arbitrary origin, and $\omega = 2\pi f$ where frequency *f = IBR*. Vectorial summation of individual phasors in a short window produces a net phasor $|BRT2 = \sum_b r_b|$. If the breath interval distribution is random (Poisson), *BTR* will approach zero; but if it breathing is highly regular with a strong periodicity, *R* will approach *IBR* that is defined as:

$$IBR = \frac{\#\ of\ breaths\ in\ each\ segment}{inter\ breath\ interval} \qquad (5)$$

We can determine inter breath interval using Hilbert transform as explained above. *BTR*2 measures how closely the phasors are clumped, a reflection of periodicity in breathing. As Figure. 3.3b shows, *BTR2* tracks EEG $\delta/\boldsymbol{\theta}$ power ratio and is relatively high when breathing is regular in NREM but drops momentarily when it is variable in REM.
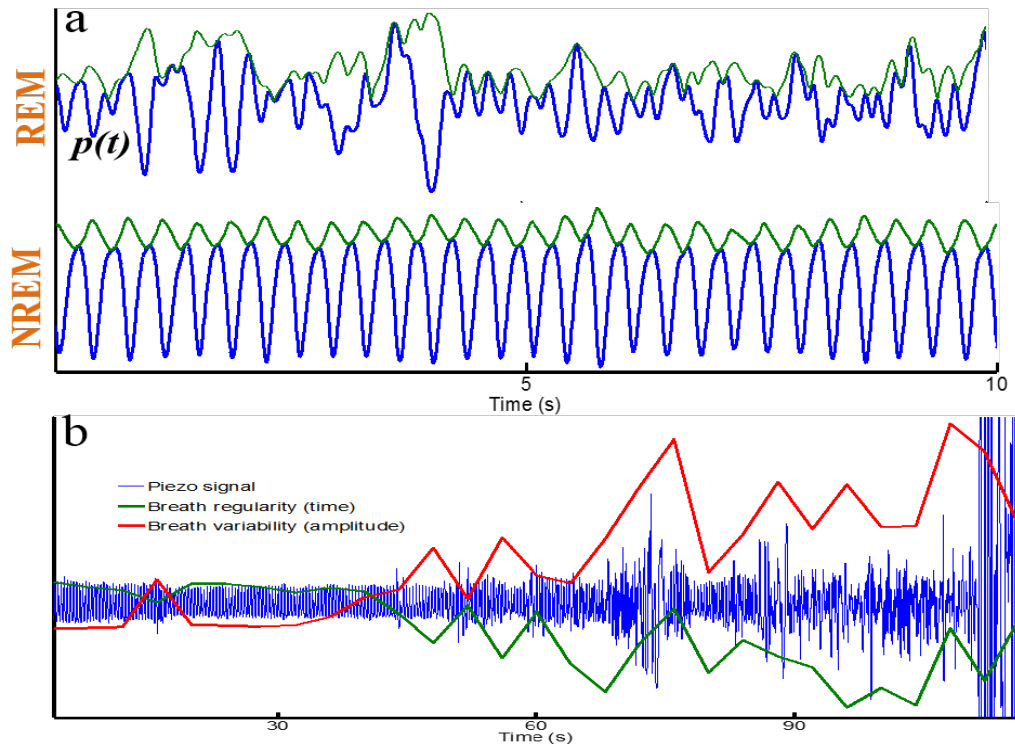
Figure 3.4 Estimation of respiration features from piezo signal. a) Breathing signal (lowpass-filtered piezo) and estimation of amplitude envelope during NREM and REM samples. Breathing amplitude changes significantly in REM while stays constant during NREM; b) Breath regularity/variability features (*BAV* and *BTR1*) in a sample sleep transition (NREM -->REM). Both breathing timing and amplitude become irregular at the transition to REM sleep.

The last approximate for *BTR* is based on the spectral analysis of breathing signal. Multitaper estimation of power spectrum can reflect the strength of a harmonic buried in noise (Thomson 1982). The same technique was used here to estimate spectral power of breathing signal. The spectral modal frequency and amplitude are equivalent to *IBR* and *BTR*3 respectively.

Automated sleep classifier: Piezo signal would be of super excellence if it could track sleep in large-scaled experiments without prior knowledge of the state (i.e. no EEG/EMG prerequisite). Thus, an unsupervised classifier becomes a desirable option. As we discussed it in previous chapter, unsupervised classifiers cannot be biased by human opinion to impose input data to regions with pre-defined boundaries in feature space. Actually they cluster input data based on intrinsic partitions in feature space without a

need for labeled training data. Similar to Chapter 2, hidden Markov models were implemented to extract mouse behavior characterized by motion and respiration features. Then, the correspondence between model output and vigilance states was evaluated. Although the modeling details were thoroughly explained in previous chapter, we review the general procedure briefly. Here we implemented HMMs to map sequential observations from piezo signal $O_t = \{TE, BTR1\text{-}3, BAV$ and $IBR\}$ to a set of discrete but 'hidden' states $S_t = \{$motion, regular breathing, irregular breathing$\}$. At any given time, the probability distribution of O_t is conditioned on the model state $P(O_t \mid S_t = i)$ and each model state has a prior probability $P(S = i)$. State transition matrix as the main feature of HMM gives the probability of a transition from state i to state j as $P(S_{t+1}=j \mid S_t=i)$. The HMM is derived from $O_{1:t}$ by maximum likelihood estimation and Markov property assumption: state S at any time t depends only on the state at $t\text{-}1$ and not on previous history. Once model parameters are determined, Viterbi algorithm is applied to decode the most likely sequence of states $S_{1:t}$ that generate new observations $O_{1:t}$.

We hypothesized that piezo motion and respiration features would separate sleep from wakefulness and REM from NREM respectively. To test this hypothesis, a correlation analysis between model prediction and human-scored vigilance state was done using performance analysis metrics and potential sources of error were also discussed.

3.4. Results and discussion

In this study, we used simultaneous recordings of EEG/EMG and piezo from 20 adult mice with 24-hour duration of each. Piezo features were first calculated in each individual mouse and then modeled using HMMs. Each trained model was also used to decode the most likely sequence of vigilance state corresponding to the feature set of that specific subject. Finally, correlation between model output and true states was evaluated using available labels for 4-s epochs by two independent scorers (consensus hypnogram considered as the reference). The performance analysis was assessed base on confusion matrix analysis (e.g. sensitivity, specificity and Cohen's Kappa) as well as sleep parameters estimations (e.g. time spent, mean bout duration and number of bouts in each state). A flow diagram of steps above is shown in Figure. 3.5.
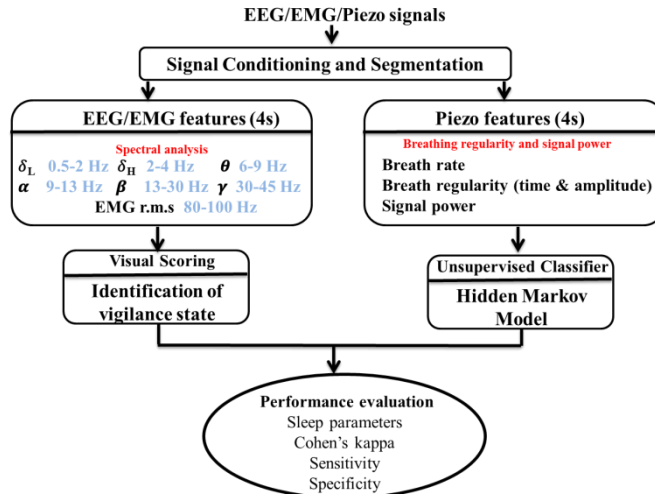
Figure 3.5 Overall procedure for noninvasive mouse sleep scoring

Figure. 3.6 demonstrates a 6-hour sample of piezo features and HMM output along with two hypnograms from independent raters. To estimate sleep parameters, all three hypnograms were filtered first to retain only prolonged bouts of sleep or wake (longer than 5 minutes). The binary hypnogram corresponding to scorer 1 is shown in Figure. 3.6. Confusion matrices on the top of Figure.3.6 show high sensitivity (~>90%) for all states except REM. Thus, we have an HMM that noninvasively classifies a set of piezo signal features (motion, breathing regularity) into behavioral states that are roughly equivalent to REM, NREM, and Wake. Though the accuracy for REM is relatively low, this could serve as a first-pass screen for a large cohort to noninvasively select potentially interesting sleep phenotypes (e.g., high REM/NREM ratio) for further analysis.

Irregular breathing during REM—just like the eye movement itself—is episodic and need not necessarily last the entire of a REM bout. However, these irregularities indicate an elevated probability of REM occurrence. This likelihood is captured here using hidden Markov models. A closer view on Figure.3.6 reveals that about 55% and 13% of detected REM episodes were actually NREM and Wake states respectively. So, REM sensitivity and NREM specificity both suffer from a misclassification error.
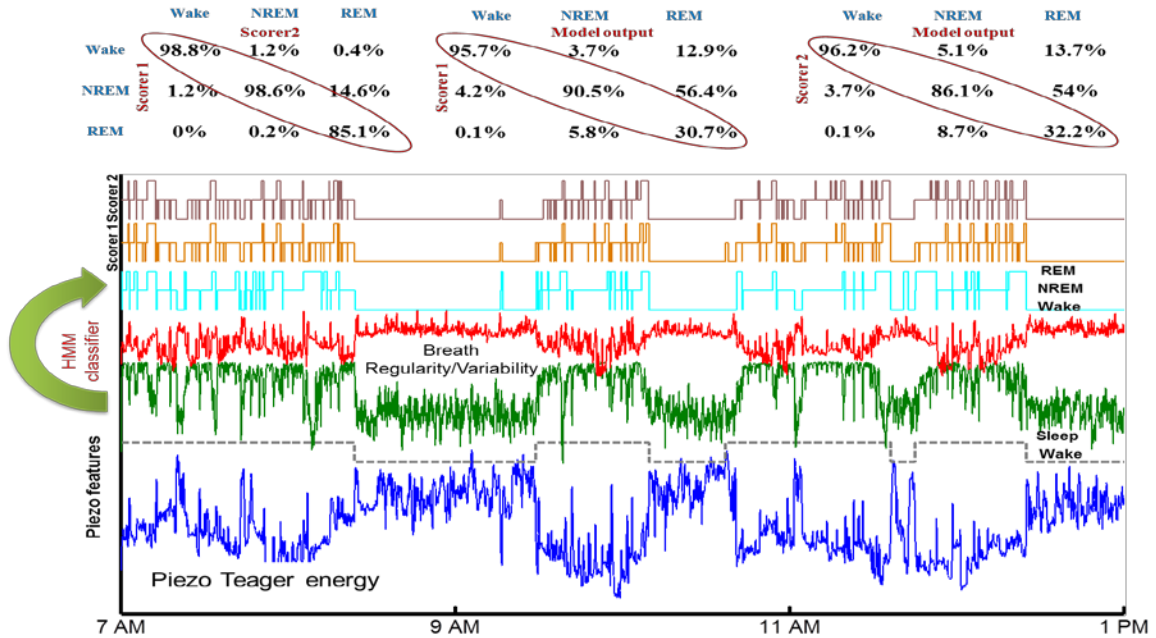
Figure 3.6 Piezo features and automated scoring. Piezo *TE* (blue) correlates with muscle tone and separates sleep from wake. Breath regularity (green) and variability (red) features are used to discriminate NREM from REM during sleep. HMM prediction of sleep-wake states is close to manual scores except that REM sensitivity is moderate.

Based on defined features for piezo signal, unsupervised HMM tends to discriminate regular and irregular breathing when animal is asleep. However, when we compared HMM-predicted states with manual scores, it seems to overestimate REM (including irregular breathing) with a portion of NREM (including regular and irregular breathing) sleep. It is known that during NREM sleep, breathing regularity increases with the depth of sleep (Long et al. 2014). Then, irregular breathing during light stages of NREM sleep might be a potential source for misclassification error for REM detection.

To evaluate the impact of light sleep on REM misclassification, average trend of piezo (*TE* and *BTR*1) and EEG (δ/θ power ratio) features for 20 mice were plotted time-locked to onset of prolonged sleep determined by human scorer. Figure. 3.7 (top) shows these trends in a one-hour time period centered at sleep onset. REM FP (or classification error) is also defined as the probability of falsely detected REM state by unsupervised HMM. At the beginning of sleep bout, *TE* suppresses because of elimination of muscle activity, and stays low for at least 10 minutes and then increases gradually as possible transitions to REM or Wake happen. Trends for *BTR*1 and EEG δ/θ power ratio are similar and start rising at sleep onset. However, there is a distinct surge in REM FP right at the sleep

onset and falls back to the baseline level gradually.
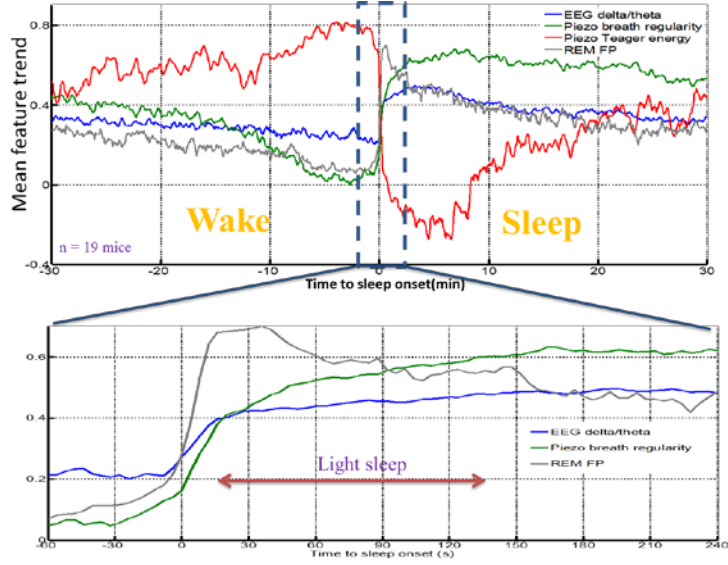
Figure 3.7 Mean piezo feature trends at sleep onset. REM specificty is low at the beginning of sleep onset where delta portion of EEG and piezo regularity are still low. Unsupervised HMM integrates this transitional state (i.e. light sleep) and REM sleep as one state.

Figure. 3.8 (lower) shows EEG δ/θ power ratio, $BTR1$ and REM FP in a 5 minute interval (sleep starts at minute 1). Mean feature trends at sleep onset revealed a gradual increase and saturation of EEG δ/θ power ratio that is mirrored by piezo breath regularity. REM detection specificity appeared to be lower at the onset of sleep following prolonged wakefulness. This period of light sleep is marked by an elevated REM false positive rate. Hence, the unsupervised HMM seems to integrate light sleep and REM into one state based on their irregular breathing patterns.

To overcome the misclassification problem, a supervised HMM was estimated and applied using n-fold cross-validation (n = 20 mice). HMM parameters were estimated using piezo features and available labels from *n-1* mice. The model then validated using piezo features from left-out test animal. Agreement between the HMMs and rater consensus scores was assessed in terms of Cohen's kappa. Cohen's kappa κ is a measurement of inter-rate agreement that corrects observed agreement (P_o) for chance agreement (P_e):

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (6)$$

κ larger than 0.8 indicates perfect agreement, 0.6 to 0.8 indicates substantial agreement, 0.4 to 0.6 indicates moderate agreement and less than 0.4 indicates slight or fair agreement between scorers. Figure. 3.8 summarizes the distribution of κ for 20 mice when we compared individual scorers with each other and model outcome with consensus scores. The range of κ for unsupervised HMMs falls within the moderate level of agreement while inter-rates agreement is almost perfect (>0.9 in average). However, supervised HMMs improved κ to the substantial level.
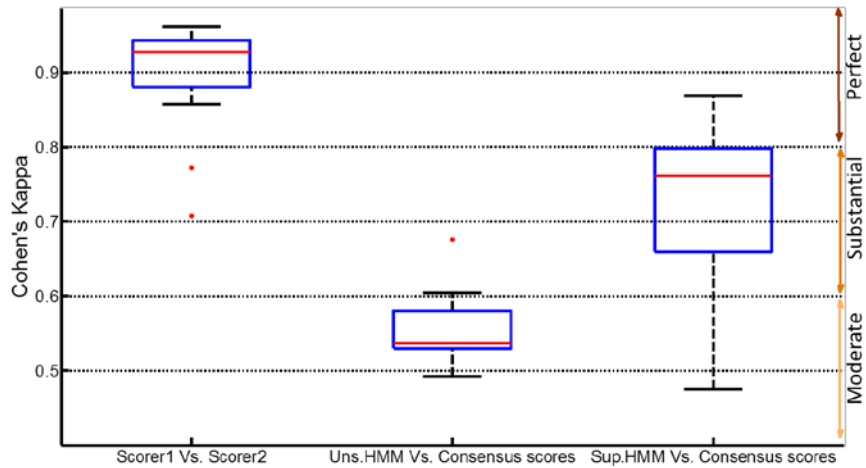
Figure 3.8 κ ranges for inter-rater agreement and unsupervised/supervised classifiers. Supervised HMM significantly improved the agreement to the substantial level.

A more detailed performance analysis was also done using state-dependent metrics from confusion matrix that are sensitivity and specificity (see Table. 3.1). As we discussed earlier, NREM sensitivity and REM specificity are the main limitations of the unsupervised classifier. Actually, the major source of misclassification error is confusion between NREM and REM states. According to Table. 3.1, supervised HMMs significantly improved NREM sensitivity and REM specificity while the rest of parameters are either remained unchanged or slightly improved.

Table 3.1 Comparison of sensitivity and specificity of classifier. Supervised HMM improved NREM sensitivity and REM specificity.

	Sensitivity (mean±s.e)			Specificity (mean±s.e)		
	Wake	NREM	REM	Wake	NREM	REM
Unsupervised HMM	88±1	51±1	73±3	96±1	96±1	75±1
Supervised HMM	90±2	80±3	66±3	91±2	93±1	92±1

Three commonly used metrics of sleep structure were also compared for Sleep and Wake bouts as well as each vigilance state within Sleep. These parameters include portion of spent time, mean bout duration and number of bouts in each state and estimated independently from the HMMs and manual scores. Sleep and wake bouts were first extracted by filtering scoring vectors (retain only episodes with a minimum of 5-minute length) and then finer comparison within sleep stages including NREM, REM and brief arousal (short wakefulness episodes) was done. As it is shown in Figure. 3.9, prediction of all parameters for sleep and wake bouts is almost the same as the ones obtained from manual scores. Similar analysis for sleep states revealed that supervised HMM gives better estimates of time spent and number of bout per state while the unsupervised HMM gives better estimates of mean bout duration.
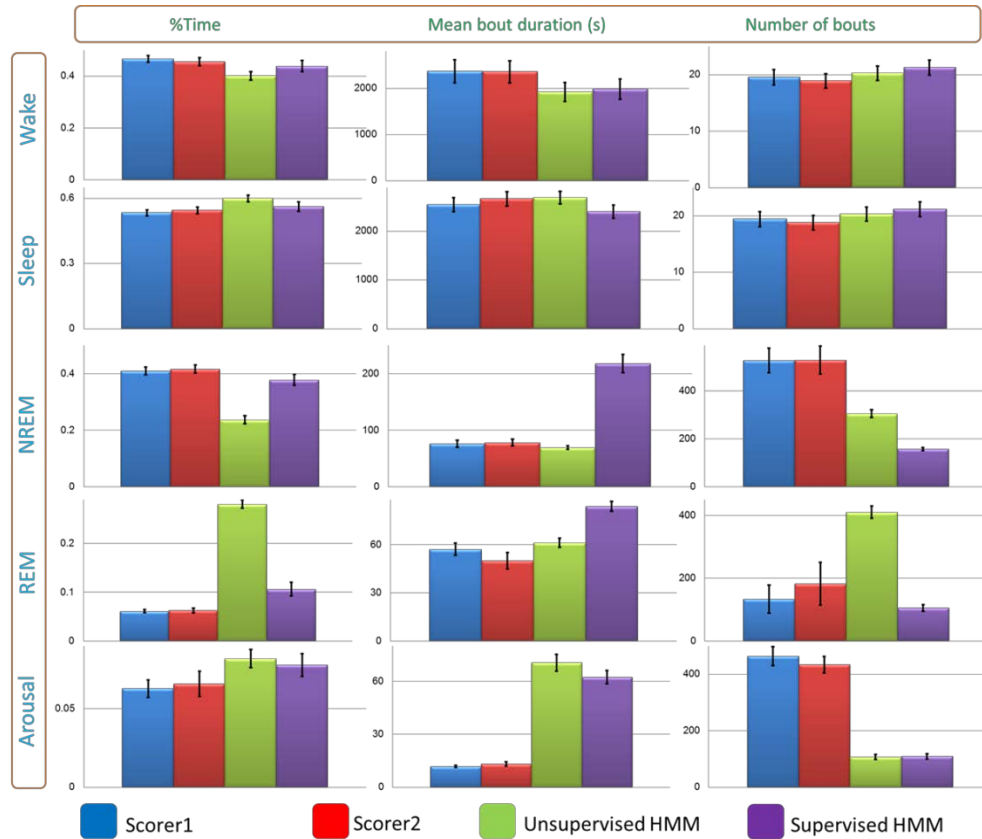
Figure 3.9 Different sleep-wake parameters are compared between vigilance states. Supervised HMM gives better predictions of % time and number of bouts while unsupervised HMM predicts mean bout durations more accurately (Yaghouby et al. 2015b).

Actigraphy, or measuring the levels of activity, in mice can be done as wheel running, photoelectric beam breaking, videography or accelerometry. The piezo system essentially overcomes limitations of actigraphy; since it is completely noninvasive and nonintrusive and provides a continuous record of sleep-wake activity. It was successfully applied to discriminate sleep from wake in mice with 95% accuracy (Donohue et al. 2008). In wakefulness, gross motor activity produces clear variable signals (Figures. 3.1 and 3.2). Even quiet wake (QW) typically has grooming, postural adjustments, or other fine movements with distinctive signatures captured by the sensor (Donohue et al. 2008).The piezo signal not only discriminates sleep from wake based on levels of activity, but also provides an excellent respiratory trace during sleep (Friedman et al. 2004 and Flores et al. 2007). The piezo sensor therefore offers the unique opportunity to noninvasively distinguish sleep from wake based on small movements undetectable by other techniques,

and subtle changes in breathing when the animal is relatively motionless. A few studies have recently tried to validate piezo merit in mice sleep/wake detection against EEG/EMG. One study (Sato et al. 2014) examined a similar piezo system against EEG and EMG in control and narcolepsy mice and reported 70% agreement for sleep/wake discrimination. Another study in mice validated the ability of piezo signal in sleep/wake discrimination with simultaneous EEG/EMG recordings and claimed the potential for finer discrimination of sleep stages (Meng et al. 2013).

Here we demonstrated that HMMs estimated exclusively from features of the piezo signal can be used to discriminate between stages of mouse sleep noninvasively. The ability to study sleep structure without EEG facilitates investigation of the genetic basis of sleep disorders such as sleep fragmentation in the elderly or REM sleep behavioral disorder. Specificity of REM classification by the unsupervised HMM suffered due to the similar breathing variability observed during light NREM at the onset of sleep. However, a supervised HMM largely alleviated this problem. Although unsupervised prediction of sleep metrics (e.g. time spent and bout duration) is less than perfect, it may be accurate enough to distinguish outliers in a cohort on the basis of a sleep trait, which is the goal of many behavioral assays.

The piezo system is a beneficial tool in epilepsy research as well. It could not only be used to track vigilance dynamics in the absence of EEG (or EEG good quality), but also as a novel biomarker to detect seizures noninvasively. In Chapter 7, we will discuss a few applications from the piezo system in epilepsy research in mice.

4. CHAPTER IV MARKOV MODELING OF SLEEP DYNAMICS FOLLOWING NEURAL INJURY

4.1. Rationale

Understanding how behavior changes over time as a consequence of degenerative neural disorders or acute neural injury involves tracking sleep-wake metrics and other markers (e.g., behavior, cognition, seizures) etc. at various points in time during disease progression in animal models. The previous chapter showed that HMMs can be useful for tracking instantaneous sleep-wake state with reasonable accuracy from continuous invasive (EEG/EMG) or even noninvasive (piezo) measurements. From the output of these classifiers, conventional sleep metrics like the % time spent, mean bout duration, and number of bouts of each state can be derived for the period under investigation.

While the methodology presented in Chapters 2 and 3 perform quite well on control animals and when behavior is stable over the diurnal cycle, it remains to be seen how it would perform when there is progressive change in behavior over time. Convenient metrics that track sleep-wake dynamics over time—beyond simplistic measures such as the percent time spent in each state or mean bout duration—are lacking and presume the ability to accurately predict the instantaneous state in chronic recordings. Disturbances in sleep and diurnal rhythms (e.g. insomnia) are common following traumatic brain injury (TBI) in 30-70% of patients and potentially undermine patient rehabilitation and recovery (Orff et al. 2009). Interestingly, recent studies have indicated that acute or mild TBI may be more associated with increased likelihood of sleep disturbances than severe forms of TBI (Orff et al. 2009). Recovery from acute TBI is also critical for avoiding development of chronic disorders such as epilepsy. The ability to track changes in the microstructure of sleep in the post-traumatic period could help assess the efficacy of intervention and perhaps offer clues about the likelihood of epileptogenesis. In this chapter, we investigated the utility of an unsupervised methodology based on hidden Markov models—estimated from invasive (EEG, EMG) and noninvasive (piezo) measurements in mice—for tracking and characterizing progressive changes in sleep-wake dynamics in the

period following neural injury.

4.2.   Animal model of acute brain injury

Invasive implantation of EEG/EMG electrodes in mice was explained in Chapter 2. It is usually recommended to give animals at least 10-15 days for recovery from surgery before recording any baseline data or performing investigative experiments. Such a craniotomy, performed here for implantation of EEG/EMG electrodes and headmount, is often used as an experimental control for investigations of brain injury in mice. In this experiment, with IACUC approval, adult C57BL/6J mice (n=6) were implanted with EEG/EMG preamplifiers and monitored immediately round-the-clock for three weeks to record and analyze progressive changes in sleep-wake dynamics.

4.3.   Model structure and features reflecting dynamics

The HMM's ability to capture brain dynamics associated with the sleep-wake cycle without human supervision makes it an appropriate tool for automated sleep scoring. The use of the HMM as an unsupervised classifier was previously validated for sequencing prolonged time series of continuous features (EEG/EMG or piezo) into discrete states, specifically NREM, REM, and Wake. The underlying Markov chain model has also been shown to reasonably represent sleep dynamics: discrete states during sleep follow a trajectory whose likelihood at any given time depends on the previous state. A graphical representation of mouse sleep dynamics is given in Figure. 4.1. The prior probability of each state and transition probabilities between them (shown by arrows in the Figure) can be extracted by HMMs. Hence, an HMM is parameterized by a vector of marginal state probabilities (P) and a matrix of state transition probabilities (S).
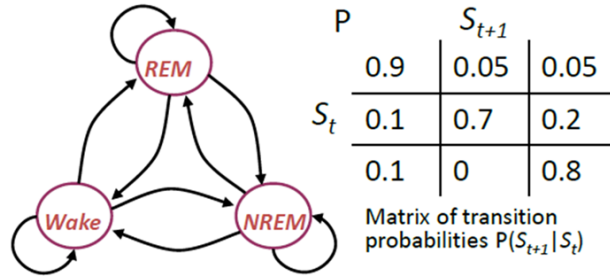
Figure 4.1 Graphical representation of sleep dynamics in mice. HMM models sleep dynamics using two parameters: a vector of marginal state probabilities (P) and a matrix of state transition probabilities (S).

These parameters can provide subtle metrics that quantify sleep quality; and continuous reestimation of the parameters over time may be useful in detecting and tracking subtle dynamical changes in behavior following an acute insult. For example, the trace T_r of the state transition matrix -- i.e., the sum of the diagonal elements -- conveys the average probability that any state will persist uninterrupted. This could be considered as a measure of sleep fragmentation: low T_r indicates lower probability of an auto-transition from each state; while high T_r reflects increased probability of transition between states. Similarly, the prior probability of each state (e.g., P_w for Wake probability) also conveys the proportion of time spent in that specific state. Trends in T_r and P_{state} reflect progressive changes in sleep-wake dynamics and can be used to track sleep quality and diurnal trends during recovery from injury.

4.4. Results and discussion

To investigate sleep dynamics after neural injury n=6 mice were continuously monitored, except short interruptions for cage cleaning and data downloads, following implantation surgery for three weeks. We recorded EEG/EMG signals simultaneously with a piezo signal (see Chapter 3). The same features used in previous chapters were estimated in 4s epochs for the recorded signals: spectral bandpower features from EEG/EMG and motion/respiration features from the piezo signal. Reestimation of HMM parameters over time (every 4 hours) yields a time series of metrics reflecting sleep quality. In this chapter, we introduced T_r and P_{state} as metrics that reflect evolving sleep-wake dynamics

across the diurnal cycle. We estimated these parameters from HMMs fitted to invasive (EEG/EMG) or noninvasive (piezo) signal features. Figure. 4.2 shows how HMM parameter time trends, estimated from EEG/EMG features, can track sleep dynamics across the light/ dark cycle.
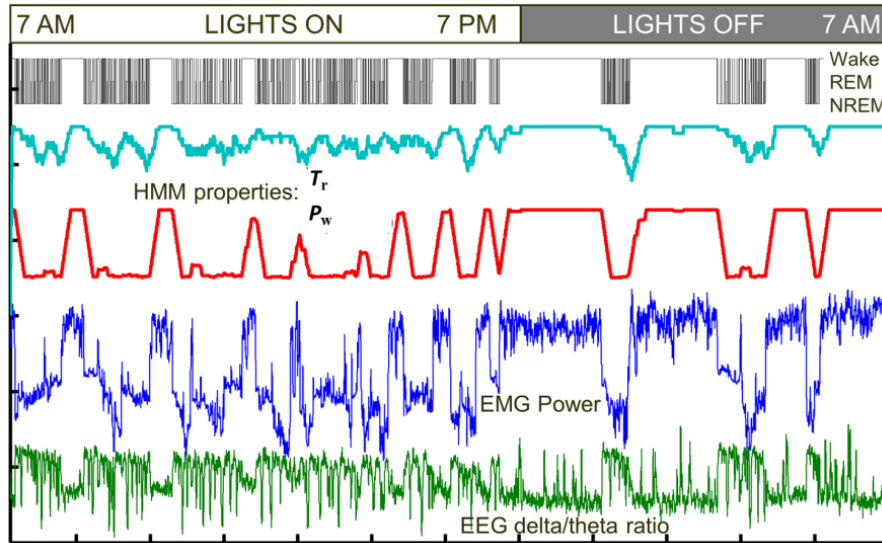
Figure 4.2 HMM metrics from EEG/EMG features track changes in sleep dynamics. T_r reflects fragmentation of sleep that stays low during light period in which several transitions happen during sleep. P_w is probability of wakefulness that is higher in dark period (Yaghouby et al. 2013).

EEG delta/theta power ratio and EMG power were estimated from 4-s epochs of a mouse recording in a 24-hour period. These feature have been used to construct an HMM that generates automated scores as a hypnogram in Figure. 2.7. Two HMM properties were estimated from this hypnogram as follows: the original hypnogram is binarized so that epochs with Wake label were 1 and the rest remained 0. The probability of Wake (P_w) is estimated as smoothed version of this binary vector (15-min moving average filter). T_r was also estimated similarly: a binary vector is derived from the hypnogram in which epochs at state transitions are set to 0 and the rest to 1. A 15-min moving average was used to extract the T_r trend from the binary vector. As shown in Figure. 4.2, P_w is high corresponding to prolonged wake bouts while high T_r reflects a stable vigilance state in the dark period. However, during the light period we have several sleep-wake cycles and

41

T_r is correspondingly lower.

During the monitoring period following brain injury, HMMs were derived separately from EEG/EMG and piezo feature sets every four hours using a maximum likelihood estimation procedure(Rabiner 1989). Figure. 4.3 demonstrates trends in HMM parameters estimated from EEG/EMG and piezo features in mice (n=6). T_r and P_w were tracked for at least three weeks after surgical implantation of an EEG headmount. The time traces are averaged over the six animals. T_r reflects the probability that any state will persist uninterrupted. As can be seen in Figure. 4.3a and Figure. 4.3c, T_r is low following surgery, indicating that sleep is more fragmented, but recovers to a stable baseline (dashed line) about a week after surgery. However, clear diurnal rhythmicity is not evident until Week 3. For HMMs estimated from piezo signals, recovery time is almost the same as for the EEG/EMG-derived HMMs but T_r periodicity is not that evident (Figure. 4.3c). One reason could be differences in sleep parameter values of the unsupervised HMMs estimated from EEG/EMG and piezo features. A quick comparison between Figure. 2.7 and Figure. 3.7 reveals the performance distinction: EEG/EMG models are highly accurate in estimation of all sleep stages while piezo models are not as good at distinguishing between NREM and REM during sleep.
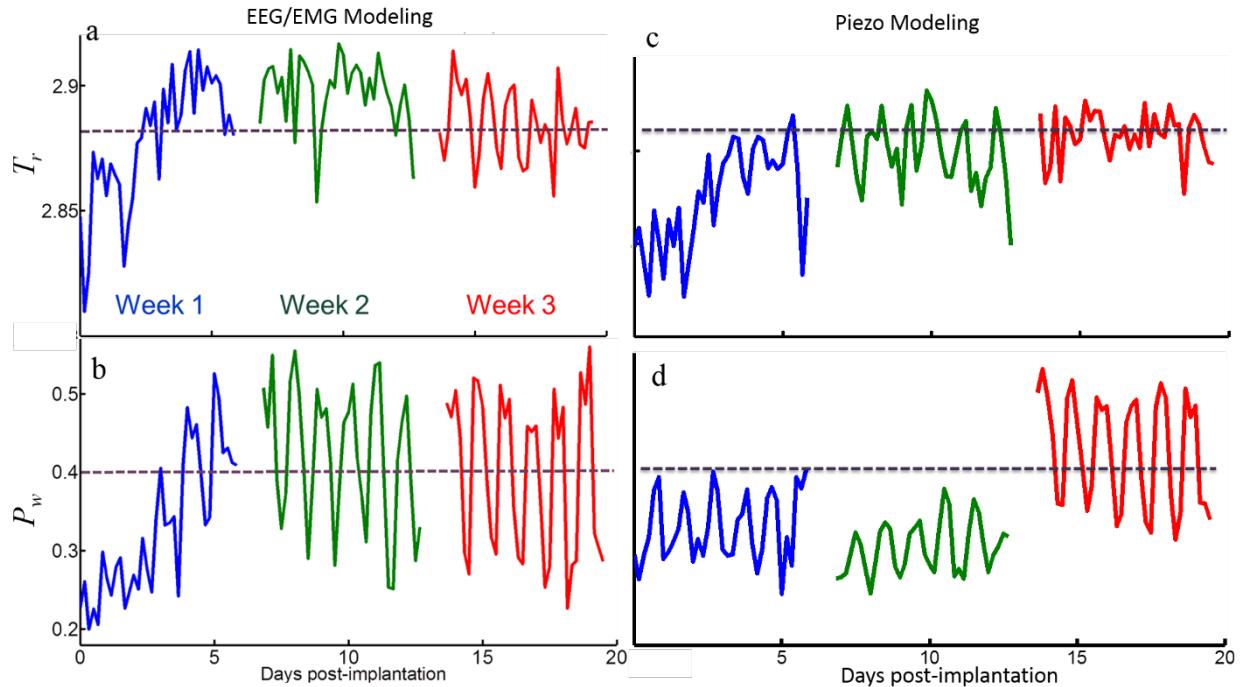
Figure 4.3 Tracking sleep recovery after neural injury. a and c (top): T_r is a measure of sleep fragmentation. It is low right after injury and increase gradually to the normal level with diurnal rhythmicity, particularly in EEG/EMG modeling, after 3 weeks. b &d (lower): P_w is percent time spent in wake that approaches to the baseline faster in EEG/EMG modeling. Trends representing means in 6 mice.

The prior probabilities of the HMM reflect the proportion of time spent in each of the vigilance states: for instance, P_w in Figure. 4.3, lower panel, conveys the % time in Wake state estimated by different models. This metric is low in the early phase of recovery, consistent with increased somnolesence, but approaches a baseline with prominent diurnal rhythmicity by Week 2 (Figure. 4.3b). The recovery time for P_w is delayed for the piezo HMMs and strong rhythms comparable to baseline are not evident until the third week. In summary, trends in Figure. 4.3 suggest increased but more fragmented sleep soon after implantation with persistent disruption in diurnal rhythm for up to two weeks. Slight differences in T_r and P_w trends between two models (EEG/EMG vs. piezo) during recovery is due to contrast in performance of unsupervised sleep classification. As discussed in Chapters 2 and 3, unsupervised HMM gives an accurate estimation of vigilance state when is modeled using EEG/EMG signals. While, performance is

moderate, particularly for NREM/REM discrimination, when an unsupervised HMM is fitted to piezo features.

However, in this specific application, though the piezo signal is a noninvasive measurement, it could potentially provide additional information that the EEG/EMG signals cannot. According to Figure. 4.3, it takes almost three weeks for a brain-injured mouse to recover to the stable dynamics of sleep-wake structure. But how do we know whether the stable sleep structure is the same as before the injury? In other words, even mild injury could induce adverse plastic changes in behavior and cause sleep quality to drift to another level permanently. It turns out that the piezo signal is a convenient tool not only to model control (i.e. baseline structure of sleep before the brain injury) where no EEG/EMG information is available, but also to track recovery from injury by comparison with the baseline. In Figure. 4.4 time trends of HMM parameters estimated from piezo features in a mouse are shown for up to five weeks: two weeks prior to surgery (baseline) and three weeks following the surgery (recovery). The animal was continuously monitored in the cage for two weeks (only piezo recording) and then implanted surgically. Monitoring was continued for three more weeks following the surgery using both EEG/EMG and piezo recordings. Baseline modeling prior to surgery shows a fairly high T_r with slight but visible diurnal variation. After surgery, T_r dramatically reduced outside the bounds of baseline circadian variation and recovers over time (Figure. 4.4a). At the end of Week 5 (3rd week following injury), T_r stays stable but the level is slightly lower than baseline.
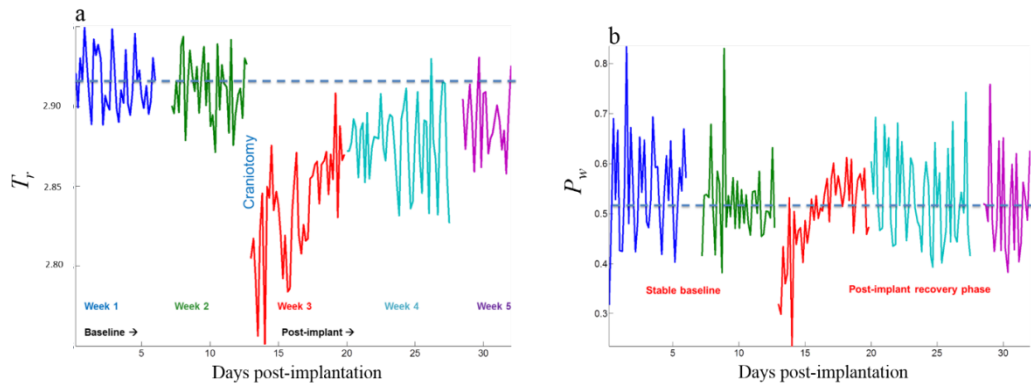
Figure 4.4 Piezo HMMs enable us to model baseline sleep dynamics prior to injury. a) T_r

shows a stable variation during baseline and progressive changes following injury. b) P_w
is recovered to the baseline level after one week (Yaghouby et al. 2012).

Similar to T_r, the proportion of time spent in the Wake state (or Pw) is high in baseline
with clear rhythms reflecting light-dark cycles. Following the injury, it drops sharply and
rebounds thereafter in about a week. Trends for Pw, after this recovery period, are very
similar to the ones estimated during baseline. Interestingly, Pw recovers to baseline
within a week but T_r shows that sleep remains fragmented and asymptotically approaches
the baseline (dashed line) on a longer timescale (see Figure. 4.4). Similar to what we
observed for EEG/EMG modeling, these trends indicate that net sleep time increases after
acute injury but that sleep is more fragmented. This supports what we concluded in post-
surgery analysis (Figure. 4.3): although time spent in Wake (or sleep) rebounds fairly
fast, the sleep fragmentation (quality) might still remain as a problem.
Preliminary results of this study suggested that HMMs estimated from physiological
measurements could provide quantitative markers of behavior and recovery from brain
injury. Recovery from acute brain injury is critical for avoiding development of chronic
disorders such as epilepsy and cognitive impairment. The search for biomarkers that
index healing using animal models of brain injury involves extensive behavioral
screening via sleep-wake and cognitive analysis. However, properties of the HMM can be
used to track dynamic changes in sleep in an unsupervised manner. This could provide
useful quantitative behavioral correlates of epileptogenesis and recovery from injury.

5. CHAPTER V SELECTIVE SLEEP RESTRICTION IN MICE USING MILD SENSORY STIMULATION

5.1. Rationale

The available knowledge about mechanisms of sleep and circadian rhythms has been widely improved through experimental manipulation of sleep in animal models. Understanding the contribution of REM or NREM stages in sleep regulation would help analyze related disorders in humans. Total sleep deprivation (TSD) has been used as an experimental tool to investigate the consequences of sleep loss on mechanisms for several years. Research done in animal models revealed a prominent rebound in characteristic EEG rhythms during NREM (slow wave activity; SWA or delta: 0.5-4Hz oscillations) and REM (theta activity: 6-9 Hz oscillations) sleep following TSD (Schwierin et al. 1999). Intermittent interruptions in human sleep do not necessarily lead to total sleep loss. Hence, TSD may not be an appropriate model to investigate physiological consequences of partial sleep loss, and selective sleep deprivation has evolved as a tool in animal sleep research. Although restriction of any sleep state will result in rebound in that specific state during the recovery period (Endo et al. 1997), it could also affect other states as well. For example, REM sleep deprivation (REM SD) could result in both REM rebound and significant suppression of SWA in NREM (Endo et al. 1997). The first generation of selective sleep deprivation protocols in rodents relied on manual sensory stimulation such as gentle handling or cage movement to compare the role of sleep states on behavior and cognition; these are still in use (Mistlberger et al. 2002 and McCoy and Strecker, 2011). However, the need for human supervision and intervention remains the main constraint in such systems. REM sleep plays a substantial role in development of the brain and memory consolidation (Mallick et al. 2010). So, investigating the consequences of REM loss in rodents has been of much interest. For REM SD a technique called "the multiple platform method" was devised in which animal sits on one of multiple platforms surrounded by water. During REM sleep and due to muscle atonia, the animal falls into a basin and awakens (Mendelson et al. 1974). Although this technique is highly stressful, it deprives the animal of almost all REM sleep and has been widely accepted as an effective tool for REM SD in rodents. Recently, programmable

devices have been designed to predict the vigilance state automatically using real-time analysis of EEG/EMG signals and rouse the animal using different types of stimulation such as cage shaking, rotating disk or activating a stir bar on the cage floor (Sahu et al. 2013; Kushida et al.1989 and Wisor et al.2011). Although automated sleep restriction provides great flexibility in timing and amount of stimulation and reduces the effects of stress, each system has specific requirements for the cage and means of stimulation delivery. Convenience of use, cost, and efficacy still remain major concerns in such systems.

In this chapter, the feasibility of a novel sleep restriction technique for mice is presented. A closed-loop sensory stimulation system was designed to detect the onset of a targeted sleep state (REM or NREM) using EEG/EMG analysis and apply tactile stimulation in the form of mechanical vibration to the cage floor. Selective sleep restriction using this system can be implemented with greater flexibility over the sleep loss proportion while other states of sleep remain intact. Here, this system was employed to selectively deprive REM sleep, which could be of interest in epilepsy research as well.

The relationship between sleep and epilepsy is complex. In general, epilepsy can impair sleep quality and different sleep stages can have a protective or precipitating influence on seizure occurrence. It is also well known that SD increases the chances of seizure occurrence by activating IEDs (Niedermeyer, 1982). Increase in cortical excitability following total SD has been noted in some clinical studies, with a greater effect resulting from selective REM SD (Placidi et al, 2013). Since hyperexcitability in brain networks during the pre-seizure (pre-ictal) period has been observed and investigated as a means of anticipating clinical seizures, it seems reasonable to expect that selective REM sleep deprivation will lead to more seizures than a corresponding amount of NREM sleep loss (Figure. 5.1). To address this hypothesis, the closed-loop sensory stimulation system can be utilized to assess the effects of chronic REM sleep restriction on epileptic mice that express spontaneous seizures (e.g., the pilocarpine-treated mouse model (Cavalheiro et al. 1996)).
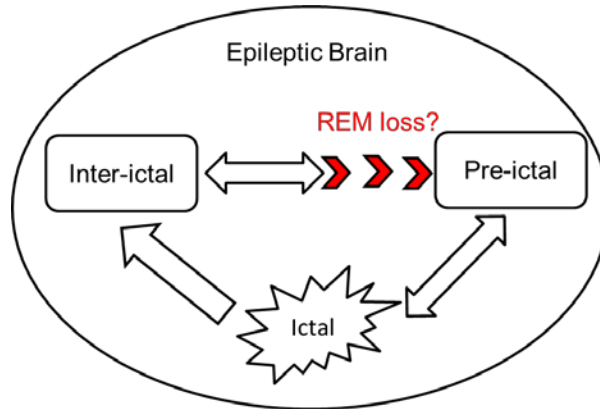
Figure 5.1 A possible model linking REM sleep loss, brain excitability and seizures. REM sleep restriction elevates cortical excitability and seizure likelihood.

5.2. Experimental design

Under IACUC approval, adult male wild type mice (C57BL/6J, 4-6 weeks old, n = 4) were surgically implanted with EEG/EMG electrodes. Prior to the experiment, each animal was kept in individual cages under consistent environmental conditions (light/dark cycle and ambient temperature/humidity) for 2-3 weeks of recovery. Details for animal care, surgery and acquisition system were discussed in Chapter 2. A 6-hour recording from EEG/EMG during baseline was acquired for each mouse (1-7 P.M.) and visually scored to train an automated REM sleep detection algorithm. Vigilance state was manually scored using defined criteria on EEG/EMG signals as Wake, REM and NREM. Hence, transition to REM sleep based on EEG/EMG feature boundaries was used to design the REM detector. Then the experimental session was performed later at the same time period as the baseline and for the same duration on a different day.

5.3. Vibro-tactile sensory stimulation system

In this experiment, a novel closed-loop sensory stimulation technique was proposed and tested to restrict sleep in mice. To study the consequences of sensory stimulation on different states of sleep, a tactile sensory stimulation system was devised using a micro-vibration motor attached to the floor of the animal's cage. This vibration motor was controlled by a computer program and generated mild intermittent sensory stimulation (MISS) that perturbed the animal's behavior. In an open-loop test, the motor was

controlled to produce a train of consecutive vibration pulses (three pulses with 1s duration each at intervals of 1s) at 15 min intermittent intervals. Along with this open-loop sensory stimulation protocol, cortical EEG and dorsal EMG signals were also recorded in mice as described in Chapter 2.

For closed-loop sensory stimulation applications, e.g. REM sleep deprivation, we designed a vibration platform by embedding eight equally spaced button microvibrator motors to a square rubber pad. This platform was attached to the base of the animal's cage and controlled by a computer program to deliver mild tactile stimulation to the animal. Each motor (10 mm diameter, 2 mm thickness) vibrates with amplitude of 0.75 g force at 12,000 r.p.m. when driven by a DC voltage (Pololu Corporation, Las Vegas, NV, USA). A driving pulse can be generated and applied to the vibration pad after a preset threshold on bout duration of any sleep state (e.g. REM) has been crossed. This real-time detection was performed by online processing of the EEG/EMG signals from the analog output of the amplifier (part # 8206, Pinnacle Technology). The acquisition system also has a screw terminal to enable external control and analog signal output. The EEG/EMG signals at these terminals were analyzed in real time for closed-loop stimulation for REM SD. EEG/EMG signals were routed to a data acquisition board (NI USB 6211, National Instruments) and digitized at 16-bit resolution and 400 samples per second. Digitized EEG/EMG signals were recorded and processed online using a custom VI to detect REM sleep and activate the stimulation system by sending a DC pulse to the vibration pad at specific times. The trigger signal was recorded in synchrony with the EEG and EMG as an additional data stream to the data acquisition board. The Labview program analyzed every 1s epoch of EEG/EMG signals (after 4-s moving average filtering) and estimates spectral band power features. The stimulation automatically stops when the state has changed and animal is awake (see Figure. 5.2).
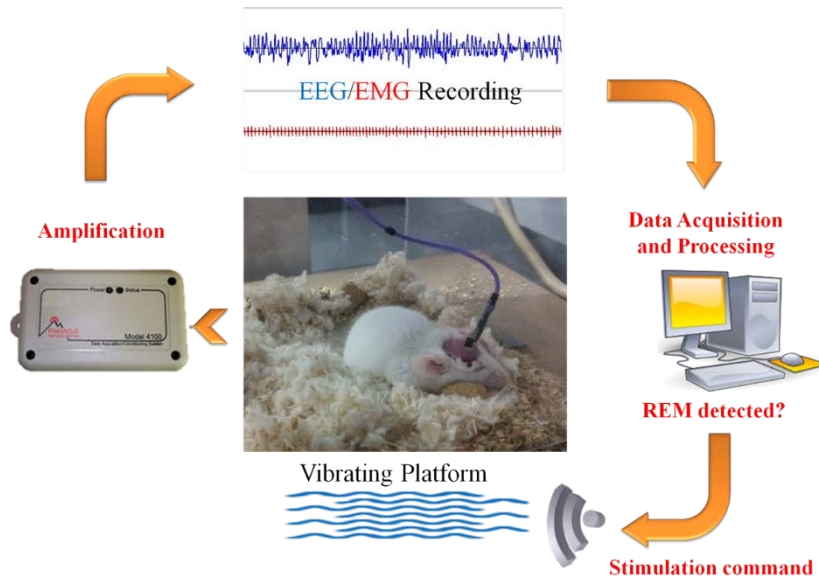
Figure 5.2 Closed-loop REM detection and restriction system. A computer program analyzes EEG/EMG signals and delivers stimulation command to vibrating platform once REM is detected (Yaghouby et al. 2014c)

Transition from wakefulness to REM sleep is not likely in normal mice; so the stimulation remains off until animal experiences a NREM to REM transition in sleep after normal sleep-wake cycles.

5.4. Sleep scoring and REM detection algorithm

A baseline recording was performed in each animal following two days of acclimatization to the cage environment. Then, training data to determine transition to REM sleep based on EEG/EMG features were selected from the baseline recording. Manual scoring of baseline sleep was based on visual inspection of video-EEG in 4 s epochs. Spectral band power features from EEG and EMG were used to construct an automated REM sleep detector in each animal. The mean power from bandpass-filtered EMG (80-100 Hz) reflects muscle tone and was used to discriminate sleep from wakefulness. During sleep, the EEG δ/θ band power ratio was estimated to detect the REM sleep onset: elevated theta rhythm in REM with reduction in delta activity reduces δ/θ intensely. Hence, REM detection thresholds were established using both features from the baseline recording and visual scores. These feature thresholds were used to

detect REM sleep onset during the REM SD experiment.
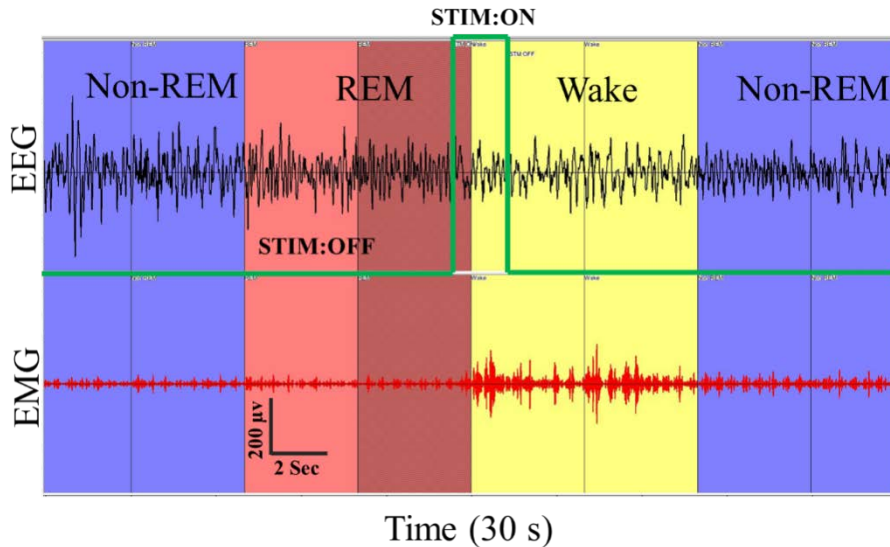
Figure 5.3 Tactile stimulation interrupts REM sleep. A brief stimulation pulse terminates REM sleep and animal falls back into NREM after a brief arousal.

5.5.   Results and discussion

Results for this section are presented in three parts. First, the state-dependent effect of sensory stimulation is demonstrated from an open-loop experiment. Then we evaluate the performance of the real-time REM sleep detector for each animal. Finally, the effect of closed-loop sensory stimulation on REM sleep restriction is discussed.

Open-loop sensory stimulation: Figure. 5.4 shows snapshots of the MISS protocol on mouse sleep and behavior. A vibrating micromotor attached to the cage generated three 1s pulses at 15 min intervals. Behavior was monitored by cortical EEG (top trace), dorsal EMG (middle), and a piezo (bottom) signal. Duration of each segment is about 10s and vertical lines mark onset and offset of the stimulation. As can be seen in Figure. 5.4 (top), MISS briefly interrupts NREM sleep. The slow delta oscillation on EEG and regular breathing signal on the piezo is interrupted by stimulation, but the animal returns from the brief arousal back to NREM sleep. Middle Figure shows how MISS disrupts REM sleep. Rhythmic theta EEG waves and relatively irregular breathing in REM are abruptly

51

terminated as the animal is roused by the stimulus, which is identical to that applied during NREM above but with a dramatically different response. This is to be expected as sensory thresholds differ for REM and NREM sleep (Rechtschaffen et al. 1966). The bottom part of Figure. 5.4 demonstrates the animal response to MISS during wakefulness. Characteristic EEG and EMG of Wake are unchanged by stimulation, which proves that MISS has minimal effects on Wake.
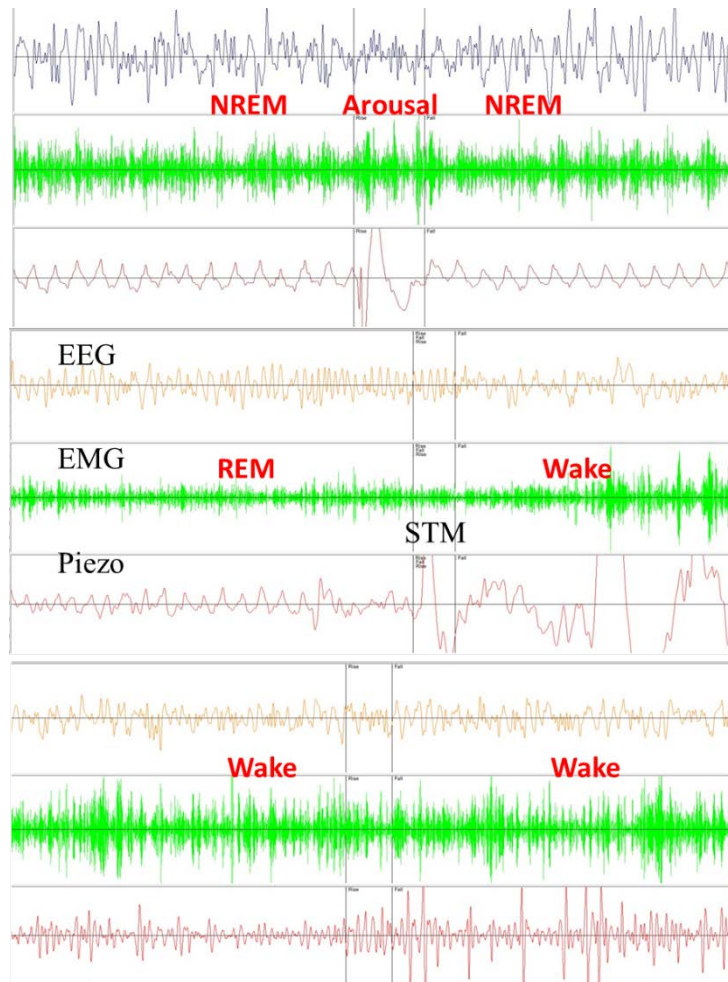
Figure 5.4 Effects of sensory stimulation on vigilance state. MISS briefly interrupts NREM sleep (top), disrupts REM sleep (middle) and has no effect on Wake (bottom).

The average effect of open-loop sensory stimulation on each vigilance state over n=4 mice was also investigated. Mice were stimulated in open-loop mode (a 1s pulse every 15 min) over a 24-hour period. Figure. 5.5 shows dorsal EMG power time-locked to the

stimulation onset and averaged over multiple stimulations during each vigilance state. Compared to what was seen in Figure. 5.4, consistent awakening is observed from REM sleep, transient arousal from NREM sleep, and no perceptible change in Wake.
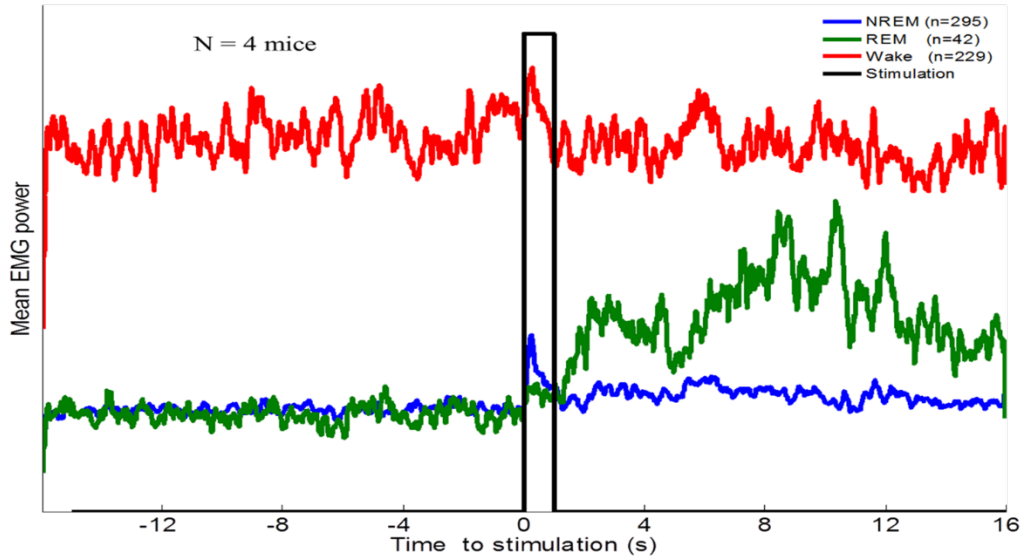
Figure 5.5 The average effect of sensory stimulation on each vigilance state. Consistent awakening from REM, transient arousal from NREM and no perceptible change in Wake are observed.

Result from open-loop sensory stimulation experiment reflected the difference in sensory thresholds for REM and NREM sleep and suggested that this technique could be useful for closed-loop REM SD in mice.

Real-time REM sleep detection performance: To evaluate the performance of our designed real-time REM detection algorithm, a human scorer blind to the state of the stimulation inspected EEG, EMG and video data for each mouse during stimulation and determined vigilance state in 4 s epochs. The closed-loop sensory stimulation was performed for each mouse over a 6-hour period in the afternoon (lights on). Then, stimulation onset and offset times were extracted and used to assess the performance of REM detection by comparing them with true REM bout onset and offset. The following events were first extracted based on this comparison for each recording: 1. True Positive (TP) events as the number of REM bouts that coincided with stimulation onset; 2. True

Negative (TN) events as the number of NREM or Wake bouts in which stimulation was off; and 3. False Positive (FP) events as the number of NREM or Wake epochs in which stimulation was on. Finally we combined these counts to estimate two commonly used performance measures:

Sensitivity = $TP / (TP + FN)$ (1)

Positive predictive value $(PPV) = TP / (TP + FP)$ (2)

In addition to the REM detection accuracy, detection latency is another important factor, particularly in real-time applications. Depending on choice of threshold, filter length and scoring resolution, there can be a delay from REM onset to stimulation triggering. This delay can be estimated as the time between REM onset and stimulation onset in TP events. Table 5.1 provides a summary of detector metrics. The REM detection sensitivity is high and over 90 %, except in Animal 1. However, the specificity (PPV) is moderate and ranged from 38% to 73 % in the same animals (Animal 1 has high specificity). The specificity range states that about one-third to one-half of all stimulations occurred during NREM or Wake. According to Figure. 5.5 tactile stimulation does not change the animal's state when awake. Thus, false stimulation could potentially affect NREM sleep.

Table 5.1 Evaluation of REM detection performance

Animal	Sensitivity (%) [Total REM bouts]	PPV (%) [Total detections]	Latency (s) (mean ± s.d.)
1	72.4 [254]	88.0 [209]	8.2 ± 4
2	93.5 [77]	38.3 [188]	7.9 ± 8
3	98.3 [114]	65.1 [172]	6.7 ± 4
4	95.7 [117]	72.7 [154]	6.7 ± 3
Mean ± s.e.m.	**95.8 ± 6**	**66.0 ± 10**	**7.4 ± 0.4**

The other metric in Table. 5.1 is latency. On average it takes about 7 s (or two epochs) for the detector to determine REM sleep and deliver a stimulation pulse. During manual scoring, brief or transitional REM episodes were scored as REM sleep while detection algorithm was tuned to detect only distinctive signature of REM sleep based on EEG/EMG features and defined thresholds. This could be one reason for this delay. As a

result, the closed-loop stimulation protocol only affected prolonged REM bouts and ignores brief episodes.

Closed-loop sensory stimulation for REM restriction: Analysis in the previous section revealed that the real-time REM sleep detector performed with reasonable sensitivity, specificity, and latency. To evaluate the effectiveness of closed-loop REM SD system, sleep parameters were estimated separately for the baseline and closed-loop stimulation recordings. Visual scores from each animal were used to extract the amount of time spent in each vigilance state as well as the distribution of bout durations in baseline and REM SD phases. Figure. 5.6 (top) shows the cumulative distribution of bout durations in different sleep/wake states where data from all animals were pooled together. As can be seen, closed-loop stimulation drastically reduced REM bout duration while the effect on NREM and Wake is relatively minor. The reduction in median REM bout duration, from 28s to 4s, supports our observation on protocol latency. The mean % time spent in each state (n=4 mice) is also shown in Figure. 5.5 (lower) for both baseline and stimulation phases. While % REM is noticeably reduced, the % time in NREM remained almost intact and % Wake increased. As discussed above, stimulation in NREM induced only brief arousal (Figure. 5.4 and Figure. 5.5), which explains why the slight reduction in NREM bouts is not accompanied by a significant decrease in % NREM. The brain's tendency to recover lost REM sleep following REM SD could be another possible reason for the shorter NREM bouts. The slight reduction in Wake bout duration might be due to an increase in brief arousals due to some false detections stimulated in NREM. However, as we pointed out, the greater % Wake matched the reduction in % REM: the amount of REM loss appeared to be compensated by a gain in Wake.
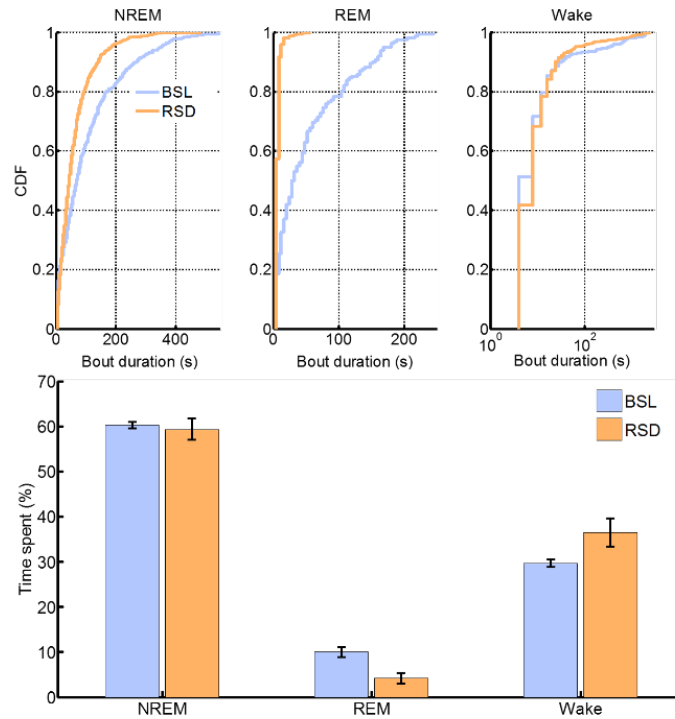
Figure 5.6 Effects of closed-loop sensory stimulation on sleep. Upper: Cumulative distribution function (CDF) of bout durations in each state. Lower: Comparison of time spent in each state for baseline and stimulation. Error bars represent standard error of the mean (n = 4).

Table 5.1 and Figure. 5.5 indicated that closed-loop stimulation protocol produced more than 50% reduction in REM sleep over a 6-hour period. The detection latency could be a reason that REM sleep is not eliminated in total. Here we implemented a simple linear threshold-based classifier to detect REM that was very sensitive to EEG/EMG signal quality. For example, poor detection specificity in Animal 2 was mainly related to original signal quality. Application of more robust machine learning techniques, such as support vector machines or hidden Markov models, could improve the real-time REM detection performance and consequently the efficacy of REM sleep restriction.

Adaptation to sensory stimulation: Although the open-loop stimulation protocol was applied for 24 hours on each mouse, the efficacy of the closed-loop system was examined for selective disruption of REM sleep in acute experiments with 6-hour durations. In addition to specificity limitations of our detection algorithm, adaptation to the tactile

stimulation may also affect performance. Depending on the frequency and duration of applied stimulation, the animal could become insensitive and unresponsive to it. Figure. 5.7 represents the mean % time spent in REM sleep for each successive hour of monitoring during the baseline and stimulation protocol. In the first hour of the experiment the % time in REM was lowest (1%). Over the next five hours of the experiment % time in REM increased and stayed at a relatively constant level (4-6%). In baseline, this value started at 10%, peaked in the mid-afternoon (14%) and dropped as evening approaches (consistent with the diurnal cycle). Hence, the amount of REM loss for the last two hours is low because the probability of REM occurrence was already low under baseline conditions.
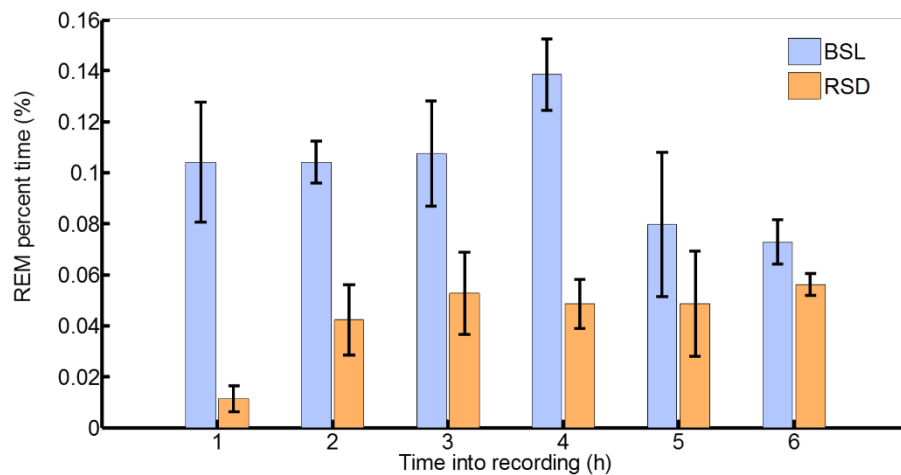
Figure 5.7 Trends in mean hourly percent time spent in REM during the baseline (BSL) and REM SD (RSD) stimulation protocol (Yaghouby et al. 2014b).

In conclusion, the efficacy of the closed-loop stimulation seems to be relatively stable except in the first hour when animal has experienced the stimulus at that first time. Hence, the animal remains responsive to the tactile stimulation in acute experiments without a change in threshold.

As explained at the beginning of this chapter, sleep deprivation is known as a provoking factor for seizures in different kinds of epilepsy. While the pathophysiologic mechanism of this effect is unexplored, selective sleep restriction in animal models can help investigators understand the underlying mechanism. Although seizure occurrence is, in

general, less likely in REM sleep, the effects of REM deficit on seizure incidents have not sufficiently been addressed. Thus, using the proposed system, the effects of selective REM sleep deprivation on seizure tendency and sleep microarchitecture in animal models can be assessed. Findings of such experiments will be useful for their relevance in human pathology.

6. CHAPTER VI INVESTIGATION OF SLEEP-SEIZURE INTERACTIONS IN A MOUSE MODEL OF TEMPORAL LOBE EPILEPSY

6.1. Rationale

In the preceding chapters, a set of experimental techniques and computational algorithms were proposed and tested for modeling and perturbation of sleep dynamics in mice. Although these novel tools are of obvious relevance to sleep research, the main purpose of this dissertation was to develop methods for facilitating animal research in epilepsy. The implications of the complex relationship of sleep with seizures in epilepsy were discussed in Chapter 1. The complex relationship between sleep dynamics and epilepsy is widely investigated in clinical and experimental studies. Pathological neural events like seizures can disturb the endogenous regulation of the sleep-wake cycle, and alterations in sleep regulation or poor sleep can precipitate seizures. In fact, vigilance dynamics biases the likelihood of seizure generation; and since seizure likelihood varies with vigilance state, it also affects the performance of seizure prediction algorithms (Schelter et al. 2006 and Bazil et al. 1997).

It is therefore important to investigate dynamical changes in the vigilance state and understand how they influence seizures or vice versa. This insight is clinically desirable for improving seizure diagnosis (or prediction) and treatment. The use of animal models, especially mice and rats, provides the opportunity to determine unknown neurobiological changes underlying sleep dynamics and seizure incidence. Data-driven algorithms developed in this dissertation (Chapters 2-4) have been shown to model vigilance dynamics in mice efficiently. In the current chapter, we try to benefit from similar tools for automated seizure detection as well. Furthermore, the proposed techniques can be applied to mouse models of epilepsy to study sleep-seizure interactions. This endeavor can be combined with closed-loop somatosensory stimulation system (explained in Chapter 4) to investigate the effects of programmed modulatory stimuli on seizure outcome and sleep quality in a chronic mouse model of epilepsy. This chapter of the

dissertation seeks to implement some of the proposed techniques in a chronic animal model of epilepsy and suggest future avenues of research.

6.2. Pilocarpine model of temporal lobe epilepsy

Temporal lobe epilepsy (TLE) is the most common type of focal epilepsy in adults. The main characteristics of TLE that can be reproduced in chronic animal models are: 1) Seizure foci located in the limbic system, specifically in hippocampus; 2) An initial precipitating event or injury is often found before the onset of TLE; and 3) A latent seizure-free period is observed after the precipitating event (Curia G. et al. 2008). Models of chronic epilepsy implemented by inducing status epilepticus (SE) in rodents as the precipitating event are widely used for preclinical analysis of mechanisms and treatment effects in several laboratories. In the pilocarpine model, SE is induced in the animal by injecting pilocarpine. After SE, there is a latent period during which neuroanatomical changes are observed, mainly in the hippocampal formation that lead to spontaneously recurring seizures (SRS) (Curia G. et al. 2008). Pilocarpine is a non-selective muscarinic receptor agonist in the parasympathetic nervous system which is widely used to induce chronic epilepsy in rodents. In this experiment, SE is induced in the animals with a single intraperitoneal (i.p.) injection of pilocarpine (290mg/kg). This dose has been shown to be effective with relatively low mortality rate (Shibley and Smith, 2002). Methylscopolamine is also injected 15 minutes before the pilocarpine treatment to suppress peripheral cholinergic effects (1 mg/kg) that could be fatal for the animal. Following pilocarpine injection, SE onset is identified by convulsive and intermittent seizures (Racine scale 3 and higher). According to guidelines formulated by Racine (Racine, 1972) seizures in rodents are categorized into five different scales based on observable behavior:

1: Mouth and facial twitches.

2: Head nodding movements.

3: Forelimb myoclonus.

4: Forelimb clonus with rearing

5: Generalized motor convulsions are observed as forelimb clonus, rearing and falling.

After pilocarpine administration, animals that survived and successfully developed SE with at least 3 convulsive seizures are considered for monitoring. The mortality rate during SE for this model is about 40% and injection dose plays a critical role (Shibley and Smith, 2002). Pilocarpine dosage should be large enough to induce verifiable hippocampal morphological changes (Curia G. et al. 2008). Hence a trade-off should be considered between the pilocarpine dosage and survival rate. SE monitoring after pilocarpine injection is performed for up to two hours and candidate animals are given softened feed soaked in sucrose water for a couple of days. Following SE, the animal enters a phase known as the latent or quiescent period. During this period, changes leading to the development of epilepsy (epileptogenesis) occur at the cellular level including: mossy fiber sprouting, interneuron loss, rewiring of synaptic circuits, glial cell activation and ectopic cell proliferation (Pitkanen and Sutula, 2002). Although some of these pathophysiological phenomena have been shown to be important in epileptogenesis, the animal generally shows normal behavior and brain electrical activity (EEG) during the latent period. Following a 4-6 week latent period, candidate mice are implanted with an EEG/EMG headmount and monitored for spontaneously recurring seizures.

Figure 6.1 Pilocarpine model of TLE. Animals develop spontaneous seizures after silent period

6.3. Automated analysis of seizures and vigilance dynamics

Once an animal survives SE and develops spontaneous seizures following a latent period, a chronic epilepsy model is available for use in experimental investigations. On average,

C57BL/6J mice have 2-3 tonic-clonic seizures (Racine 3 or higher) per day during their chronic epileptic phase (Shibley and Smith,2002). The first step in investigating the sleep-seizure interplay is to identify seizure onset. Figure. 6.2 demonstrates a set of physiological recordings--including EEG, EMG and piezo signals--during a spontaneous seizure in a mouse treated with pilocarpine. A seizure is typically characterized by concurrent high-amplitude and high-frequency epileptiform spikes on the EEG. This activity is usually accompanied by convulsions reflected in EMG and piezo signals as high muscle activity. Hence, EEG features that are descriptive of this seizure signature (large amplitude, rhythmic spiking) can be used to automate the seizure detection process. Here, we apply a simple threshold-based algorithm using EEG features to identify seizure candidates followed by a quick verification using raw signals and recorded video. Two commonly used features from the EEG signal have been chosen here to detect seizure onset: Teager energy (or *TE*) and line length (or *LL*).
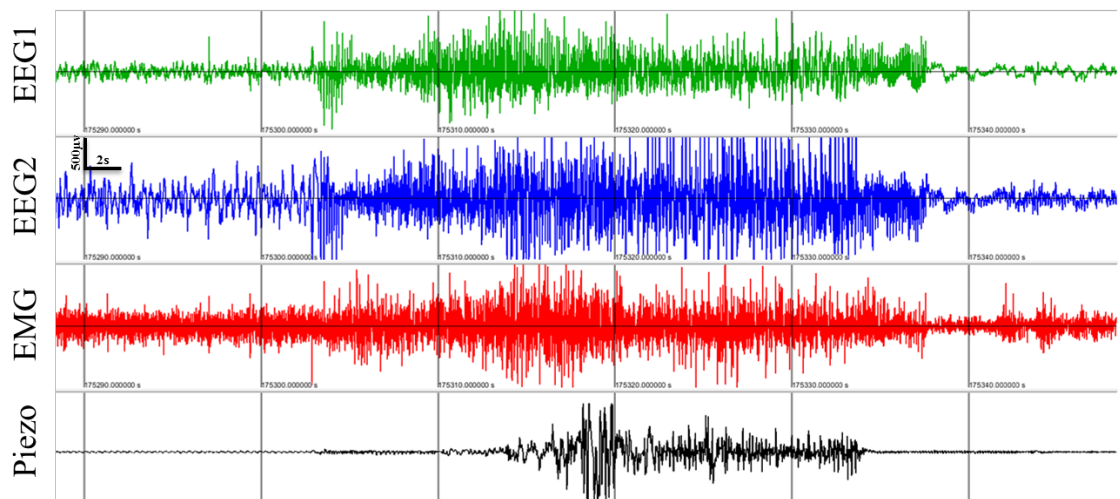
Figure 6.2 Electrophysiological recordings during a seizure incidence in mouse model of TLE. Concurrent high-amplitude and high-frequency EEG spikes are the main characteristics of seizures. This is also accompanied with convulsions obvious in EMG and piezo signals.

TE (for the piezo signal) was proposed in Chapter 3 as a surrogate for muscle tone (EMG signal). *TE* is a nonlinear operator that approximates the instantaneous energy of an

oscillator (Kaiser 1990) and reflects changes in amplitude and frequency in a time series. Here, we extracted *TE* from the broadband EEG signal (0.5-45Hz). In a continuous EEG time series $x(t)$, *TE* is computed as:

$$TE(t) = x^2(t) - x(t-1)x(t+1) \qquad (1)$$

The mean value of *TE* over each epoch is estimated as a feature of the data. The EEG line length (*LL*) is also a computationally simple feature that has been used before for seizure detection (Esteller et al., 2001). Similar to *TE*, *LL* grows as signal magnitude and frequency increases. *LL* of a signal (e.g. EEG time series $x(t)$) is estimated as:

$$LL(t) = \sum(|x(k) - x(k-1)|), \; k = t - N + 1, \dots, t \quad (2)$$

That is, *LL* is the total distance traversed by the signal from sample to sample over the epoch. Figure. 6.3 illustrates trends in *TE* and *LL* features for an almost 4-day long recording in a mouse. There are nine verified seizures during this recording that are noticeable as sharp spikes in both features. However, *LL* has smaller dynamic range in the baseline and seems to be less susceptible to other factors such as sleep and circadian rhythms. Because of the available contrast with the baseline, seizure detection can be accurately performed using EEG line length. This seizure detection algorithm has been applied to a handful of mice and showed 100% sensitivity and higher than 90% specificity.
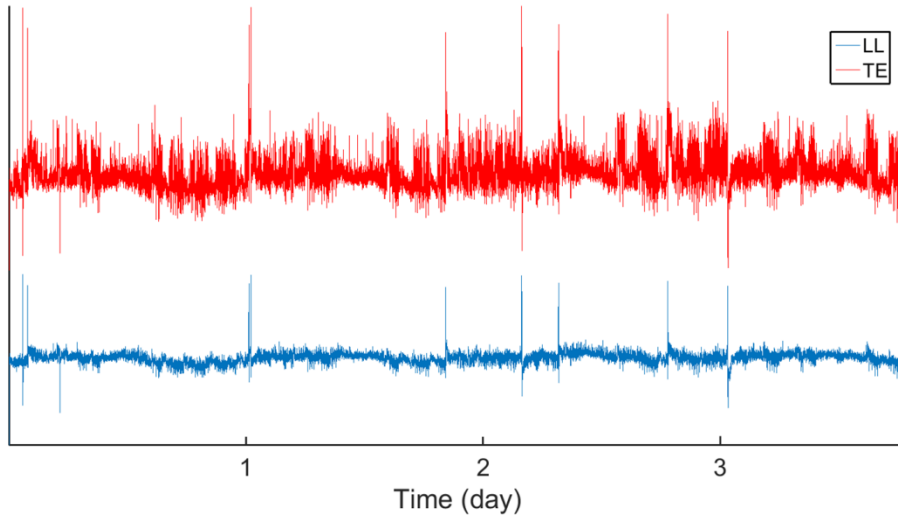
Figure 6.3 Seizure detection features from EEG. *TE* and *LL* show sharp spikes at seizures. 9 seizures are evident during the first 3 days.

Once seizures were detected accurately in recorded data, the temporal correlation between seizure events and sleep states can be determined by scoring the EEG/EMG signals. From Chapter 2, we know that determination of vigilance state in mice is feasible using unsupervised HMMs built on EEG/EMG features. Here, we utilize the same framework in epileptic mice. Discriminative features from EEG and EMG signals, i.e. EEG delta/theta power ratio and EMG power, were estimated in fixed epochs (4 s) and modeled using an HMM. Figure. 6.4 shows trends of EEG/EMG features as well as automated detection of vigilance state by an HMM. We also show how to apply a threshold to EEG line length and detect seizures. In Figure. 6.4, *LL* is normalized with respect to a smoothed version of it to correct for baseline variation (here a 3-minute long median filter was used). This step emphasizes the contrast between seizures and baseline in *LL*.
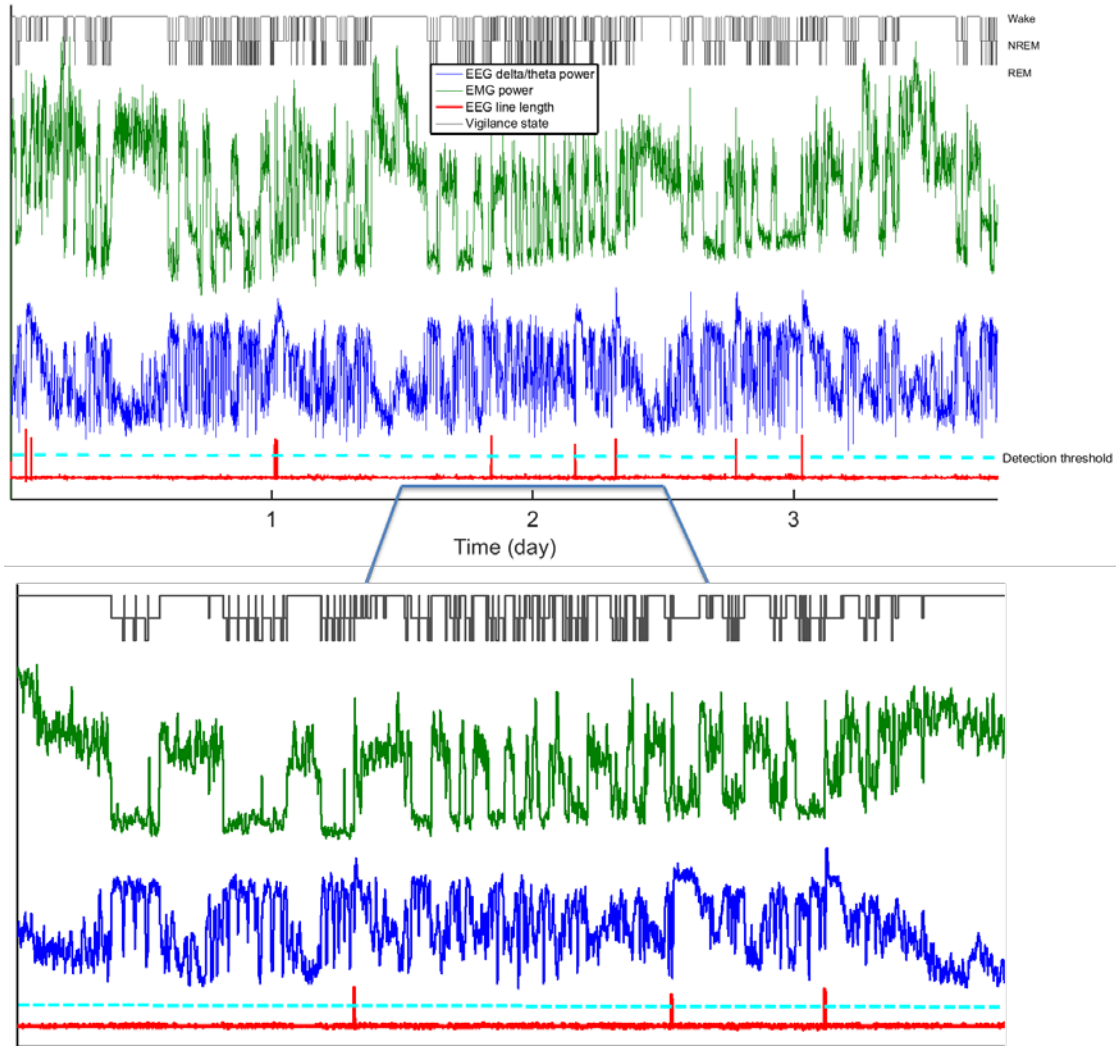
Figure 6.4 Sleep and seizure features from a TLE mouse. Vigilance state can be determined using an HMM fitted to EEG/EMG features. 9 seizures are also detected using EEG line length.

Once seizures and vigilance state are determined, we can study the passive correlation between them. For example, out of 9 seizure incidences seen in Figure. 6.4, two seizures happened when animal was awake and the rest happened when the animal was asleep (NREM). A similar analysis on prolonged recordings will tell us how seizures are clustered, when they usually happen, how sleep/wake cycles are modified following seizures, and how changes in vigilance state elevate or suppress seizure likelihood.

6.4. Noninvasive detection of seizures

A closer look at Figure. 6.2 reveals that convulsive seizures impose a characteristic pattern on the piezo signal that is different from the effect of other movements. It makes us wonder if the piezo signal can be used for detection of seizure onset. Noninvasive detection of seizures without the help of EEG/EMG measurements would enable us to continuously monitor an animal's behavior during the silent period and identify viable candidates with spontaneous seizures for EEG/EMG surgery and experimentation. In fact, the mortality rate of SE using pilocarpine is relatively large (~30-40%) and about 50% of surviving animals will develop spontaneous seizures (Shibley and Smith, 2002). Hence, on average one of three mice treated with pilocarpine will survive and become chronically epileptic. Given the cost and time spent for EEG/EMG surgical implantation, it is desirable to identify epileptic models during the silent period and then implant them with electrodes; rather than implanting them first with EEG/EMG electrodes and then injecting with pilocarpine and hoping for survival. Here, we investigate the feasibility of a similar detection algorithm to that described earlier but using noninvasive piezo measurements. Features reflecting seizure onset (TE and LL) are computed from the broadband piezo signal (0.5-20 Hz) in 4s epochs. Figure. 6.5 shows a feature from the piezo signal (TE) during the fourth week post-implantation with pilocarpine. The animal was not instrumented with the EEG/EMG headmount prior to pilocarpine injection. However, recording of the piezo signal along with video was started immediately after SE at the beginning of the silent period.
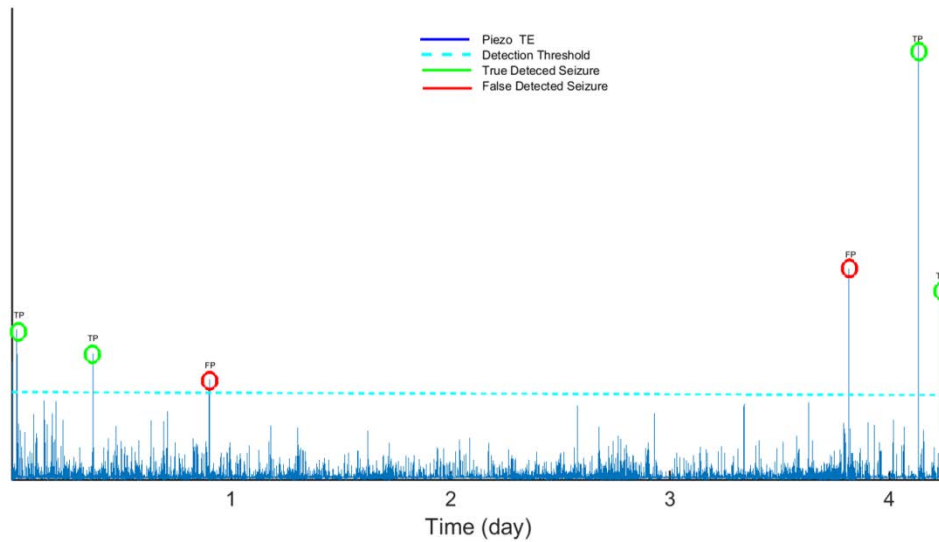
Figure 6.5 Noninvasive detection of seizures using piezo *TE*. 6 seizure candidates were detected with applying an arbitrary threshold and 4 seizures were verified using video.

A threshold-based detection was performed and candidate events were verified using the simultaneous video recording. Figure. 6.5 demonstrates that six candidate events were detected during a 4-5 day recording and after verification from video, four of them were identified as convulsive seizures and the other two were false detections (motion). Although the specificity of piezo signal features for seizure detection is not as high as EEG features, it is still a very useful tool to determine onset of spontaneous seizures following SE and identify animal candidates for EEG/EMG surgery.

In Chapter 3, we introduced an unsupervised technique for automated extraction of vigilance state from noninvasive piezo measurements. The same is applicable here to model vigilance dynamics in an epileptic animal without the need for EEG/EMG measurements. Noninvasive determination of vigilance state along with noninvasive detection of seizure onset would help us investigate their relationship either during the latent period prior to EEG/EMG implantation or in the event that the EEG/EMG signals lack the required quality. This endeavor can lead to the design of an effective tool to investigate behavioral indices of epileptogenesis even in large sample animal experiments. The same features presented in Chapter 3 were extracted here from the piezo signal and modeled using a 3-state HMM. Figure. 6.6 shows trends of two selected piezo feature along with HMM-decoded vigilance state. Seizures are also evident in the

Figure. A simple correlation analysis in a sample of reasonable size will tell us whether (and in what proportions) seizure occurrence is biased by vigilance state. Hence, noninvasive analysis of sleep and seizures could help us correlate seizure onset with vigilance state and explore the effect of seizures on sleep structure without the need for invasive EEG/EMG measurements.
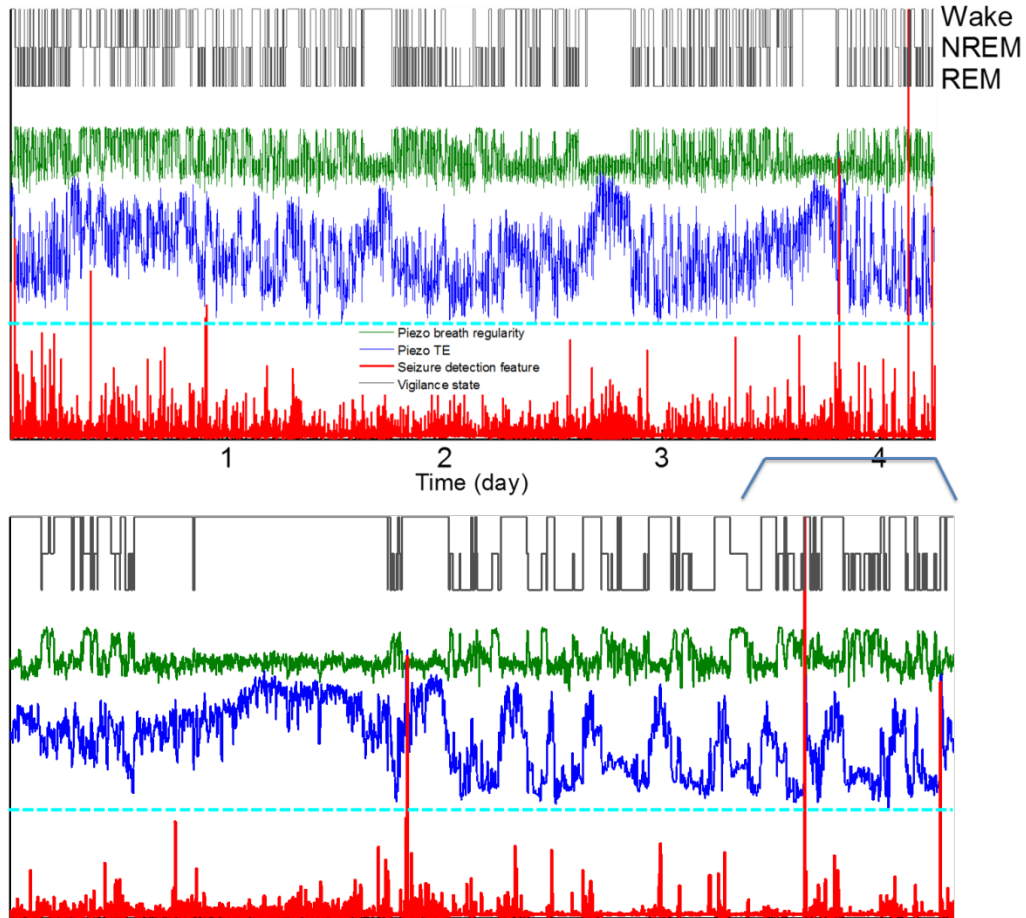
Figure 6.6 Piezo signal features and unsupervised detection of vigilance states using an HMM. Features reflecting motion and respiration were used to model vigilance dynamics.

6.5. Conclusions

This chapter of the dissertation was intended to examine the utility of the techniques previously introduced in an animal model of temporal lobe epilepsy. In fact, the

feasibility of such techniques was shown in a few simple applications; other applications along with statistical inference based on larger samples are deferred to future investigations. Understanding the relationship between seizures and any vigilance state in a baseline recording is critical to explaining how seizures are clustered and generated with respect to circadian or ultradian rhythms. The hidden Markov model -- and its usefulness as an efficient sleep classifier as proposed in Chapter 2 -- was successfully applied to EEG/EMG features from an epileptic mouse and a continuous hypnogram of sleep/wake states predicted without the availability of expert-scored training data (Figure. 6.4). Once we incorporated seizures into the analysis with the help of a simple but efficient seizure detection algorithm, the correspondence between seizures and vigilance can be studied. We have also shown that the correlation analysis can be done even without EEG/EMG signals to some extent. The noninvasive sleep scoring system proposed in Chapter 3 was applied here to extract vigilance states from the piezo signal recorded from a mouse treated with pilocarpine. This provided us with a reasonable partitioning of sleep-wake cycles along with fairly accurate seizure detection. This preliminary analysis showcases the potential of noninvasive piezo signals for detection of epileptic seizures in rodents.

The piezo system has never been utilized to detect seizures and successful implementation of this would lead to a completely noninvasive system for seizure monitoring that eliminates limitations of available noninvasive seizure detection systems (e.g. video monitoring). While the ability of HMMs for instantaneous prediction of vigilance state was shown in epileptic animals, continuous re-estimation of HMMs over a prolonged recording-- similarly to what was proposed in Chapter 4--can be performed to extract dynamical features related to sleep and epilepsy. This could serve as a useful noninvasive tool to track biomarkers of sleep or seizure dynamics particularly during epileptogenesis.

In addition to passive correlation between sleep states and seizures, sleep structure can be modulated actively to investigate the consequences on seizures. For example, the selective sleep restriction system proposed in Chapter 5 is one way to manipulate one variable (i.e., sleep) and study responses in another one (i.e., seizures). The introduced

system is flexible enough to modify sleep architecture in several controlled ways such as: reducing percent time or bout durations in different stages of sleep (e.g. REM or deep NREM) or inducing fragmentation (e.g. brief arousals). However, it is also useful to modulate sleep quality in an opposite direction; i.e. increasing sleep propensity or alleviating fragmentations or interruptions in sleep. We have shown that by regulating the ambient temperature using a thermostatic control system, it is possible to induce those effects on mouse sleep (Abbas et al. 2015). Thus, both sensory stimulation and thermoregulation systems are useful tools for modulating sleep quality in epilepsy models and investigating the consequences on seizures.

7. CHAPTER VII CONCLUSIONS

7.1.   Overview

Available treatments for epilepsy can have mild or severe side effects (Ortinski et al. 2004) such as poor sleep quality, which affects many epilepsy patients (Dingledine et al. 2007). On the other hand, it is well known that seizures disrupt normal sleep and sleep in turn influences seizure generation: 1) Light or transitional NREM sleep can facilitate seizures while REM sleep can inhibit them; and 2) Sleep deprivation can elevate brain excitability and thus precipitate seizures. A thorough understanding of sleep-seizure interactions using animal experiments would help scientists balance their reciprocal effects and improve the quality of life in patients with epilepsy. This dissertation focused on experimental analysis of sleep including novel computational techniques, empirical designs and tools for use in animal models and specifically mice. The ultimate goal of this endeavor is to develop a framework for programming seizure therapy without compromising sleep quality. This research offers a clear translational path toward epilepsy therapies such as automated neuromodulation for seizure control and programmed timing/dosage of medications. In this chapter a summary of the main findings of this dissertation is given. In terms of computational modeling, hidden Markov models were employed as unsupervised classifiers of sleep states and dynamics. Specific experiments were also designed to study sleep recovery following acute brain trauma and effects of sensory stimulation on sleep. The piezo signal as a novel measurement for scoring mouse sleep was also introduced and its potential for noninvasive analysis of sleep and behavior was evaluated.

7.2. Unsupervised scoring of mouse sleep using EEG/EMG measurements

Experimental and clinical evaluation of sleep and related disorders commonly involves visual scoring of usually prolonged EEG/EMG recordings which indeed is a tedious task. So, automated scoring of sleep using computers is highly desirable in sleep research and

several machine learning algorithms have been developed to implement this. In unsupervised classification, a model finds natural partitions in the input feature space without the need for human supervision. Sleep is a complex phenomenon and incorporates dynamics that can be modeled using Markov chains. Hence, Hidden Markov models are appropriate models for classifying sleep. We showed that unsupervised HMMs estimated from EEG/EMG features can accurately predict stages of sleep in mice (Yaghouby et al. 2012 &2013). On average ($n = 6$ mice), the sensitivity of our classifier was 93%, 96% and 88% for NREM, REM and Wake respectively. The specificity of the model was correspondingly high at 90%, 96% and 98% for NREM, REM and Wake respectively. The unsupervised model also gave accurate estimates of metrics of sleep dynamics such as the % time spent in each state. Successful implementation of this model in epileptic animals, similar to the example shown in Chapter 7, would greatly facilitate analysis of sleep structure and its correlation with epileptogenesis and epilepsy.

7.3. Noninvasive scoring of mouse sleep and behavior using a piezoelectric motion sensor

Many sources of variation in sleep architecture can be explained by heredity. Thus, investigation of genes that contribute to normal and abnormal sleep and wake behaviors would improve our knowledge of mechanisms and functions of the underlying vigilance states. Mice are considered the best genetic models for characterizing sleep through large-scale screening and behavioral experiments. However, screening still requires expensive and labor-intensive animal experimentation with EEG/EMG analysis. We proposed a noninvasive technique based on a piezoelectric motion sensor for scoring stages of sleep and behavior in mice. Measures of broadband activity and breathing regularity derived from the piezo signal showed potential for automated sleep scoring with reasonable accuracy when fitted with an unsupervised HMM (Yaghouby et al. 2011, 2012 and 2015b). The classifier distinguished Wake with high (89% sensitivity, 96% specificity) and REM with moderate (73% sensitivity, 75% specificity) accuracy, but NREM with poor sensitivity (51%) and high specificity (96%). The main source of error appeared to be the variability in breathing regularity associated with both REM and light NREM sleep. A supervised HMM classifier corrected the NREM-REM discrimination

problem and gave sensitivities of 90, 81, and 67% and all specificities above 90% for Wake, NREM, and REM respectively. Estimation of sleep metrics by the proposed model, i.e. % time spent in each state etc., was also comparable with human assessments. As we showed in Chapter 6, the piezo system can also be used to detect seizures and sleep states noninvasively. The findings of this research will open up new avenues for high throughput analysis of sleep or seizure phenotypes while alleviating the need for EEG recordings in small animals like rodents.

7.4. Unsupervised tracking of sleep dynamics during recovery from brain trauma

The ability to model and track brain dynamics from continuous physiological measurements will benefit investigations of related neural disorders such as brain injury and epilepsy. Hidden Markov models were shown to track the dynamics of sleep efficiently without the need for observer supervision (Yaghouby et al. 2012 and 2013). We applied HMMs for tracking sleep dynamics following acute brain injury in mice. Reestimation of HMM parameters over time provides us with useful metrics for sleep quality and dynamics. Such metrics can track progressive changes in behavior over time. The results suggested that HMMs estimated from both invasive and noninvasive signals could reveal peculiarities in sleep-wake dynamics. This approach can also be used to model and continuously track the effects of interventions such as sensory or electrical stimulation on mouse sleep regulation and dynamics.

7.5. Tactile sensory stimulation for selective sleep restriction in mice

Deeper understanding of sleep mechanisms and related disorders requires experimental techniques for manipulating sleep in animal models. A considerable amount of research has been done to propose sleep restriction systems in rodents. Each system comes with specific trade-offs such as implementation, adaptation, intrusiveness, stressfulness, and so on, that must be considered when designing an experiment to investigate a research question. Hence, novel techniques for experimental sleep manipulation are always of interest. To this end, we evaluated a promising technique for selective sleep restriction in mice that employs vibratory tactile stimulation triggered by automated detection of a

particular sleep state (in this case, REM). The effectiveness of the proposed technique in terms of sensitivity and specificity of real-time REM detection, the responsiveness of the animal to the stimulus and the chance of adaptation and getting desensitized to repeated stimulation were carefully evaluated and discussed (Yaghouby et al. 2014b&c). On average ($n = 4$ mice) the closed-loop sensory stimulation system consistently reduced % time spent and bout durations in REM. The real-time detection algorithm performed with high sensitivity (96%), moderate specificity (66%) and reasonable latency (7.4s) when compared with manual scores. Selective restriction of REM sleep in epilepsy models will improve our understanding of sleep-epilepsy interactions and its translatability across species. While REM sleep is known to prohibit seizures, its loss hasproportionally a greater effect on cortical excitability compared to NREM loss. On the other hand, REM sleep deprivation can occur as a consequence of a person's lifestyle: keeping up late hours and getting up early would eliminate majority of REM sleep, because the proportion of REM increases toward the end of sleep cycle. Repetition of such behaviors may cause a cumulative effect and promote seizures. Besides, understanding the effects of selective REM SD (totally or partially) may be useful for better adjustment of timing and dosage of seizure and sleep medications.

7.6. Other applications

The reciprocal model for sleep-seizure interactions (proposed in Chapter 1) was the base for different research questions addressed in this dissertation and to be explored in future work. One question was: "Can sleep quality be modulated to indirectly control seizure likelihood?" In Chapter 6 we proposed a sensory stimulation technique to alter sleep architecture and test the effect on seizure likelihood. A similar design, with more flexibility in vibration intensity, was also implemented and applied in a pilot study (not reported here) to restrict deep sleep in rats. The purpose of that study is to use mild tactile stimulation to prevent rats from awakening while reducing the amount of deep NREM sleep. Although this experiment is still at a preliminary stage, the effectiveness of closed-loop sensory stimulation for deep sleep restriction has been shown in rats (Huffman et al. 2015).

In general, the ability to regulate sleep quality can be considered as a non-

pharmacological therapy for sleep abnormalities. Dynamical modulation of sleep could also serve as a useful approach to control or reduce seizure likelihood in epilepsy. Besides sensory stimulation, regulation of ambient temperature can alter sleep dynamics. In a preliminary study we showed the ability to titrate REM and NREM sleep by controlling ambient temperature in mice (Abbas et al. 2015). By designing a thermostatic control system, we were able to adjust the ambient temperature of animal's cage at certain levels and study the consequences on sleep structure. The results showed that an elevation in temperature will change both REM and NREM proportions and bout durations in sleep, which can be a useful approach for investigating the reverse of the effect of tactile stimulation. Both sensory stimulation and temperature regulation offer low-intrusive techniques to mosulate sleep and investigate outcomes. I expect my research to have set the tone for further investigations whose results will have broad implications for the dose and timing of sleep and anti-seizure medication, and for the design of neuromodulation systems for individuals with epilepsy.

APPENDIX A

Naive Scoring of Human Sleep Based on a Hidden Markov Model of the Electroencephalogram*

Farid Yaghouby, Member IEEE-EMBS, Pradeep Modur and Sridhar Sunderam, Member IEEE-EMBS

ABSTRACT

Clinical sleep scoring involves tedious visual review of overnight polysomnograms by a human expert. Many attempts have been made to automate the process by training computer algorithms such as support vector machines and hidden Markov models (HMMs) to replicate human scoring. Such supervised classifiers are typically trained on scored data and then validated on scored out-of-sample data. Here we describe a methodology based on HMMs for scoring an overnight sleep recording without the benefit of a trained initial model. The number of states in the data is not known a priori and is optimized using a Bayes information criterion. When tested on a 22-subject database, this unsupervised classifier agreed well with human scores (mean of Cohen's kappa > 0.7). The HMM also outperformed other unsupervised classifiers (Gaussian mixture models, k-means, and linkage trees), that are capable of naive classification but do not model dynamics, by a significant margin ($p < 0.05$).

INTRODUCTION

Sleep quality is a critical determinant of human health and performance. Clinical evaluation of disordered sleep involves overnight polysomnography (PSG) following specific guidelines [1]. A PSG recording includes electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and other measurements, and is

scored by an expert in 30 s epochs into discrete vigilance states, namely wakefulness (Wake), rapid eye movement (REM) sleep, and non-REM (NREM, stages 1-3) sleep [2]. Scoring sleep is difficult and tedious. Many statistical classifiers have been developed to automate this process and replicate human performance [3], sometimes from a single EEG channel alone [4-5]; most require supervision in the form of expert heuristics or a statistical model derived from expert-scored training data to stage sleep; and all are used in essentially the same manner: i.e., by fitting a model to scored data from one set of subjects and validating it on out-of-sample data from another set [3-5]. This gives confidence that the model will work reliably on future subjects.

Supervised classifiers are constrained by the need for (and subjectivity/variability of) human scoring of training data. No method to date generates a reasonable first-pass hypnogram from a sleep recording without supervision: i.e., without previous training. Even hidden Markov models (HMMs), which, strictly speaking, are unsupervised classifiers, are first fitted to training data in which all vigilance states are known to occur, and then used to score test data [6-8]. But in the naive scenario, no initial model is available; nor may all vigilance states occur. Here, we propose a method for using HMMs to score overnight sleep without the benefit of a trained classifier. While supervised classifiers need labeled training data, unsupervised classifiers like the HMM find natural partitions in data that could map signal features onto distinct hidden states. In principle, PSG epochs can be mapped onto vigilance states without prior training-which a supervised classifier cannot do. This could yield a useful first-pass score for a new patient, to be refined by an expert if reasonably accurate.

Implicit in HMMs is the notion of dynamics, that the state follows a trajectory whose likelihood depends on the previous state at any instant. In contrast, most classifiers are "static", i.e., they do not incorporate context when determining state, unless subsequent steps filter classifier output: for instance, a minimum duration criterion, median filtering, exponential updating, and so on. Research on sleep dynamics suggests that human sleep is fairly well represented by a Markov chain model [9]. Since HMMs are built on Markov chains, this may explain their popularity in sleep scoring. However, other unsupervised but "static" classifiers (e.g., Gaussian mixture models or GMMs, k-means, k nearest neighbors, linkage trees, etc.) that cluster the feature space to score sleep from PSG

features have been investigated in the past [10-11]. Whether the assumption of Markov dynamics in HMMs truly translates into better predictive performance compared to other unsupervised static classifiers has not been verified. Here, we test a methodology for naive scoring of human sleep using HMMs. We also compare HMM performance with three unsupervised static classifiers to see if the added computational burden imposed by Markov dynamics is justified by classification performance.

METHODS

Signal features extracted from 30s epochs of overnight PSGs were modeled using four unsupervised classifiers: an HMM, a GMM, a k-means classifier, and a linkage tree. The number of states in each was optimized by an information criterion. Classification accuracy was assessed against expert-scored hypnograms.

Data source and feature extraction

This analysis is based on a Physionet database of 22 overnight expert-scored PSGs (6-9 h each; 100 Hz sampling) of healthy subjects (male/female, 18-79 years old, mean ~40) without medications [12-13]. All analysis was performed using MatlabTM (Mathworks, Natick, MA). The hypnograms, which mapped 30s epochs of data onto six states (NREM 1-4, REM, and Wake) were relabeled per the current guidelines of the American Academy of Sleep Medicine [2] by combining NREM stages 3 and 4. Hence, each hypnogram contained up to five labels: N1, N2, N3 for NREM, R for REM, and W for Wake. The Fpz-Cz signal from each subject was bandpass-filtered into seven distinct frequency bands, specifically: δL (0.5-2 Hz), δH (2-4Hz), $\boldsymbol{\theta}$ (4-9Hz), $\boldsymbol{\alpha}$ (9-12Hz), $\boldsymbol{\sigma}$ (12-16Hz), $\boldsymbol{\beta}$ (16-30Hz) and $\boldsymbol{\gamma}$ (30-45Hz) using 3rd order Butterworth IIR filters. The mean power in these bands was estimated in 30s epochs and combined into "sleep variable" ratios:

$$S1 = (\delta L + \delta H) / \boldsymbol{\theta} \qquad\qquad (1)$$

$$S2 = (\boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\sigma} + \boldsymbol{\gamma}) / (\delta L + \delta H + \boldsymbol{\theta}) \qquad\qquad (2)$$

$$S3 = \delta L / \boldsymbol{\sigma} \qquad\qquad (3)$$

Each variable is designed to emphasize contrast between EEG rhythms observed in different states of vigilance: S1 captures differences between N3 (strong delta) and R

(strong theta), S2 distinguishes N3 (low frequency) from W (broadband activity), and S3 discriminates N2 (spindle activity). This three-dimensional vector of features was expressed on a logarithmic scale, which makes the observation distribution approximately Gaussian, and used as the input to the unsupervised classifiers to be evaluated.

Modeling the data using unsupervised classifiers

The main aims of this analysis are: 1. To perform unsupervised sleep scoring using an HMM; and 2. To compare HMMs, which incorporate dynamics as Markov state transitions, with other unsupervised but static classifiers (GMMs, k-means clustering, and linkage trees) that do not have dynamics. In effect, GMMs and HMMs are parametric since they are based on a probability model, while k-means and linkage trees are nonparametric since they are based solely on proximity in the feature space.

Gaussian mixture models. A GMM expresses the distribution of $\mathbf{S} = [S_1 \ S_2 \ S_3]^T$ as a linear mixture of Gaussians: $p(\mathbf{S} \mid \Theta) = \sum \alpha_k \, p(\mathbf{S} \mid \theta_k)$. Each component k corresponds to one of ns model states, and θ_k is parameterized by a mean vector and covariance matrix; α_k is a mixing coefficient. Once ns is fixed, model parameters are determined from sample PSG data using maximum likelihood estimation. Assuming samples are independent and identically distributed, optimal parameters are those that maximize the function $L(\Theta \mid \mathbf{S}_{1:N}) = \prod p(\mathbf{S}_i \mid \Theta)$, which expresses the joint likelihood of all samples i = 1:N. L (or more commonly, log L) is optimized via an Expectation-Maximization (E-M) algorithm [14], in which an initial parameter guess is iteratively refined in a way that local convergence is guaranteed. For each subject, we used multiple randomized seeds and selected the solution with largest log L. Then, we labeled each epoch by the GMM component that maximized its probability density.

Hidden Markov models. An HMM is a dynamical model of a sequence or time series [15] that assumes each observation $\mathbf{S}_k$ in a sequence to be randomly drawn from a probability distribution conditioned on an underlying nominal state Q_k. $\mathbf{S}_k$ is conditionally independent of $\mathbf{S}_{k-1}$ given Q_k. The evolution of state Q_k over time follows the Markov property: i.e., given Q_k, the distribution of Q_{k+1} is independent of Q_{k-1}, Q_{k-2}, and so on

[16]. Here, we model the observation density p($\mathbf{S}$ | Q) as a Gaussian distribution where Q is one of ns discrete model states that relate to the different states of vigilance. To model a PSG recording using an HMM, its parameters must be fixed: namely, a set of priors π and emission models p($\mathbf{S}$ | Q), one for each of the ns states; and a matrix of transition probabilities P_{tr} between any two states. Algorithms are available for statistical inference using HMMs [16] that generally involves the recursive application of Bayes rule to compute the probability of a sequence of emissions from an arbitrary sequence of states, and for decoding the most likely sequence of states given an arbitrary sequence of emissions (the Viterbi algorithm). An E-M variant known as the Baum-Welch algorithm is used to estimate HMM parameters for a sample observation sequence $S_1, S_2,...,S_N$ [14]; since the source states are not known a priori, the HMM is an unsupervised model. Since we have chosen a Gaussian emission model for each state, we used the GMMs described in the previous section as the initial guesses of the priors and observation densities of the HMM. Once the model is determined, the Viterbi algorithm is used to decode the sequence of hidden states $Q_0, Q_1,...,Q_N$ most likely to have generated the sequence of emissions. As for GMMs, the likelihood L associated with the model can be computed for a sequence of observations.

k-means clustering. This is a well-known unsupervised algorithm, used here to cluster sample vectors of sleep variables into different states. The algorithm starts with k randomly selected prototypes or centroids (for k states), and then associates each data sample with a centroid based on the Euclidean distance between them in the feature space. The centroids are then recomputed based on the newly determined membership of each state. State labels and centroids are recursively updated until convergence [17].

Hierarchical clustering. A linkage tree is a clustering technique that builds a hierarchy of clusters using a "bottom up" approach. It starts with each observation forming its own cluster and then merges clusters based on their proximity to each other to move up the tree [18]. The tree therefore contains successively smaller numbers of clusters (states) at each level until there is only one cluster encompassing all the data at the top. The level at which the tree is "cut" or terminated determines the number of states, and their descendants on the tree inherit their labels.
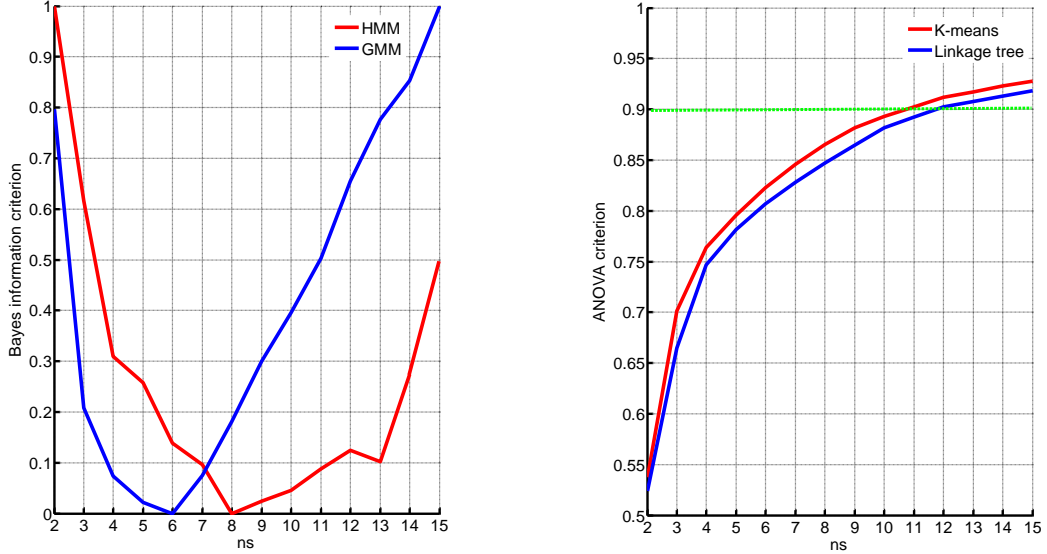
Figure 1. Criteria (shown for one sample subject) for selecting the number of model states ns that best fits the data. Left: Bayes information criterion (scaled by the dynamic range) passes through a minimum that determines ns for GMMs and HMMs. Right: Optimal ns for k-means and linkage tree classifiers is chosen as the lowest value for which an F-statistic representing relative variance between states exceeds 90%.

Optimization of the number of classifier states

For an unsupervised classifier, the number of model states *ns* must first be specified. Since the optimal number of states is not known a priori, a criterion is needed for the value of ns that best predicts the scatter observed in the data. While a large *ns* may give a better fit, the parameter space needs to be kept manageable and overfitting avoided. Also, ns should be close-but not necessarily equal-to the actual number *ms* of vigilance states in the sample: some model states may be sub-states of one vigilance state that together determine its distribution in the feature space.

For the parametric classifiers (GMM and HMM), we constructed models with *ns* varying from 2 to 15 Gaussian components. Then the optimal model was chosen by using the Bayes information criterion (BIC) [19], which balances conflicting terms representing the goodness-of-fit of the model and the degrees of freedom respectively:

$$BIC = -2 \log L + k \log n \qquad (4)$$

L is the likelihood of the data given the probability model, n is the number of observations (i.e., epochs of data), and *k* is the model degrees of freedom based on the

total number of fitted parameters in the model. Fig. 1a demonstrates how *BIC* varies with ns in a GMM fitted to data from an arbitrary subject (blue graph) whose recording contained all five vigilance states (*ms* = 5). A GMM with *ns* = 6 seems optimal for this subject. For an HMM of the same subject's data a choice of eight model states (*ns* = 8) is deemed optimal. The excess model states turn out to be subcomponents of vigilance states. For the nonparametric classifiers (k-means clustering and linkage trees) there is no probabilistic model, so a likelihood measure cannot be defined. Instead, we specify a criterion inspired by the F-statistic typically used in analysis of variance. We selected the optimal ns as the smallest value for which the ratio *R* of the variance between clusters to the total variance crossed 90%. For the sample subject in Fig. 1b, *R* monotonically increases with ns for the k-means algorithm and crosses 90% at *ns* = 11. Similarly, *ns* = 12 is optimal for a linkage tree classifier extracted from the same subject's data.
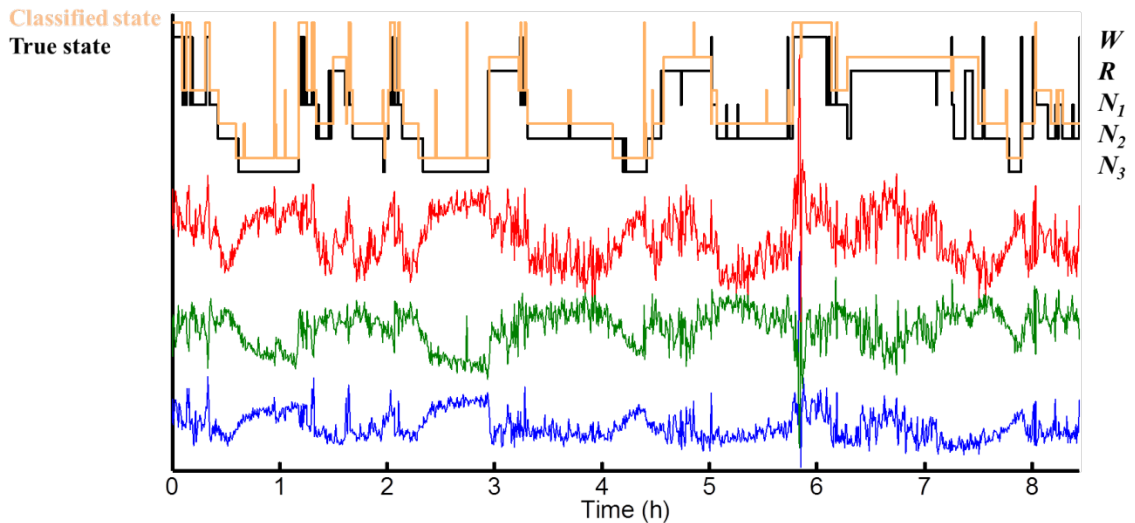
Figure 2. HMM classifier output for a sample overnight sleep recording. Input features S_1, S_2, and S_3 are shown below the model-generated (black) and true (beige) hypnograms for the data (Cohen's kappa = 0.8).

Mapping the model states to vigilance states

For each sleep record, dynamic (HMM) and static (GMM, k-means and linkage tree) unsupervised classifiers with ns optimized by *BIC* or *R* were constructed. The mapping

between model states and vigilance states is not known a priori. In fact, multiple model states may form sub-states of a particular vigilance state; and not all vigilance states may occur in a sleep record (e.g., subject never reaches N_3, or the recording does not include W). Whichever the case, we assume that a sleep physician could quickly inspect a few samples of each model state and fix the true vigilance state, based on which the hypnogram can easily be relabeled. In our analysis, we determine the mapping from model states to vigilance states by computing Cohen's kappa [20], which is a widely used statistical measure of inter-rater agreement. Since kappa takes chance agreement between the nominal states into account, it is a more reliable measure than just the overall proportion of agreement between labels. We applied the mapping that optimized Cohen's kappa for each subject before assessing the performance of each classifier.

 Assessment of classifier performance

Classifier performance was assessed by comparing model-predicted labels against true hypnogram labels using conventional metrics of detection sensitivity and specificity. The sensitivity (expected true positive rate) of a specific vigilance state reflects the proportion of actual sample epochs of that state correctly identified by the classifier. Conversely, the specificity (expected true negative rate) for a particular state is the proportion of other states not wrongly classified as the state of interest. Overall model performance was gauged by kappa while the ability to detect specific states was assessed using sensitivity and specificity.

RESULTS

Fig. 3 gives the performance of optimal static and dynamic classifiers on a 22-subject database in terms of Cohen's kappa. The static classifiers appeared to have similar performance with kappa of about 50%, which is considered moderate agreement with expert sleep scores. GMMs and linkage trees performed slightly but not significantly better than k-means. HMMs significantly outperformed the static classifiers ($p < 0.05$ by ANOVA), with a median kappa of over 70% (substantial agreement).

Trends in classifier performance in terms of sensitivity and specificity for each vigilance state (Table I) mirrored overall agreement (kappa), with some differences. Linkage trees and k-means gave very similar sensitivity and specificity for all five states. GMMs performed significantly better overall, except for lower sensitivity and higher

specificity to N_2, than the other static classifiers. HMMs gave comparable or significantly higher sensitivity and specificity for all states than any of the static classifiers.

Table I. Performance of unsupervised classifiers by vigilance state.

	Sensitivity (mean ± standard error)					Specificity (mean ± standard error)				
	N1	N2	N3	R	W	N1	N2	N3	R	W
K-means	27.8±3.4	81.3±2.4	61.3±5.6	52.8±4.7	32.9±3.8	93.3±1.2	79.1±2.3	93.9±0.9	91.6±1.1	93.6±0.9
Linkage tree	32.2±4.1	82.4±2.5	59.9±6.4	52.1±5.7	35.6±4.8	94.4±1	78.8±2.9	93.2±1.6	92.9±1	94.4±1.3
GMM	39.8±5.5	68.4±3.8	75.9±5.5	63.4±4.6	57.1±7.3	91.2±1.1	89.5±1.9	91.5±1.7	91.7±1.9	91.9±1.4
HMM	41.8±4.9	84.1±2.2	68.9±7.4	86.7±1.8	73.3±4.7	95.3±0.6	89.2±1.2	96.1±0.8	95.6±0.9	96.7±0.5

DISCUSSION

In this work, we compared HMMs with multiple static classifiers for clinical sleep scoring. The presumptive advantage gained by the empirical Markov chain representation of the dynamical sleep state transitions in the HMM has never been verified, but are now clear. Our other goal, to propose and test a means for obtaining reasonable initial sleep scores for an overnight recording without a previously trained model, also appears feasible. In this regard, we proposed a criterion for optimizing the number of states modeled by the classifier from the data without a priori information. This approach improved classification performance compared with similar studies [6-8], which are few in number and presume without justification that all stages of sleep are presented in each recording. Since the purpose of our HMM is to generate a first-pass segmentation, a human expert can quickly match up the model states with conventional vigilance states by reviewing a random sample of each model state. Moreover, our use of three simple power spectral features rather than a wide range of spectral /nonlinear EEG features [3-8] or auxiliary EMG/EOG features [8] results in a simple but more efficient automated sleep scoring technique. Use of the initial band power variables did not improve classifier performance despite the greater dimensionality.
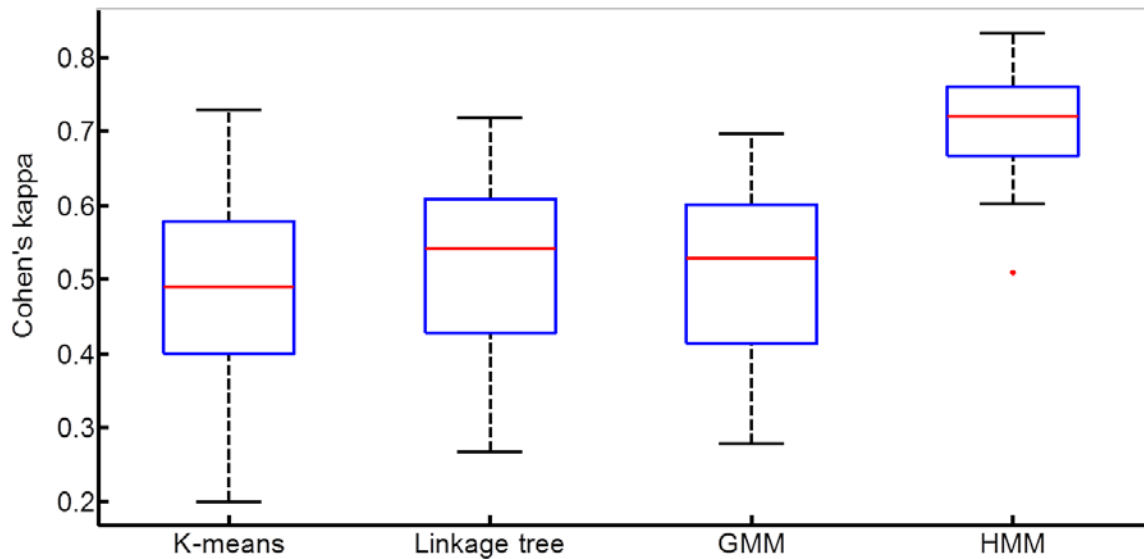
Figure 3. Overall performance of sleep classifiers assessed using Cohen's kappa (n = 22 subjects). HMM performance is significantly better than GMM, linkage tree and k-means classifiers (p < 0.05).

REFERENCES

[1]   A. Rechtschaffen , A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects", CA: BI/BR, Los Angeles, 1968.

[2]   C. Iber, S. Ancoli-Israel, A. Chesson, and S.F. Quan, "The AASM manual for the scoring of sleep and associated events", American Academy of Sleep Medicine, 2007.

[3]   S. Khalighi , T. Sousa, G. Pires and U. Nunes, "Automatic Sleep Staging: A Computer Assisted Approach for Optimal Combination of

Features and Polysomnographic Channels", Expert Syst Appl, vol. 40, pp.7046–7059, 2013.

[4]   B. Koley and D. Dey, " An ensemble system for automatic sleep stage classification using single channel EEG signal", Comput Biol Med, vol. 42, pp. 1186–1195, 2012.

[5]   L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier", Comput Meth Prog Bio,vol. 8, pp. 10-19, 2012.

[6]   A. Flexer, G. Grubera and G. Dorffner , "A reliable probabilistic sleep stager based on a single EEG signal," Artif Intell Med, vol. 33, 199-207, 2005.

[7]   L.G. Doroshenkov, V.A. Konyshev and S.V. Selishchev, "Classification of Human Sleep Stages Based on EEG Processing Using Hidden Markov Models", Biomedical Engineering, vol. 41, pp. 25-28, 2006.

[8]   S. Pan, C. Kuo, J. Zeng and S. Liang, "A transition-constrained discrete hidden Markov model for automatic sleep staging", BioMed Eng OnLine., vol. 11, 2012.

[9]   J.W. Kim, J.S. Lee, P. A. Robinson, and D.U. Jeong, "Markov Analysis of Sleep Dynamics", Phys Rev Lett, vol .102, pp.178104, 2009.

[10]   I. Gath, C. Feuerstein, and A. Geva. "Unsupervised classification and adaptive definition of sleep patterns". Pattern Recognition Lett, vol. 15, 977-984, 1994.

[11]   H. Escola, E. Poiseau, M. Jobert, and P. Gaillard. "Classification using distance-based segmentation application to the analysis of EEG signals", Pattern Recognition Lett, vol. 12, 327-333, 1991.

[12]   A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark , J Mietus, G. Moody , C. Peng, H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals", Circulation, vol .101, pp. 215-220, 2000.

[13]   B. Kemp, A.H. Zwinderman, B. Tuk, H.A.C. Kamphuisen, J.J.L. Oberyé, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG", IEEE-BME vol. 47, 1185-1194, 2000.

[14]   JA. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models" International Computer Science Institute, 1998.

[15]   A.   Krough, "An introduction to hidden Markov models for biological sequences", Computational Methods in Molecular Biology, 1998.

[16]   L.R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition", Proc. IEEE, vol. 77, 257-286, 1989.

[17]   J. B. McQueen, "Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1. University of California Press. pp. 281–297, 1967.

[18]   G. J. Székely and M. L. Rizzo, "Hierarchical clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method", J Classif, vol. 22, pp. 151-183, 2005.

[19]   D. Posada and T.R. Buckley, "Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests", Syst Biol, vol. 53, pp. 793-808, 2004.

[20]   J. Cohen, "A coefficient of agreement for nominal scales". Educ Psychol Meas, vol. 20, pp. 37-46, 1960.

APPENDIX B

Validation of a Closed-Loop Sensory Stimulation Technique for Selective Sleep Restriction in Mice*

Farid Yaghouby, Member, IEEE-EMBS, Christopher J. Schildt, Kevin D. Donohue, Senior Member, IEEE, Bruce F. O'Hara, and Sridhar Sunderam, Member, IEEE-EMBS

ABSTRACT

Experimental manipulation of sleep in rodents is an important tool for analyzing the mechanisms of sleep and related disorders in humans. Sleep restriction systems have relied in the past on manual sensory stimulation and recently on more sophisticated automated means of delivering the same. The ability to monitor and track behavior through the electroencephalogram (EEG) and other modalities provides the opportunity to implement more selective sleep restriction that is targeted at particular stages of sleep with flexible control over their amount, duration, and timing. In this paper we characterize the performance of a novel tactile stimulation system operating in closed-loop to interrupt rapid eye movement (REM) sleep in mice when it is detected in real time from the EEG. Acute experiments in four wild-type mice over six hours showed that a reduction of over 50% of REM sleep was feasible without affecting non-REM (NREM) sleep. The animals remained responsive to the stimulus over the six hour duration of the experiment.

INTRODUCTION

Since circadian and homeostatic modulation of sleep is similar across different mammalian species, animal models may be useful in unraveling the mechanisms of sleep

in humans. The use of animal models, particularly rodents, in sleep research provides scientists with the opportunity to investigate the genetic and neurobiological changes underlying sleep abnormalities.

Total sleep deprivation (TSD) has long been used for investigating sleep regulation mechanisms and the effects of sleep loss. The main effects of TSD on rodent sleep are a prominent increase in electroencephalogram (EEG) slow wave activity (SWA; 0.75-4Hz oscillations) during non-rapid eye movement (NREM) sleep as well as theta activity (6-9 Hz oscillations) in rapid eye movement (REM) sleep [1]. However, TSD is not an adequate model for interrupted sleep in humans, which does not always feature total sleep loss. Hence, selective sleep deprivation has been investigated in animal models and humans to evaluate its effects on physiology.

Selective sleep deprivation protocols permit comparison of the roles of each state of sleep on behavior and cognition. Deprivation of NREM (REM) sleep is followed by NREM (REM) rebound during the recovery period [2]. However, deprivation of either state may affect the other one as well. REM sleep deprivation (REM SD) is not 100% selective and results in REM rebound and extensive suppression of SWA in NREM [2]. Manual experimental methods have been employed to target particular stages of sleep that may involve gentle handling [3] or cage movement [4]. To avoid the need for human supervision and intervention, other methods have been devised, especially for REM SD: for instance, the "flower pot" or "multiple platform" method, in which the animal is placed on a platform and falls into a basin when it becomes atonic during REM sleep. This method essentially deprives the animal of all REM sleep and is highly stressful [5].

Programmable computer-controlled devices have been employed to detect sleep state automatically from the EEG and rouse the animal using some form of stimulation, such as cage shaking [6,7], a rotating disk over water [8] or a slowly rotating stir bar on the cage floor [9]. Automated sleep restriction may provide greater flexibility and selectivity compared to previous techniques, and perhaps limit the confounding effects of hormonal stress. However, each manipulation technique has advantages and limitations related to convenience, intrusiveness, cost, and efficacy. Here, we test the feasibility of a novel sleep restriction technique in mice. The basis of this technique is to detect the onset of a targeted

sleep state (REM or NREM) from EEG signals using a computer algorithm and apply tactile stimulation in the form of vibrations transmitted through the cage floor to rouse the animal. Using this system, we can implement selective sleep restriction relatively easily and with greater flexibility over the proportion of sleep loss. The application of our technique for selective REM SD in mice is described in this paper.

METHODS

Animals, care and protocols

All experimental procedures in this study were conducted with the approval of the Institutional Animal Care and Use Committee (IACUC) at the University of Kentucky. The experiments were performed on adult male wild type mice (C57BL/6J, Jackson Labs), the most widely used inbred strain (4-6 weeks old, $n = 4$). Each animal was housed independently with 14h/10h light/dark (ambient temperature: $20 \pm 2℃$, humidity: $50 \pm 10\%$) and free access to food and water at all times. A baseline EEG recording of 6 h duration (1-7 P.M.) during subjective night was acquired from mice and used to tune an automated REM sleep detector. An experimental session was performed at a later date, at the same time of day and for the same duration, in which tactile stimulation was applied whenever REM sleep onset was detected from the EEG.

Surgical implantation and signal acquisition

Electrodes for monitoring brain and muscle activity were implanted under 2.5% isoflurane anesthesia. A head-mounted preamplifier (8201; Pinnacle Technology, Inc, Lawrence, KS) was affixed directly over bregma using four miniature silver screws that serve as two differential cortical EEG derivations with a common reference and ground. Teflon-coated leads were inserted bilaterally into the dorsal neck muscle posterior to the skull to provide an electromyogram (EMG). Then, the animals were allowed to recover and adapt for two weeks before collecting data.

Our chronic acquisition system includes tethered EEG/EMG (8206; Pinnacle Tech., Lawrence, KS), with a USB camera (Microsoft LifeCam VX-6000) and infrared (IR) illumination source to enable continuous video recording across light and dark periods. Input signals were digitized at 14 bits and a sampling rate of 400 Hz under software

control (Sirenia[TM], Pinnacle Tech.). A custom LabVIEW[TM] interface (National Instruments) captured video in synchrony with EEG/EMG acquisition.
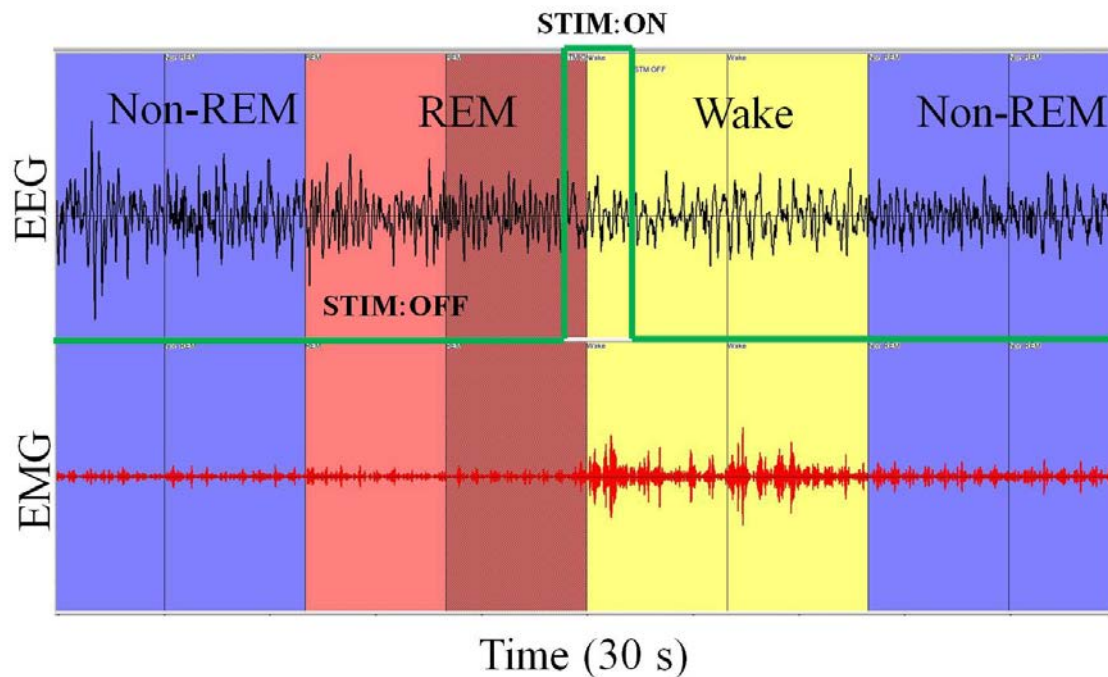
Figure 1. Snapshot of EEG and EMG during a NREM-REM state transition during sleep interrupted by a brief tactile stimulus. The animal is aroused briefly as seen from the elevated EMG amplitude, and then drifts back into NREM sleep.

EEG/EMG signals were analyzed in real time using LabVIEW to detect REM sleep onset and trigger stimulation for REM SD. A baseline recording was first processed offline to tune the REM classifier. This was then used to detect REM sleep onset in real time and trigger the stimulation system in closed-loop. The trigger signal was recorded synchronously with the EEG and EMG as an additional data stream.

Sleep scoring and REM detector training

Training data were selected from a baseline recording in each mouse (after allowing it to adapt to the recording cage for two days) to determine transition to REM sleep based on EEG/EMG features. The vigilance state was manually scored based on well-established criteria using a video-EEG viewer in 4 s epochs as Wake, REM and NREM. Wake is identified by low amplitude, desynchronized EEG and relatively high amplitude EMG. NREM stages have increasingly prominent delta (δ: 0.5-4 Hz) while

93

REM has a prominent theta oscillation (θ: 6-9 Hz) similar to activity during Wake; EMG amplitude is low in both REM and NREM. Hence, spectral band power estimates from EEG and EMG were used as features to construct an automated REM sleep detector for each animal. The mean power from band-pass filtered EMG (80-100 Hz) was used to detect low muscle tone in sleep. Within sleep, the δ/θ band power ratio was estimated to detect the onset of REM sleep. Thus, REM detection thresholds were established for both features using the baseline recording and manual scores. These feature thresholds were used to detect REM sleep onset during the REM SD experiment.

Real time REM sleep detection and stimulation

REM SD experiments were performed using a closed-loop system that applies a vibratory tactile stimulus to the animal when REM sleep is detected. Eight equally spaced button-type shaftless vibration motors (No. 1638, Pololu Corporation, Las Vegas, NV, USA) are attached to the underside of a rubber pad on the floor of the animal's cage. Each motor (10 mm diameter, 2 mm thickness) vibrates with an amplitude of 0.75 g at 12,000 r.p.m. when driven by a 3 V DC supply. The vibration is transmitted to the animal's body via the pad and produces tactile stimulation. A LabVIEW program calculates frequency band power features from EEG and EMG in 1 s epochs (4 s moving average) and activates the stimulation when preset thresholds on the EEG/EMG features are crossed indicating that REM sleep onset has been detected. The stimulation is automatically stopped when the state has changed and the animal is awake (see Fig. 1). Using this technique we were able to selectively reduce the proportion of REM sleep dramatically without affecting NREM sleep. The performance of the system in detecting REM sleep in real time and reducing its proportion was verified against manual scoring of the data from the experiment.

RESULTS

First, we evaluate the performance of the real-time REM sleep detector for each animal. Then we assess the effects of closed-loop sensory stimulation on REM and NREM sleep.

Assessment of real-time REM sleep detection

A human scorer inspected EEG, EMG and video data for each mouse during the experimental stimulation phase and determined vigilance state in sequential 4 s epochs. The state of the stimulation trigger was not visible to the scorer. Stimulation onset and offset times were also extracted from the recordings. To assess the performance of real-time REM detection in each animal, true REM incidents (bouts of continuous REM) as determined by visual scoring were compared against stimulation times. Numbers of the following events were determined for each recording: 1. True Positive (TP) detections, i.e., REM bouts that overlapped with stimulation onset; 2. True Negative (TN) events, i.e., NREM or Wake bouts (i.e., other than REM) in which stimulation was already off or switched off; and 3. False Positive (FP) detections, i.e., NREM or Wake epochs in which stimulation was activated or already on. These counts were combined into two commonly used performance measures:

$$\text{Sensitivity} = TP / (TP + FN) \qquad\qquad (1)$$
$$\text{Positive predictive value } (PPV) = TP / (TP + FP) \qquad (2)$$

In addition, depending on the choice of threshold or time constants related to filtering and the 4 s resolution of manual scoring, there can be a finite delay from the true onset of REM to when the detector is triggered. The REM detection latency was estimated for each TP detection as the time between REM onset and stimulation onset. A summary of detector metrics is presented in Table I.
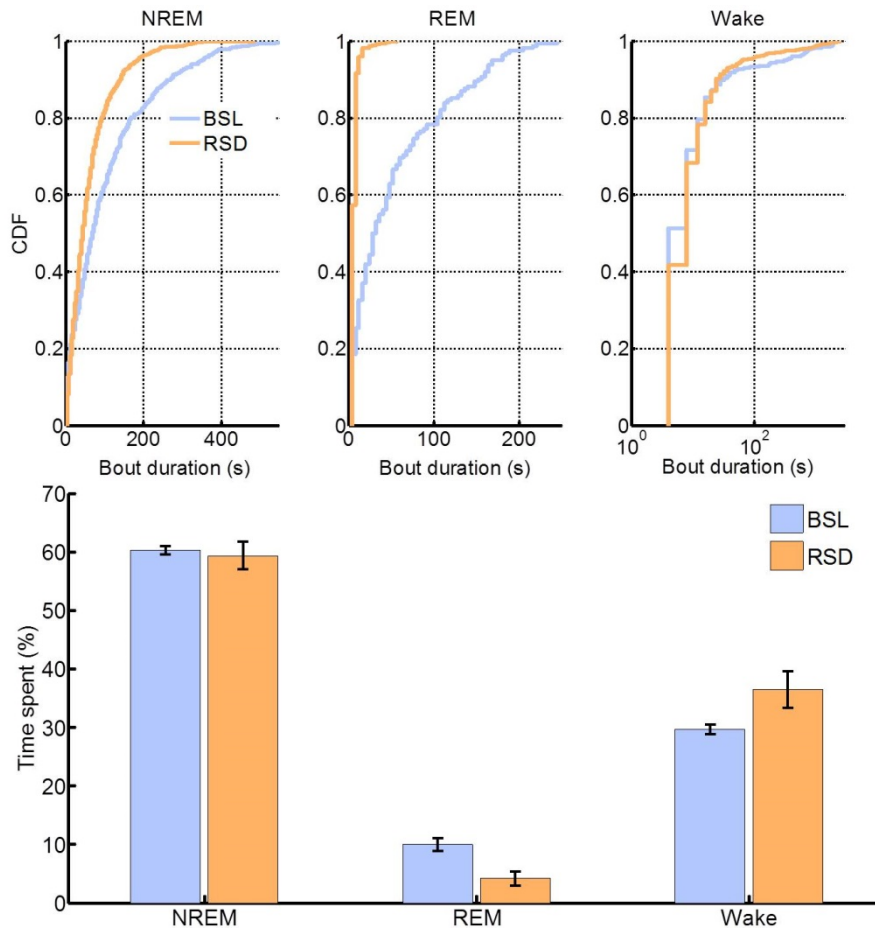
Figure 2. Effects of closed-loop sensory stimulation on sleep. Upper: Cumulative distribution function (CDF) of bout durations in each state (data pooled from all four mice). Lower: Comparison of mean percent time spent in each state for baseline and stimulation periods. Error bars represent standard error of the mean ($n = 4$).

Performance of REM sleep restriction system

To study the effectiveness of closed-loop REM SD system, sleep parameters were estimated and compared for the baseline and experimental REM SD recordings. Visual scores from each animal in baseline and REM SD phases were used to estimate percent time spent in REM, NREM, and Wake as well as the distribution of bout duration for each state. In Fig. 2 (Upper) we plot the cumulative distribution of bout duration for each state of vigilance (data from all animals are pooled together). It shows that stimulation drastically reduced REM bout duration while its effect on NREM and Wake is relatively

small. In Fig. 2 (Lower) the mean % time spent in each state ($n = 4$ mice) is shown for the baseline and stimulation phases. The % time in NREM is almost intact while % REM is clearly reduced and % Wake increased. The amount of REM loss appears to be compensated by a gain in Wake.

DISCUSSION

Sleep is a delicate and complex dynamical process. Disruption of sleep due to stress, injury, medication, disease, lifestyle, and environmental factors can have serious health consequences. Experimental manipulation of sleep can help us understand how sleep and health are inter-related and discover new treatments for sleep-related disorders. Many methods have been proposed for sleep restriction in rodents, and each one comes with unique trade-offs—ease of implementation, flexibility, stressfulness, intrusiveness, efficacy, arousal threshold, adaptation over time—that must be considered in selecting one that is appropriate for the research question under investigation. New additions to the arsenal of techniques for sleep manipulation are therefore always welcome.

In this paper a promising new technique for selective sleep restriction in mice was evaluated that employs vibratory tactile stimulation triggered by automated detection of a particular phase of sleep (in this case, REM). As with any closed-loop sleep restriction method, the effectiveness of this technique relies on how sensitive and selective the detector is to REM sleep, the responsiveness of the animal to the stimulus, and whether the animal is likely to get desensitized to the stimulus with repeated exposure to it over the course of the experiment. We consider these factors below in light of our experimental results.

The ability of our algorithm to detect REM sleep in real time is summarized in Table I. The sensitivity of the detector to REM sleep onset was high, over 90 %, in 3 of 4 animals. But this is balanced by a much more moderate specificity (PPV), which ranged from 38 to 73 % in the same animals. (The situation is reversed in Animal 1, which had relatively poor detection sensitivity but high specificity.) This means that roughly one-third to one-half of all stimulations occurred during a state other than REM (NREM or Wake). Since tactile stimulation does not change the animal's state when awake, the

slight reduction in mean NREM bout duration relative to baseline (Fig. 2 Upper) could have been due to these false REM detections.

Table I. Evaluation of REM detection performance.

Animal	Sensitivity (%) [Total REM bouts]	PPV (%) [Total detections]	Latency (s) (mean ± s.d.)
1	72.4 [254]	88.0 [209]	8.2 ± 3.8
2	93.5 [77]	38.3 [188]	7.9 ± 8.4
3	98.3 [114]	65.1 [172]	6.7 ± 4.8
4	95.7 [117]	72.7 [154]	6.7 ± 3.3
Mean ± s.e.m.	**95.8 ± 5.9**	**66.0 ± 10.4**	**7.4 ± 0.4**

Stimulation during NREM appeared to induce only brief arousal, which may explain why the fragmentation of NREM sleep is not accompanied by a significant reduction in % NREM sleep in Fig. 2 (Lower). Another possible reason for the shorter NREM bouts could be the brain's homeostatic tendency to try to recover lost REM sleep as the protocol is continued. Metrics for Wake show a lower bout duration, which may be due to increased brief arousals during NREM triggered by FP stimulation, but a greater % Wake, which matches the reduction in % REM closely. The other detection metric in Table I is latency. On average it takes about 7 s (two 4 s epochs) for the detector to determine that the animal is in REM sleep and then deliver a stimulation pulse. A possible reason for this delay is the way data was scored. Brief or transitional episodes of REM are manually scored as REM while the detector may wait for a more distinctive signature of REM sleep based on EEG/EMG features and the preset thresholds on them. As a consequence, the protocol only affects prolonged REM bouts and ignores brief episodes. The reduction in median REM bout duration (Fig. 2 Upper) compared to baseline (from 28 s to 4 s) supports this observation.

Taken together, Table I and Fig. 2 indicate that the stimulation protocol produced a reduction in REM sleep of over 50 % on average over a 6 h period. That REM sleep is not eliminated altogether may be attributed in part to the latency of detection. In this preliminary study, we have implemented a simple linear thresholding approach for REM onset detection that is very sensitive to EEG/EMG signal quality, which was poor in

Animal 2 and adversely affected detection specificity (PPV). Taking advantage of better supervised machine learning techniques, such as support vector machines or hidden Markov models, could improve the performance of online REM detection and thereby the efficacy of REM sleep restriction.
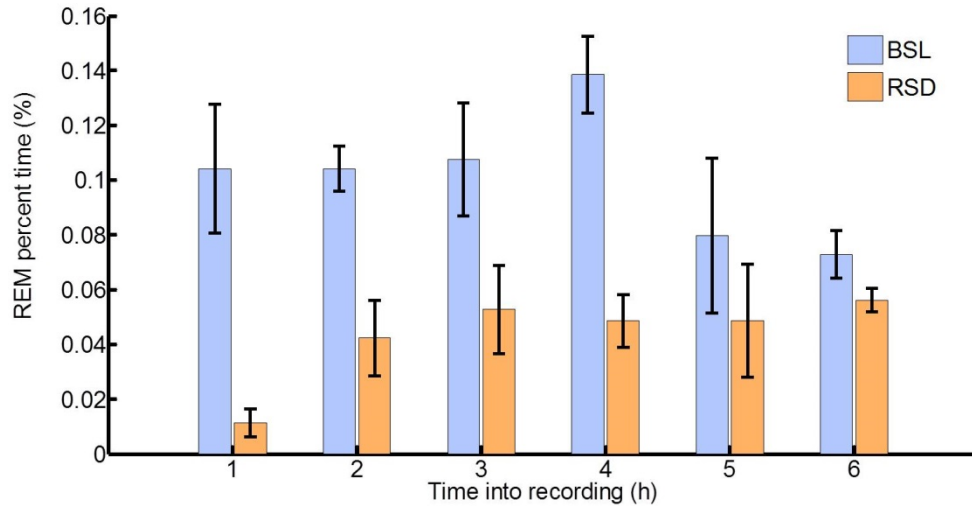
Figure 3. Trends in mean hourly percent time spent in REM during the baseline (BSL) and REM SD (RSD) stimulation protocol.

The effectiveness of this system was examined for selective disruption of REM sleep in mice in acute experiments 6 h in duration. Besides the limitations on performance of the detector, the animal could become desensitized and eventually oblivious to the stimulus depending on the frequency and duration over which it is applied. Fig. 3 presents the mean % time spent in REM for each consecutive hour of monitoring during the baseline and RSD protocols. The % time in REM is lowest (1 %) in the first hour and higher but relatively constant (4-6 %) over the next five hours. By comparison the baseline value starts at 10 %, peaks in the mid-afternoon (14 %), and starts to fall as the evening approaches; this is consistent with diurnal variation. It is logical that the achievable reduction in REM later in this period is low simply because the probability of its occurrence is already low under baseline conditions. In conclusion, the efficacy of the stimulation protocol appears to be relatively stable except in the first hour since the animal is naive to the stimulus at that time. This suggests that, in the acute experiments performed, the animal remains responsive to the stimulus without a change

in threshold. However, it remains to be seen whether the effect will persist with more frequent application over a longer monitoring period. More experimentation is under way to better characterize the performance and limitations of this new system for chronic selective sleep restriction in mice.

REFERENCES

[1] B. Schwierin, A.A. Borbély, I. Tobler, "Prolonged effects of 24-h total sleep deprivation on sleep and sleep EEG in the rat," *Neurosci Lett,* vol.261, pp. 61–64, 1999.

[2] T. Endo, B. Schewierin, A. Borbely, I. Tobler, "Selective and total sleep deprivation: effect on the sleep EEG in the rat," *Psychiatry Res*, vol. 66, pp. 97-119, 1997.

[3] A. Ocampo-Garcés, E. Molina, A. Rodríguez, E.A. Vivaldi, "Homeostasis of REM Sleep After Total and Selective Sleep Deprivation in the Rat," *J Neurophysiol*, vol. 84, pp. 2699-2702, 2000.

[4] S. Datta, F. Desarnaud, "Protein Kinase A in the Pedunculopontine Tegmental Nucleus of Rat Contributes to Regulation of Rapid Eye Movement Sleep," *J Neurosci*, v. 30, pp. 12263–12273, 2010.

[5] W.B. Mendelson, R.D. Guthrie, G. Frederick, R.J. Wyatt, "The flower pot technique of rapid eye movement (REM) sleep deprivation," *Pharmacol Biochem Behav*, vol. 2, pp. 553-556, 1974.

[6] J.G. McCoy and R.E. Strecker, "The cognitive cost of sleep lost," *Neurobiol Lear Mem*, vol. 96, pp. 564-582, 2011.

[7] S. Sahu, H. Kauser, K. Ray, K. Kishore, S. Kumar, U. Panjwani, " Caffeine and modafinil promote adult neuronal cell proliferation during 48 h of total sleep deprivation in rat dentate gyrus," *Exp Neurol*, vol. 248, pp. 470-481, 2013.

[8] C.A. Kushida, B.M. Bergmann and A. Rechtschaffen, "Sleep deprivation in the rat: IV. Paradoxical sleep deprivation," *Sleep*, vol. 12, pp. 22-30, 1989.

[9] J.P. Wisor, W.C. Clegern, M.A. Schmidt, "Toll-Like Receptor 4 is a Regulator of Monocyte and electroencephalographic Responses to Sleep Loss," *Sleep*, vol. 34, pp. 1335-1345, 2011.

APPENDIX C

Quasi-Supervised Scoring of Human Sleep in Polysomnograms Using Augmented Input Variables*

Farid Yaghouby and Sridhar Sunderam

ABSTRACT

The limitations of manual sleep scoring make computerized methods highly desirable. Scoring errors can arise from human rater uncertainty or inter-rater variability. Sleep scoring algorithms either come as supervised classifiers that need scored samples of each state to be trained, or as unsupervised classifiers that use heuristics or structural clues in unscored data to define states. We propose a quasi-supervised classifier that models observations in an unsupervised manner but mimics a human rater wherever training scores are available. EEG, EMG, and EOG features were extracted in 30s epochs from human-scored polysomnograms recorded from 42 healthy human subjects (18 to 79 years) and archived in an anonymized, publicly accessible database. Hypnograms were modified so that: 1. Some states are scored but not others; 2. Samples of all states are scored but not for transitional epochs; and 3. Two raters with 67% agreement are simulated. A framework for quasi-supervised classification was devised in which unsupervised statistical models—specifically Gaussian mixtures and hidden Markov models—are estimated from unlabeled training data, but the training samples are augmented with variables whose values depend on available scores. Classifiers were fitted to signal features incorporating partial scores, and used to predict scores for complete recordings. Performance was assessed using Cohen's K statistic. The quasi-supervised classifier performed significantly better than an unsupervised model and sometimes as well as a completely supervised model despite receiving only partial scores.

The quasi-supervised algorithm addresses the need for classifiers that mimic scoring patterns of human raters while compensating for their limitations.

INTRODUCTION

Sleep is increasingly the subject of debate in the context of public health [1, 2]. Disorders of sleep [3] are not only unique in the spectrum of illnesses but also accompany and complicate the management of other serious neurological conditions such as epilepsy [4], Parkinson's [5] and Alzheimer's disease [6]. Human sleep has been dissected broadly into five distinct states of vigilance: Wakefulness (*W*), rapid eye movement or REM sleep (*R*), and non-REM sleep (*N*) with stages *N1*, *N2*, and *N3* that reflect increasing sleep depth. Sleep analysis typically involves overnight monitoring in a sleep lab resulting in a polysomnogram: i.e., a suite of continuous measurements that may include an electroencephalogram (EEG), electromyogram (EMG), electrooculogram (EOG), and electrocardiogram (EKG), among other physiologically derived signals. The polysomnogram is inspected by a human expert, who labels the predominant vigilance state in sequential epochs, each typically 30s in duration, for the entire recording. Despite the adoption of detailed guidelines [7] for labeling each vigilance state by practitioners of sleep medicine, and continuing efforts to automate the process, scoring sleep in polysomnographic recordings remains a tedious and subjective exercise. Even expert raters can be uncertain about the presentation of certain vigilance states and may vary widely in their assessment of specific recordings [8].

Computational tools that segment sleep either look for intrinsic patterns in the data [9-11] to define the predominant vigilance states or model a human rater's scoring of sample data and try to mimic her performance when applied to future recordings [12, 13]. These contrasting approaches, referred to as unsupervised and supervised classification respectively, are mutually exclusive; moreover, they do not explicitly address issues of rater uncertainty and disagreement. Here we propose a simple modification to the way classifiers are applied to sleep data to address three specific scenarios:

1. A human rater is more certain about the symptoms of some vigilance states than others;

2. A rater labels all the states, but only in samples where the evidence is unambiguous; and

3. One classifier needs to mimic a panel of raters with some variance in their scoring patterns.

In our algorithmic solution to these distinct but related problems, a set of features computed from each epoch of the polysomnogram is augmented, or tagged, with a vector variable whose value depends on the available score(s). This sequence of score-augmented input variables is used to train an unsupervised classifier—Gaussian mixture models (GMMs [14]) and hidden Markov models (HMMs [15]) are used here as illustrative examples—to map the continuous-valued features onto discrete vigilance states. Minor variations on this theme are used to address each of the scoring scenarios identified above and the performance of the classifier compared with appropriate reference methods.

METHODS

Overview

Descriptive features were extracted from sequential signal epochs of overnight polysomnograms derived from an online database. For each recording, the hypnogram—i.e., the sequence of vigilance state labels assigned by a human rater—was systematically modified to simulate situations in which the rater was uncertain about the identity of certain states or epochs. The vector time series of features was fitted to two different statistical classifiers, a GMM and an HMM, using a novel quasi-supervised algorithm and used to predict the sequence of true vigilance states. The predictions were compared against the hypnogram to assess the ability of the proposed algorithm to compensate for missing or imprecise scores, and tested on a second night's recording from each subject when available. The performance of fully supervised and unsupervised classifiers on the same data were also assessed as reference cases.

Description of human subject data

This analysis is based on the Sleep EDF database [16] (available from www.physionet.org [17]). The database has a total of 61 overnight expert-scored PSG recordings from healthy individuals acquired with institutional oversight and informed consent. The data were collected from two different studies: 1. Sleep cassette (SC), which includes two successive overnight in-home recordings (except in one case) from 20 subjects (10 male and 10 female, 25-34 years old) without any medications; and 2. Sleep telemetry (ST), in which PSGs were recorded in-hospital, from 22 healthy subjects (15 female and 7 male, 18-79 years old) with mild difficulty falling asleep, for two nights, one after temazepam intake. However only the placebo night was available and used in our analysis. Besides the cohort and data acquisition methods, there are no other differences between the SC and ST data sets. The entire duration of each PSG (mean duration 8.3±1.1 h, $n = 61$) was used in our analysis and contains EEG (Fpz-Cz and Pz-Oz channels), EOG (horizontal) and submental EMG signals (100 Hz sampling rate) as well as a hypnogram of manual scores by a trained technician. The hypnograms, which mapped 30s epochs of data onto six states (non-REM 1-4, REM, and Wake), were relabeled per the current guidelines of the American Academy of Sleep Medicine [7] by combining non-REM stages 3 and 4. Hence, each hypnogram contained up to five labels: *N1*, *N2*, *N3* for non-REM, *R* for REM, and *W* for Wake.

Signal feature selection and extraction

All analysis was performed using custom-written code on the Matlab^{TM} environment (Mathworks Ltd., Natick, MA). Frontal EEG (Fpz-Cz) from each subject was bandpass-filtered into seven distinct frequency bands, specifically delta-low (0.5-2 Hz), delta-high (2-4 Hz), theta (4-9 Hz), alpha (9-12 Hz), sigma (12-16 Hz), beta (16-30 Hz), and gamma (30-45 Hz) using Butterworth IIR filters. The mean power fraction in each band was estimated in 30s epochs and combined into a vector of seven EEG features. The root-mean-squared (r.m.s.) values of broadband EMG and EOG were also included to give a vector *X* of nine features for analysis. All feature values were converted to a decibel scale, i.e., $10 \log_{10}(\cdot)$, to make the distributions more symmetric over their dynamic range and less sensitive to outliers. The choice of spectral bands reflects commonly recognized EEG rhythms; other selections of features may be used within the same modeling and analysis framework.

Sleep scoring algorithms

Supervised and unsupervised classification: A statistical classifier assigns sample measurements $\boldsymbol{X}$ to one of N discrete categories or classes $S \in \{1, \cdots, N\}$ by assuming a (usually parametric) statistical model of $\boldsymbol{X} \rightarrow S$. Examples of statistical classifiers are linear discriminant analysis (LDA), artificial neural networks (ANN), and support vector machines (SVM). In order to construct the statistical model, class-labeled training samples are usually required to estimate the parameters, and the model is referred to as a supervised classifier; all the above examples belong to this category.

Other models known as unsupervised classifiers can be used to fit models to unlabeled training data and predict the class membership of future observations. Such classifiers typically look for natural clusters in the data that may coincide with the classes of interest, in this case the sequence of vigilance states underlying the polysomnogram. Of course, the states modeled by an unsupervised classifier may not conform completely to an individual human rater's perceptions of class differences and are determined by the measurements and features used to estimate the model parameters. But such classifiers can still be very useful, especially when no prior class definitions are available; common examples are k-means, linkage trees, GMMs, and HMMs—though some of these may be supervised as well.

Here we describe a method for constructing *quasi-supervised* classifiers: models that tend to mimic a human rater's behavior when scoring information is available but look for structural clues in the training data when the available scores are selectively applied or uncertain. To demonstrate the feasibility of this approach, we use models that rely on Bayesian inference, specifically GMMs and HMMs.

Bayesian models, GMMs, and HMMs: We provide a brief overview of Bayesian models in the context of sleep scoring and the issues relevant to GMMs and HMMs. We emphasize intuition over mathematical rigor, and refer the interested reader to other sources for a formal theoretical treatment [14, 15, 18, 19].

First, we assume that the subject is always in one of N discrete, mutually exclusive vigilance states $S \in \{1, \cdots, N\}$, and that a vector of M features $\boldsymbol{X} = [x_1, \cdots, x_M]^T, \boldsymbol{X} \in \mathbb{R}^M$ (T = transpose), is extracted from samples of the signals in a polysomnogram in successive windows of time (e.g., 30 s duration), so that we have a set of observations

$X_{1:T} = \{X_1, \cdots, X_T\}$ that are made in states $S_{1:T} = \{S_1, \cdots, S_T\}$. Each value in set S represents a modeled state that may—but does not necessarily—correspond directly to a human rater-scored vigilance state ($N3$, $N2$, etc.). At an arbitrary time t, the subject may be in a vigilance state S_t but the state is quantified by the observation X_t. The classifier's task is to infer S_t from X_t with acceptable accuracy. It is expected that there will be some variability and noise in the estimation of X, and this is described by a probability density function $f(X)$ which, when integrated over a region of X, gives a probability measure $P(X)$.

Since the N states are mutually exclusive, the probability associated with an observation X integrates the probability that X is observed in any of the states: i.e.,

$$P(X) = \sum_S P(X \cap S) \qquad (1)$$

The probability that X is observed, when the state is known to be S, is the conditional:

$$P(X|S) = P(X \cap S)/P(S) \qquad (2)$$

where $P(S)$ represents the prior probability of state S in the absence of information about X. Eq. 2 is known as Bayes rule. From the above, we get an expression for the probability distribution of X in terms of the conditional and prior probabilities:

$$P(X) = \sum_S P(X|S)P(S) \qquad (3)$$

Starting from an observation X, we can now compute the posterior probability of state S as:

$$P(S|X) = P(S \cap X)/P(X) = P(X|S)P(S)/P(X) \qquad (4)$$

A reasonable prediction of state is the one that maximizes the posterior:

$$\hat{S} = \underset{S}{\operatorname{argmax}} P(S|X) \qquad (5)$$

A Bayesian model must assume knowledge of the conditional $P(X|S)$, usually in a standard parametric form, in order to make predictions. The GMM is one such model [14], in which $P(X|S)$ is expressed as a Gaussian distribution parameterized by a state-dependent mean vector $\mu_S \in \mathbb{R}^{N \times M}$ and covariance matrix $\Sigma_S \in \mathbb{R}^{M \times M}$. Each Gaussian component contributes to the mixture to a degree expressed by a linear coefficient α_S, which replaces the state prior $P(S)$ in Eq. 3. A GMM constructed from sleep data would assume that the observation X can be modeled as a mixture of Gaussian components, and

that each component corresponds to one of the known vigilance states (or perhaps their sub-states).

HMMs [15] can be used to capture the evolution of a process over time and have been used for modeling the dynamics of sleep [10, 20-22] An HMM adds a layer of complexity to Eq. 3 by linking the model states to one another. The purpose is to model not just independent observations but the distribution $P(\boldsymbol{X}_{1:T})$ of the ordered sequence (i.e., time series) of observations generated by a latent state sequence $S_{1:T}$. In this model, the current state exclusively determines the distribution of future states (viz. the Markov property):

$$P(S_t|S_{1:t-1}, \boldsymbol{X}_{1:t}) = P(S_t|S_{t-1}) \qquad (6)$$

This quantity is known as a state transition probability; its values for all possible combinations of S_{t-1} and S_t constitute an $N \times N$ state transition matrix γ, an essential property of the HMM. In addition to Markov transitions, the current observation is assumed conditionally independent of previous observations and states given the current state:

$$P(\boldsymbol{X}_t|S_{1:t}, \boldsymbol{X}_{1:t-1}) = P(\boldsymbol{X}_t|S_t) \qquad (7)$$

Along with a set of state priors $\pi = P(S)$, fixing γ and the conditional $P(\boldsymbol{X}|S)$ completely specifies the structure of an HMM; an assumption of stationarity makes these properties independent of time t. In our treatment, the observation $\boldsymbol{X}$ is multivariate Gaussian, and the model is therefore a Gaussian observation HMM (GO-HMM) [23].

The simplifying assumptions made above permit the recursive application of elementary rules of probability (the product rule and Bayes' theorem) to make inferences regarding the dynamics of the process underlying observations $\boldsymbol{X}_{1:T}$. A common problem solved using HMMs is to decode the sequence of states $S_{1:T}$ most likely to have generated $\boldsymbol{X}_{1:T}$. This is commonly accomplished using the Viterbi algorithm [15]. The algorithm is initialized by computing the distribution of the first observation $\boldsymbol{X}_1$ as $\delta_1(S) = P(\boldsymbol{X}_1|S)$, for $S \in \{1, \cdots, N\}$, and keeping track of the preceding state that maximizes the probability of each successive observation $\delta_t(S') = \max_S[\delta_{t-1}(S)\gamma(S, S')] P(\boldsymbol{X}_t|S')$. At termination, the optimal path probability is $P^*(S) = \max_S \delta_T(S)$ and the terminal state is the one that maximizes $P^*(S)$. We can now backtrack along the sequence δ_t to identify the most likely predecessor at each step and recover the best state sequence $S_{1:T}$.

GMM and HMM parameters are estimated from training data using maximum likelihood (ML) techniques. In ML estimation [18], a likelihood function L is defined as the joint probability density of a set $\boldsymbol{X}_{1:T}$ of independent and identically distributed observations for the chosen model with parameter set Θ (e.g., $\Theta = \{\alpha_S, \boldsymbol{\mu}_S, \Sigma_S\}$ for a GMM):

$$L(\Theta|\boldsymbol{X}_{1:T}) = P(\boldsymbol{X}_{1:T}|\Theta) = \prod_{t=1}^{T} P(\boldsymbol{X}_t|\Theta) \qquad (8)$$

Taking the logarithm on both sides converts the product into a sum over the sample data:

$$\log L = \sum_{t=1}^{T} P(\boldsymbol{X}_t|\Theta) \qquad (9)$$

The likelihood function L expresses the parameters as a function of the fixed observations. ML estimation proceeds by taking the partial derivative of $\log L$ with respect to each parameter, equating it to zero, and solving the resulting system of equations for the unknown parameters Θ that maximize $\log L$ (hence the name ML).

When labeled training data exist, ML estimates of GMM and HMM parameters are relatively easy to derive and compute: for instance, the ML estimate of the true mean of state S is merely the arithmetic average of independent training samples labeled as S by a human rater; similarly for the covariance matrices, state priors, and transition matrix.

If no labeled training data are available, the observations become related to the parameters through hidden variables (the states $S_{1:T}$) apart from the unknowns Θ, and we have:

$$\log L = \sum_{t=1}^{T} P(\boldsymbol{X}_t, S_t|\Theta) \qquad (10)$$

with unknowns on either side of the conditional. This is often intractable, since $\log L$ must now be maximized over all possible state paths for $S_{1:T}$ to determine the correct maximum. One solution to this problem is to use an E-M algorithm (for Expectation-Maximization) [18]. E-M is an iterative process that converges to a local maximum when given an initial guess of the model parameters. In order to avoid getting trapped in a local trough, several initial guesses within the search space are tested and the solution with greatest likelihood is selected. A popular version of E-M used for HMMs is the Baum-Welch algorithm [15].

A framework for quasi-supervised classification: We have seen how GMMs and HMMs can be estimated and used to predict state when labeled or unlabeled training data are

available. Though such models are widely used, there are no methods to address situations in which sample scores are limited or uncertain. Here we propose a simple method for building quasi-supervised classifiers that use partial scores to stage sleep.

Consider a scored polysomnogram from which a sequence of labeled observations $X_{1:T}$ is derived. Let each $X_{1:T}$ be augmented with another vector $e = [e_1, \cdots, e_K]^T$ so that

$$Z^T = [X^T e^T] \implies Z \in \mathbb{R}^{M+K} \qquad (11)$$

where K is the number of unique states labeled by the human rater (in the hypnogram). For instance, $K = 3$ if the rater labels R and W but does not distinguish between $N1$, $N2$, and $N3$ in non-REM sleep.

Just as for $X_{1:T}$, we can model $Z_{1:T}$ as an N-state GMM with parameters $\Theta = \{\alpha_S, \mu_S, \Sigma_S\}$ by initializing the parameters with randomized seeds and following the E-M algorithm until it converges to the solution with greatest likelihood. The N modeled states are not necessarily identical to the K states scored by the rater. They must be selected by the user to suit the problem at hand. This flexibility is important in different scoring scenarios, as we will see below. Finally, the values in e are chosen based on the state label S_t assigned by a human rater to each observation X_t.

Let us start with $K = 5$ vigilance states (for $N3$, $N2$, $N1$, R, and W) scored from a polysomnogram in 30 s epochs. The time series $X_{1:T}$ extracted from the signals can be fitted using an E-M algorithm to a GMM or HMM with $N = 5$ states. If the value of e_t is uncorrelated with S_t (for instance, always a zero vector), then the E-M algorithm simply yields an unsupervised classifier that optimizes the fit of the model to the observed data. If, on the other hand, e_t bears some correlation to the scored state S_t, we can expect the model to tend toward the human rater's scoring patterns. But S is a categorical variable, and therefore incompatible with X in the augmented vector Z. So what form should e take?

Recall that S takes on values from $\{1, \cdots, K\}$. Let us define e so that:

$$e_j = \begin{cases} 1 \text{ if } S = j \\ 0 \text{ otherwise} \end{cases}, j \in \{1, \cdots, K\} \qquad (12)$$

Each state S is now identified by a unit vector e in K dimensions. It follows that for two observations at times t and t':

$$e_t e_{t'} = \begin{cases} 1 \text{ if } S_t = S_{t'} \\ 0 \text{ otherwise} \end{cases} \qquad (13)$$

That is, the set of values assumed by $\boldsymbol{e}$ form an orthogonal basis. This lets us incorporate the state label S, a categorical variable, into the quantitative description $\boldsymbol{X}$ of a sample without otherwise altering its properties or imposing an artificial ordering on the states. Adopting this definition for $\boldsymbol{e}$ in Eq. 11, intuition tells us that if $\boldsymbol{X}$ is now set to zero, the E-M algorithm will cluster the data strictly on the basis of scores S_t—in effect, a supervised classifier. Observations augmented with similar tags $\boldsymbol{e}$ will cluster since they are closer to each other in the augmented feature space $\mathbb{R}^{M+K}$ than in $\mathbb{R}^M$; by the same logic, samples with unlike tags are farther apart and less likely to form a cluster. Hence tagging the training samples makes an unsupervised classifier behave like a supervised one. If the tags are excluded in the training step (or all set to be identical), the E-M algorithm converges to the unsupervised model. The tags incorporate the knowledge and intuition of a human rater into the parameter estimation. While the unsupervised and supervised asymptotes are illustrative and set bounds on the resulting model, it is situations where only partial scoring information is available that determines the utility of the quasi-supervised algorithm.

To conclude, the algorithm proceeds as follows (see Fig. 1): Available categorical scores S are transformed into vector "tags" $\boldsymbol{e}$ of length equal to the number of scored vigilance states K. The tags are attached to the vector of training observations $\boldsymbol{X}$ to give augmented input variables $\boldsymbol{Z}$. Starting with randomized initial guesses for the model parameters, a GMM or HMM is estimated from $\boldsymbol{Z}$ using the appropriate E-M algorithm with the desired number of states N specified. After stripping entries corresponding to the tag $\boldsymbol{e}$ from parameters $\boldsymbol{\mu}_S$ and Σ_S, the model is then used to predict the state in epochs for which scores are unavailable or uncertain based on un-augmented observations $\boldsymbol{X}$ (i.e., not $\boldsymbol{Z}$). This approach is quasi-supervised in that model parameters are estimated using exactly the same methods as for unsupervised classifiers—except that the samples are tagged with a score-based vector—but converges to a strictly supervised classifier when complete scoring information is incorporated into the training data. The choice of score tags $\boldsymbol{e}$ is critical and can be tailored to address different typical scoring scenarios, as illustrated below.
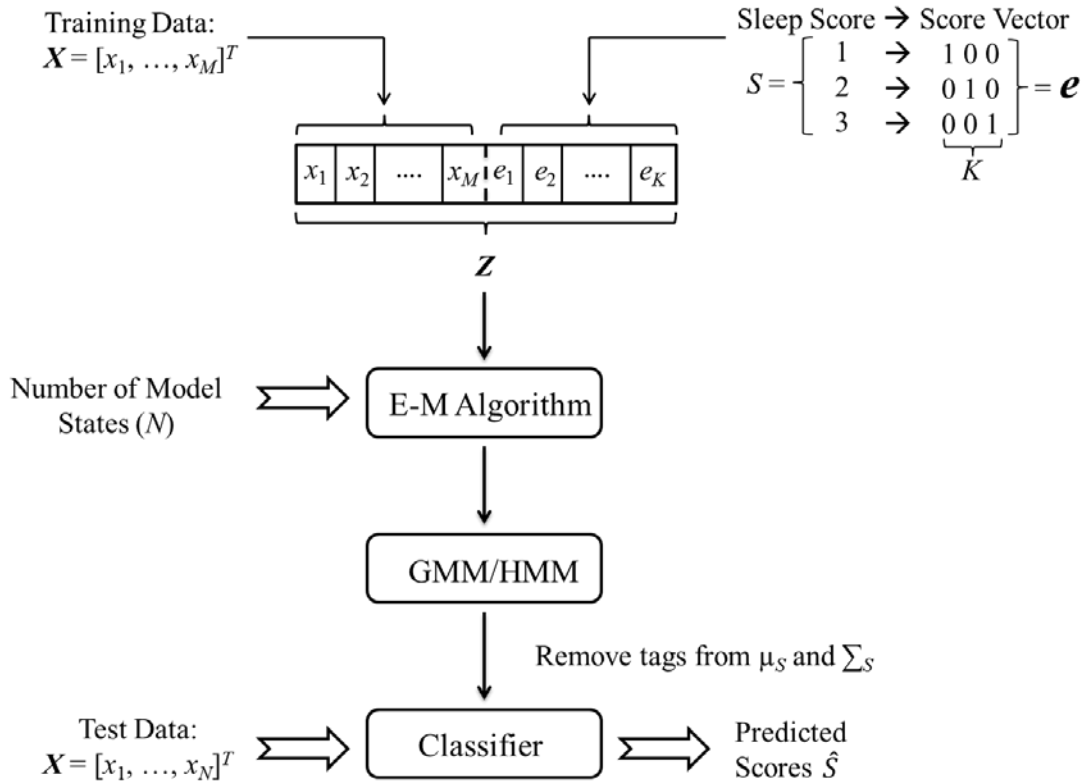
Figure 1. Flow diagram for quasi-supervised classification. A vector X of M features is computed from each epoch of a polysomnogram. The sleep score S is converted into a unit vector e whose length depends on the number K of states scored by the rater. X is augmented with e to give Z, the input to an E-M algorithm, which estimates the parameters of the GMM or HMM that maximizes the likelihood that a model with $N \geq K$ states explains the data. The excess dimensions are removed from the mean vector $\boldsymbol{\mu}_S$ and covariance matrix Σ_S of each state in the model. The model is then used to classify new unlabeled inputs X, or the same data in which only K states were previously labeled, into N states.

Analysis procedure

The general procedure followed for analysis is common for Problems 1 to 3 below except where noted. First, surrogates were prepared from the available hypnograms based on the requirements of each problem. Then samples of the observation vector X were augmented with a scoring vector e chosen from one of K unique values corresponding to the states

scored on the surrogate hypnogram (Fig. 1). The number of states, N, to be modeled was fixed and the score-augmented variables used to estimate GMM and HMM parameters through an E-M algorithm. The models were used to predict the sequence of vigilance states in each polysomnogram and on a second night's data when available. Performance was assessed in terms of Cohen's K statistic [24], which measures the agreement in categorical scores on a sample scored by two independent raters. K was used here to assess concordance between the model predictions and true hypnogram, separately for each vigilance state and then for all states pooled together. These metrics were compared for the quasi-supervised method against reference methods in which the same algorithm was applied, but in a completely unsupervised (no tags) and completely supervised (unit basis vector tags used for all five states: $K = 5$). This is intended to help evaluate the extent to which the quasi-supervised classifier is able to compensate for incomplete score information in the training data. Since each polysomnogram is analyzed independently by the three algorithms, differences in K for the cohorts (same night and second night) were investigated using a Wilcoxon sign rank test separately for the quasi-supervised classifier versus the unsupervised and supervised classifiers respectively. In each comparison, a false positive probability p under 0.01 was considered statistically significant.

Problem 1: Human rater is uncertain about certain vigilance states. Here we consider the situation in which the rater is confident of identifying some states but not others. For instance, she is sure of the distinction between *W*, *R*, and *N*, but not stages of *N*, i.e., *N1*, *N2*, and *N3*. Hence labels are not available for three of the five states and completely supervised classification is not possible. On the other hand, unsupervised classification does not take advantage of the available scores for *W*, *R*, and *N*. In our quasi-supervised approach, we collapse stages of *N* into one label on the hypnogram ($K = 3$), and tag *W*, *R*, and *N* with unit vectors $\boldsymbol{e}$ (specifically $[1\ 0\ 0]^T$ for *N*, $[0\ 1\ 0]^T$ for *R*, and $[0\ 0\ 1]^T$ for *W*) but fit the data to a GMM or HMM with $N = 5$ since we wish to recover all the vigilance states. The expectation is that *W*, *R*, and *N* will be separated by the E-M algorithm based on their disparate tags, but that three natural partitions or sub-states corresponding to *N1*, *N2*, and *N3* will be required to adequately fit the model to samples of *N* based on the distribution of $\boldsymbol{X}$.

We test the utility of this approach in situations where the rater does not distinguish between the following states: I. *N1*, *N2*, and *N3*; II. *N1* and *W*; III. *W* and *R*; IV. *N1* and *N2*; and V. *N1* and *R*. These choices reflect typical sources of confusion faced by human raters in scoring sleep [8, 33, 40].

Problem 2: Human rater scores all vigilance states, but only labels epochs with clear manifestations. Suppose that the rater labels samples of all five vigilance states, but only those epochs for which he is sure of the predominant state. This can happen at the transitions between different states or in the presence of artifacts. We simulate this situation by deleting the scores from three successive epochs at each state transition in the hypnogram. In the solution, the score tags e are set to orthogonal unit vectors of length $K = 5$ but to a zero vector for unscored epochs. In the modeling step, as in Problem 1, we specify $N = 5$ states. Since e for unscored epochs is equidistant from all the unit vector tags in $\mathbb{R}^K$, the E-M algorithm allocates scored epochs to the five states according to the tag e, but distributes the unlabeled epochs among these states based on X.

Problem 3: Two or more raters score a polysomnogram and one model is to be trained, but there is some level of disagreement between them. Here, each rater produces a hypnogram but there is only one sequence of observations to be modeled. Since only one rater's scores were available for each recording, we simulated a scenario in which two or more human raters disagree about one-third of the time by generating surrogate hypnograms in which 33% of randomly selected epochs had their scores deleted. The quasi-supervised classifier was then used to complete the scores and its performance evaluated against the original hypnogram. While this is not strictly identical to the case of inter-rater disagreement, it is expected that it is a reasonable simulation of that scenario.

RESULTS

Table I summarizes the incidence of states *N1*, *N2*, *N3*, *R*, and *W* in each hypnogram in terms of the number of 30 s epochs and the percent time spent in that state. Results of analysis for Problems 1, 2, and 3 using HMMs are presented in Tables II-IV. The corresponding results obtained using GMMs are presented in the Supplement and are

referred to as Tables S1, S2, and S3. The performance of HMMs was consistently better than GMMs, with the same trends being observed in different scenarios.

Problem 1. Only some vigilance states are scored by the human rater

Tables II and S1 give the performance of the quasi-supervised algorithm in terms of Cohen's K, compared to completely unsupervised and supervised implementations, for a GMM and HMM. Results are presented separately for each state and finally for all states together. Four different scenarios are explored in which some of the vigilance states were assigned identical scores to simulate scoring uncertainty: Case I. *N1*, *N2*, and *N3*; Case II. *N1* and *W*; Case III. *W* and *R*; and Case IV. *N1* and *N2*. Each entry in the table represents Cohen's K averaged over 42 overnight PSGs along with the standard error of the mean.

Table 1 Distribution of sleep states per PSG

State	**First night** (*n* = 42)		**Second night** (*n* = 19)	
State	Epochs	% Time	Epochs	% Time
N3	146±12	15±1	149±17	16±2
N2	445±20	45±2	460±33	45±2
N1	79±7	8±1	61±8	6±1
R	188±9	19±1	188±13	18±1
W	127±10	13±1	153±9	15±1
All values reported as mean±standard error.				

In general—with a few exceptions for individual states—the proposed quasi-supervised classifier performs significantly better in terms of K than the unsupervised model but not as well as the completely supervised model, which represents the maximum attainable performance when complete scoring information is available. When all states are considered, K for the quasi-supervised classifiers is within the 60-80% range, which is

thought to indicate excellent agreement [25]; in fact, K of 80% for five states in equal proportion would mean perfect agreement, which is highly unlikely in practice. In contrast, K for the unsupervised classifiers is close to 50% in all cases, i.e., moderate agreement.

The HMM almost always outperformed the GMM but only by a small margin. When examining the predictions for each hypnogram, the difference was attributed to noise fluctuations in the GMM predictions that are smoothed by the HMM, which optimizes the entire sequence rather than the state in each epoch without context (see Fig. 2).

Table 2 HMM classifier accuracy κ^a for Problem 1 (first night PSG; n = 42).

	Scenario:	I (N1, N2, N3 pooled)	II (N1 and W pooled)	III (W and R pooled)	IV (N1 and N2 pooled)	V (N1 and R pooled)
	Unsupervised	63±4[*]	64±3	64±3	64±3	64±3
N3	Quasi-supervised	60±4	73±4	75±4	76±4	73±4
	Supervised	83±2	83±2	83±2[†]	83±2[†]	83±2
	Unsupervised	51±2[*]	51±2	50±2	51±2	49±2
N2	Quasi-supervised	57±3	70±2	73±2	69±2	69±2
	Supervised	82±1	82±1	82±1	82±1	82±1
	Unsupervised	14±3	16±3[*]	14±3	16±3	13±3
N1	Quasi-supervised	35±4	22±4	52±4	34±4	6±3
	Supervised	66±2	66±2	66±2	66±2	66±2
	Unsupervised	59±3	58±4	57±3	60±3	57±3
R	Quasi-supervised	90±1	89±1	68±4	91±1	74±2

	I	II	III	IV	V
	Supervised	90±1†	90±1†	90±1	90±1†	90±1
	Unsupervised	51±4	50±5	51±5*	49±5	51±4
W	Quasi-supervised	81±2	62±5	46±6	80±3	79±3
	Supervised	87±1	87±1	87±1	87±1	87±1
	Unsupervised	50±2	50±2	49±2	50±1	49±1
All	Quasi-supervised	65±2	68±2	68±2	73±2	65±2
	Supervised	83±1	83±1	83±1	83±1	83±1

[a]Cohen's kappa (mean± standard error). *Quasi-supervised model is not significantly different ($p > 0.01$) from unsupervised model according to a Wilcoxon sign rank test (matched samples). †Quasi-supervised model is not significantly different ($p > 0.01$) from supervised model according to a Wilcoxon sign rank test (matched samples).

Table S1 GMM classifier accuracy κ[a] for Problem 1 (first night PSG; n = 42).

		I	II	III	IV	V
	Scenario:	**(N1, N2, N3 pooled)**	**(N1 and W pooled)**	**(W and R pooled)**	**(N1 and N2 pooled)**	**(N1 and R pooled)**
	Unsupervised	63±3*	64±3	64±3	64±3	63±3
N3	Quasi-supervised	60±4	75±4	77±3	77±3	75±4
	Supervised	82±1	82±1†	82±1†	82±1†	82±1†
	Unsupervised	49±2*	49±2	49±2	49±2	48±2
N2	Quasi-supervised	54±2	67±3	69±3	64±2	64±3
	Supervised	81±1	81±1	81±1	81±1	81±1
	Unsupervised	13±3	14±3*	14±3	15±3	13±3*

N1	Quasi-supervised	30±3	21±4	50±4	31±3	2±2
	Supervised	62±2	62±2	62±2	62±2	62±2
	Unsupervised	54±3	54±3	53±3[*]	54±3	53±3
R	Quasi-supervised	85 ±1	85±1	62±5	85±1	71±2
	Supervised	86 ±1[†]	86±1	86±1	86±1[†]	86±1
	Unsupervised	52±4	52±4[*]	52±4[*]	50±4	52±4
W	Quasi-supervised	81±2	58±5	47±6	79±3	77±3
	Supervised	85±1[†]	85±1	85±1	85±1[†]	85±1
	Unsupervised	48±2	48±2	48±2	48±1	47±2
All	Quasi-supervised	63±2	65±2	65±2	69±2	61±2
	Supervised	81±1	81±1	81±1	81±1	81±1

[a]Cohen's kappa (mean± standard error). *Quasi-supervised model is not significantly different (p > 0.01) from unsupervised model according to a Wilcoxon sign rank test (matched samples). †Quasi-supervised model is not significantly different (p > 0.01) from supervised model according to a Wilcoxon sign rank test (matched samples).

In each of the four case studies of selective scoring examined, the quasi-supervised classifier significantly improved on the unsupervised model for states that were not scored (in the surrogate hypnogram), but not to the extent that it matches the supervised model; for the scored states however, the quasi-supervised classifier rivals the supervised classifier in performance. This indicates that the proposed algorithm is able to track the human rater when scores are available but can still uncover the unscored states by modeling variability in the observed data. Fig. 2 illustrates this using a spectrogram derived from a sample polysomnogram. Although the scores used to construct the quasi-supervised models did not differentiate between *N1*, *N2*, and *N3*, the GMM and HMM are both able to recover the scores for these states quite well, thus saving the human rater

the inconvenience of having to make these distinctions. K appeared to be relatively low for *N1*, even for the supervised classifier, in all four case studies. This is easily explained by the very low incidence of *N1* in the data (see Table I), which means that there are few samples for any of the classifiers to train on or distinguish from the other vigilance states. In truth, stage *N1*, occurring at the transition between *W* and *N2*, is notoriously hard to distinguish. While *W* is more easily characterized by elevated muscle tone and active EOG, and *N2* displays distinctive transients such as sleep spindles and K complexes, *N1* is in a gray area that human raters find hard to demarcate. These factors taken together contribute to the poor classification performance on *N1*. A second night's recording was available in 19 of the 42 subjects analyzed. For these subjects, Tables III and S2 give the performance of each classifier trained on the first night of recording but applied blind to data from the second night. Unlike Tables II and S1, which represents a composite of performance with and without scoring information on the same data set, the results in Tables III and S2 are strictly derived from out-of-sample classification. As expected, K for all states together was lower for all three approaches, unsupervised, quasi- and supervised while following similar trends to those noted in Tables II and S1 when comparing scored versus unscored states and GMMs versus HMMs. K for the quasi-supervised classifier was close to 60%, which is lower than in Tables II and S1 but still acceptable, especially when considering that K for the unsupervised classifier now dwells close to 45%; nor is the supervised classifier that much better at 65-70%.
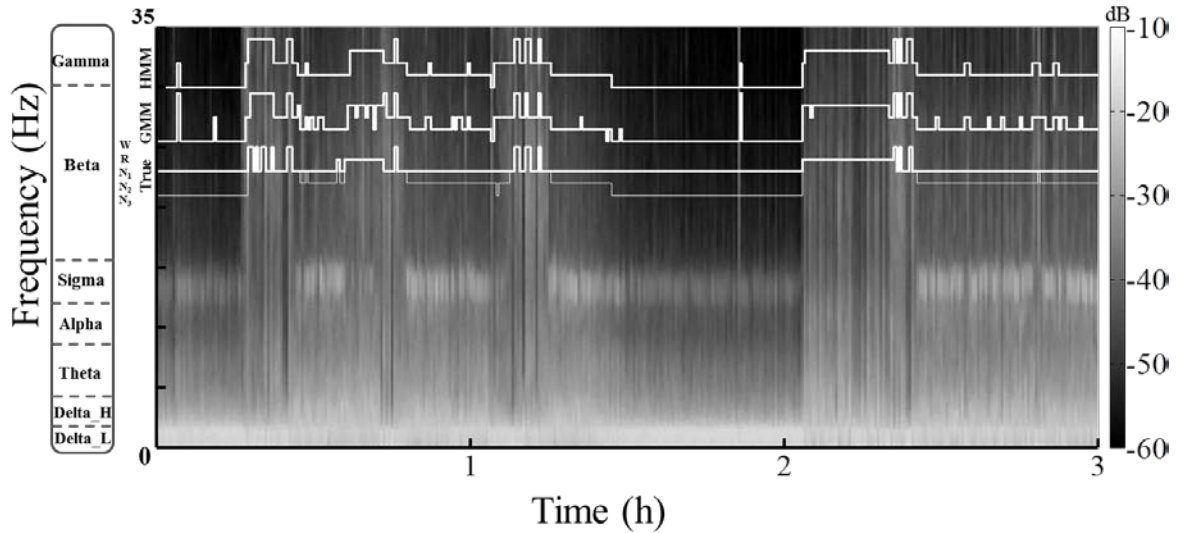
Figure 2. Automatic sleep scoring when only some states are labeled by the human rater in the training data (Problem 1). The figure shows a 3 h sample (starting at 2 a.m.) from a spectrogram , i.e., the distribution of signal power in decibels (dB) by frequency over time, computed for an 8 h recording in 30 s epochs of EEG from Fpz-Cz. Overlaying the image are staircase plots of the True five-state hypnogram (thin line); the surrogate three-state hypnogram (thick line), which does not differentiate between N1, N2, and N3; and the hypnograms predicted by the quasi-supervised GMM and HMM, which were trained using input features augmented with a score vector derived from the surrogate hypnogram. A comparison of model predictions with the true hypnogram shows that the GMM and HMM are able to reconstruct the unlabeled states with reasonable accuracy even as they track the human rater's scores of the labeled states. The HMM is less susceptible to noise fluctuations than the GMM, resulting in slightly better performance.

Table 3 HMM classifier accuracy κ^a for Problem 1 (second night PSG; n = 19).

	I	II	III	IV	V
Scenario:	(N1, N2, N3 pooled)	(N1 and W pooled)	(W and R pooled)	(N1 and N2 pooled)	(N1 and R pooled)
Unsupervised	58±7[*]	59±6	59±6	57±7[*]	56±7[*]

N3	Quasi-supervised	59±6	74±5	72±5	69±6	67±5
	Supervised	76±4	76±4[†]	76±4[†]	76±4[†]	76±4[†]
	Unsupervised	44±5[*]	46±5	47±5	44±5	46±6
N2	Quasi-supervised	55±4	66±4	66±3	62±4	65±4
	Supervised	72±3	72±3[†]	72±3[†]	72±3	72±3[†]
	Unsupervised	8±3[*]	7±3[*]	6±3	6±3[*]	7±3[*]
N1	Quasi-supervised	18±5	15±5	26±5	14±4	4±3
	Supervised	23±5[†]	23±5[†]	23±5[†]	23±5[†]	23±5
	Unsupervised	41±7	40±7	39±7	40±7	40±7
R	Quasi-supervised	62±7	67±5	58±6	67±6	59±6
	Supervised	69±6[†]	69±6[†]	69±6	69±6[†]	69±6
	Unsupervised	54±6	51±7[*]	54±6[*]	56±5	57±5
W	Quasi-supervised	65±5	63±5	55±7	71±5	71±5
	Supervised	74±5[†]	74±5	74±5	74±5[†]	74±5[†]
	Unsupervised	44±4	44±4	45±4	44±3	45±4
All	Quasi-supervised	56±3	63±3	61±3	62±3	60±3
	Supervised	69±3	69±3[†]	69±3[†]	69±3	69±3

[a]Cohen's kappa (mean± standard error). *Quasi-supervised model is not significantly different ($p > 0.01$) from unsupervised model according to a Wilcoxon sign rank test (matched samples). †Quasi-supervised model is not significantly different ($p > 0.01$) from supervised model according to a Wilcoxon sign rank test (matched samples).

Table S2 GMM classifier accuracy κ[a] for Problem 1 (second night PSG; n = 19).

		I	II	III	IV	V
	Scenario:	**(N1, N2, N3 pooled)**	**(N1 and W pooled)**	**(W and R pooled)**	**(N1 and N2 pooled)**	**(N1 and R pooled)**
N3	Unsupervised	61±6[*]	60±6[*]	58±5	58±6[*]	59±6[*]
	Quasi-supervised	60±6	69±6	71±5	71±5	66±6
	Supervised	75±4	75±4[†]	75±4[†]	75±4[†]	75±4[†]
N2	Unsupervised	46±5[*]	44±5	44±5	43±5[*]	46±5[*]
	Quasi-supervised	49±3	63±4	64±3	57±3	63±3
	Supervised	70±3	70±3[†]	70±3[†]	70±3	70±3[†]
N1	Unsupervised	8±3[*]	8±3[*]	7±3	9±3[*]	9±3[*]
	Quasi-supervised	14±4	13±4	23±4	11±4	1±2
	Supervised	25±4	25±4	25±4[†]	25±4[†]	25±4
R	Unsupervised	41±6	37±7	38±6[*]	40±6	39±6
	Quasi-supervised	64±5	63±5	53±6	64±5	57±5
	Supervised	63±5[†]	63±5[†]	63±5	63±5[†]	63±5
W	Unsupervised	53±5	52±5[*]	53±6[*]	54±5	55±5
	Quasi-supervised	69±5	65±4	54±6	71±4	72±4
	Supervised	71±4[†]	71±4[†]	71±4	71±4[†]	71±4[†]
	Unsupervised	45±3	43±3	43±3	44±3	45±3

All	Quasi-supervised	54±3	60±2	58±3	59±3	58±3
	Supervised	67±3	67±3[†]	67±3[†]	67±3	67±3[†]

[a]Cohen's kappa (mean± standard error). *Quasi-supervised model is not significantly different (p > 0.01) from unsupervised model according to a Wilcoxon sign rank test (matched samples). †Quasi-supervised model is not significantly different (p > 0.01) from supervised model according to a Wilcoxon sign rank test (matched samples).

Problem 2. Only some epochs are scored, but for all vigilance states

Results for Problem 2 are presented in Tables IV and S3. The overall performance of the quasi-supervised classifier is somewhat improved by a few points relative to Problem 1 for the first night analysis as well as for the second night, which is completely out-of-sample data. This is to be expected since sample scores are available here for all five vigilance states (except at the transitions between states) and the algorithm is not forced to come up with its own definitions. Of course, the unsupervised and supervised classifiers perform about the same as before since the scoring information provided to them is unchanged. From the spectrogram in Fig. 3, it can be seen that the model appears to fill in the missing scores at the transitions between states in a reasonably satisfactory manner.

Problem 3. One classifier must be constructed based on the sample scores of multiple raters

Tables 4 and S3 also summarize results for Problem3. The performance of the GMM and HMM classifiers for in-sample and out-of-sample data is very similar to that obtained for Problem2. It shows that even when a full third of the data is left unscored, the model is still capable of filling the blanks with reasonable accuracy.
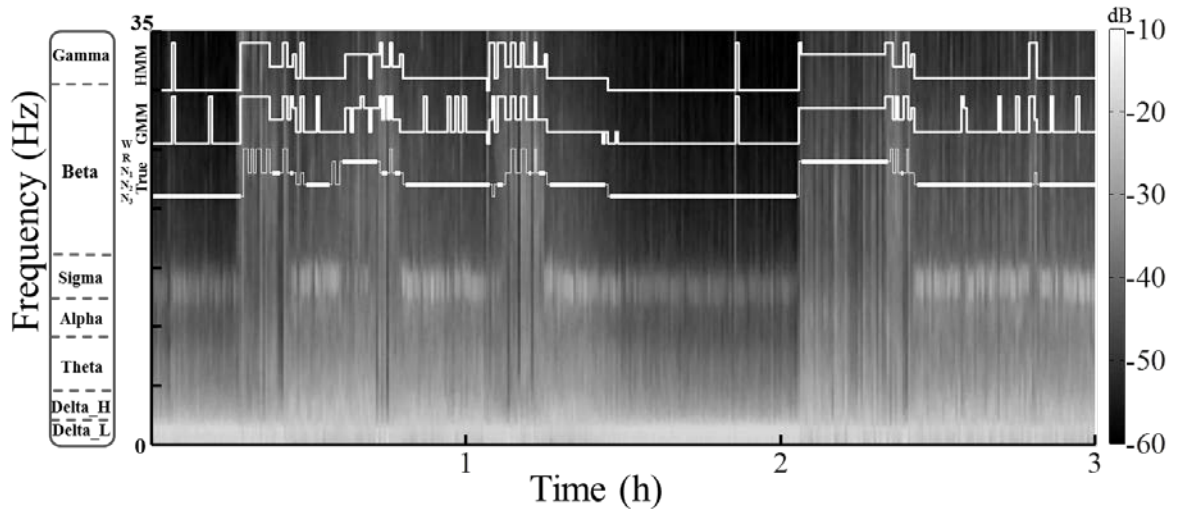
Figure 3. Automatic sleep scoring when all states are labeled by the human rater, but not for all epochs in the training data (Problem 2). The figure shows a 3 h sample (starting at 2 a.m.) from a spectrogram , i.e., the distribution of signal power in decibels (dB) by frequency over time, computed for an 8 h recording in 30 s epochs of EEG from Fpz-Cz. Overlaying the image are staircase plots of the True five-state hypnogram (thin line); the surrogate five-state hypnogram (thick line), in which scores are deleted for three successive epochs at each state transition; and the hypnograms predicted by the quasi-supervised GMM and HMM, which were trained using input features augmented with a score vector derived from the surrogate hypnogram. A comparison of model predictions with the true hypnogram shows that the GMM and HMM are able to track changes in vigilance state across state transitions.

Table 4 HMM classifier accuracy κ^{a} for Problem 2 and Problem 3

		Problem 2		Problem3	
		First night PSG **(in-sample test; $n = 42$)**	**Second night PSG** **(out-of-sample test; $n = 19$)**	**First night PSG** **(in-sample test; $n = 42$)**	**Second night PSG** **(out-of-sample test; $n = 19$)**
	Unsupervised	63±3	57±6[*]	64±4	62±4
N3	Quasi-supervised	73±4	66±7	73±4	74±3
	Supervised	83±2	76±4[†]	83±2	80±2
	Unsupervised	50±2	46±5	50±2	51±3
N2	Quasi-supervised	75±2	71±3	77±1	72±3
	Supervised	82±1	72±3[†]	82±1	77±2
	Unsupervised	12±3	9±4	13±3	8±2
N1	Quasi-supervised	36±4	18±4	35±4	24±4

	Supervised	66±2	23±5[†]	66±2	45±4
	Unsupervised	58±3	39±7	59±3	52±4
R	Quasi-supervised	88±1	66±6	81±3	70±4
	Supervised	90±1	69±6[†]	90±2	80±3
	Unsupervised	51±4	56±6[*]	53±4	61±4
W	Quasi-supervised	73±3	64±5	66±5	73±4
	Supervised	87±1	74±5	87±1	83±3
	Unsupervised	49±2	45±4	50±1	50±3
All	Quasi-supervised	74±1	64±3	74±1	69±3
	Supervised	83±2	69±3	83±1	77±3

[a]Cohen's kappa (mean± standard error). *Quasi-supervised model is **not** significantly different ($p > 0.01$) from unsupervised model according to a Wilcoxon sign rank test (matched samples). [†]Quasi-supervised model is **not** significantly different ($p > 0.01$) from supervised model according to a Wilcoxon sign rank test (matched samples).

Table S3 GMM classifier accuracy κ^{a} for Problem 2 and Problem 3

	Problem 2		Problem3	
	First night PSG	**Second night PSG**	**First night PSG**	**Second night PSG**
	(in-sample	**(out-of-sample**	**(in-sample**	**(out-of-sample**

		test; $n = 42$)	test; $n = 19$)	test; $n = 42$)	test; $n = 19$)
	Unsupervised	63±3	58±5[*]	63±3	63±4
N3	Quasi-supervised	73±4	62±7	68±4	71±3
	Supervised	82±1	75±4[†]	82±1	80±2
	Unsupervised	49±2	44±5	49±2	50±3
N2	Quasi-supervised	73±2	67±3	70±2	66±3
	Supervised	81±1	70±3[†]	81±1	75±2
	Unsupervised	13±3	10±4[*]	14±3	8±2
N1	Quasi-supervised	32±4	16±3	25±4	17±3
	Supervised	62±2	25±4[†]	62±2	44±4
	Unsupervised	54±3	39±6	53±3	48±4[*]
R	Quasi-supervised	82±1	60±6	69±3	54±4
	Supervised	86±1	63±5[†]	86±1	76±3
	Unsupervised	50±4	52±6[*]	52±4[*]	61±3[*]
W	Quasi-supervised	70±3	65±4	62±5	65±4
	Supervised	85±1	71±4[†]	85±1	80±2
	Unsupervised	48±2	44±3	48±1	49±3

All	Quasi-supervised	71±2	60±2	66±1	61±3
	Supervised	81±1	67±3	82±1	75±3

[a]Cohen's kappa (mean± standard error). *Quasi-supervised model is *not* significantly different ($p > 0.01$) from unsupervised model according to a Wilcoxon sign rank test (matched samples). [†]Quasi-supervised model is *not* significantly different ($p > 0.01$) from supervised model according to a Wilcoxon sign rank test (matched samples).

DISCUSSION

Computerized sleep scoring is desirable because with it comes the prospect of objective, data-driven segmentation of vigilance states that can consistently be applied to get reproducible output. Unsupervised sleep scoring has been pursued almost since the advent of digital EEG. The earliest efforts encoded heuristics used by experts in their visual analysis to process spectral measures or other quantitative features of polysomnographic signals and divide them into different states of vigilance [26, 27]. The goal was to produce a reasonable first pass segmentation that could quickly be refined by an expert into a final sequence of scores. Not surprisingly, advances in machine learning techniques have prompted various approaches—particularly probabilistic models—to the task of finding natural partitions in sleep data that could correspond to different vigilance states. The HMM is one such modeling technique that maps continuous observations onto discrete hidden states [15]. Early statistical models of sleep dynamics used Markov chain models to represent probabilistic transitions between stages of sleep extracted from expert-scored hypnograms [28]. These models have become more refined and are being used to characterize disordered sleep and the effect of medication [29, 30]. The HMM is a natural extension of the Markov chain that assumes the polysomnogram to comprise a sequence of observations generated by Markov states that are hidden from view [15]. This has contributed to its popularity in automatic sleep scoring [10, 20-22]. HMM parameters are usually estimated using unsupervised ML techniques; so the modeled

states are not biased by human opinion. They are, however, dependent on the features chosen to represent the data and how much they vary between vigilance states.

Unsupervised scoring can give very reasonable results without prior training, but must ultimately satisfy the gold standard of human assessment. Despite well-defined guidelines—first suggested in the 1960s [31]—that have evolved over time to reflect a growing consensus [7, 32], agreement between human raters scoring the same recording is hardly perfect and can be quite variable. One recent study comparing sleep scores between raters from two laboratories in different countries [8] found only moderate agreement for controls (mean K = 0.57) that was still lower for a cohort with narcolepsy (mean K = 0.54). The greatest disagreement was seen between scores on stages *N1* and *N2*, *N2* and *N3*, and *N1* and *W*; in Problem 1, we used our algorithm to distinguish between these states without supervision. A larger study [33] with independent raters from nine centers found better overall agreement (mean K = 0.63) although agreement by sleep stage still varied over a wide range. A rater has opinions forged by his or her training that can mutate over time and with experience. For this reason it is difficult to predict to what extent an unsupervised classifier will agree with a particular human rater.

There is another perhaps more obvious motivation for automatic sleep scoring: a computer algorithm may never be perfect in the eyes of one rater or another, but it can be programmed to behave like one. Models built for this purpose are known as supervised classifiers. A statistical model can be trained to mimic the scoring habits of a particular human rater, thus alleviating (if not eliminating) the burdensome task of manual scoring. Supervised sleep scoring also has a long history. Early efforts have used discriminant analysis [34] and distance metrics [35] of from samples of human-scored vigilance states to determine the scores of incoming data. Fisher discrimination, in which the input features are transformed to optimize the separation between samples of different states, has also been employed. More recent supervised schemes continue to make their way into this domain as and when they are developed or as increases in computing power makes it feasible to do so: these include linear discriminant analysis [36], neural networks and their variants [37], support vector machines [38], and random forest classifiers [39]. Increased computing power has also made it feasible to enlarge the feature space in a bid to better fit training data and improve performance. But the true measure of a supervised

classifier remains its ability to accurately score new recordings, i.e., out-of-sample data. The ability of a classifier trained on one cohort (e.g., healthy controls) to score data from another cohort (e.g., individuals with possible sleep disordered breathing) remains a concern.

We have discussed how unsupervised classifiers can model observations unconstrained by human-defined vigilance states, and how supervised classifiers can encode and mimic a specific human rater's scoring patterns. The middle ground in which a classifier seeks its own definitions but defers to human judgment when required has not been explored. In this manuscript, we have described an algorithmic framework that compensates for rater uncertainty and incomplete training data to automatically score sleep in a polysomnogram.

We accomplish this quasi-supervised classification by transforming categorical sleep scores into numerical variables or *tags* that link the scores to continuous-valued features extracted from the data. This sleight of hand allows an essentially unsupervised classifier to compensate for scoring uncertainty and for partial or incomplete scores in the training data. Three problem scenarios were explored using this framework:

1. In which only some states are scored by the human rater: Here the quasi-supervised model recognizes that the system may have more states than identified by the scorer. By augmenting samples of the scored states with unique tags, the classifier identified scored states with accuracy comparable to a completely supervised classifier but still distinguished unscored states in the manner of an unsupervised classifier. Consequently, overall performance on in-sample and out-of-sample data is somewhere between these extremes.

2. In which all vigilance states are scored, but not all of the epochs: In this scenario, the rater is uncertain of the prevailing state during some periods of the recording. We make the reasonable assumption that this is most likely during transitions between states and do not use those scores in the modeling step. The results demonstrate that the quasi-supervised classifier was able to fill in the blanks with reasonable accuracy, sometimes as well as the supervised classifier.

One question that might arise is whether a quasi-supervised method is really needed for addressing Problem 2. Since training samples are available for all the vigilance states, it

seems that a completely supervised classifier of any sort could be trained to predict sleep scores. This is true, but only for "static" classifiers such as LDA, which model individual observations and not sequential data. For an incomplete state sequence, a supervised HMM cannot be constructed without additional considerations. The quasi-supervised algorithm proposed here allows us to proceed using an E-M algorithm for unsupervised model estimation by augmenting observations from scored and unscored epochs with distinctive tags that reflect the rater's opinion when available.

3. In which multiple raters score all the epochs and states, but sometimes disagree: Since only one professional scoring was available for the analyzed data, we generated surrogate hypnograms from the available ones to simulate the scenario in which raters disagree one-third (33%) of time. Then a GMM/HMM was constructed using the quasi-supervised algorithm from the incomplete hypnograms in which scored epochs represent the putative consensus between multiple human raters. As was seen in Problem 2, the algorithm performed reasonably well in completing the scores.

We have treated Problems 1, 2, and 3 in isolation, but they could co-occur in a given scenario: for instance, multiple raters partially score each hypnogram based on their certainty/uncertainty with respect to some states/epochs, but with some level of disagreement. Although this composite scenario certainly merits discussion, a rigorous analysis would be more useful when two or more independent raters are actually available (rather than the simulation of consensus hypnograms that we have used in Problem 3).

In conclusion, we have described a framework for quasi-supervised classification that may prove useful for clinical sleep scoring and also for investigating the properties of vigilance dynamics through polysomnographic recordings. The proposed method is flexible enough to accommodate different situations in which scoring uncertainty occurs and computer assistance is desirable. There are some limitations in the method as presented at this time: First, since the classifier is constructed around an unsupervised learning algorithm, states that are not previously labeled by the human rater must still be identified with known vigilance states (or sub-states thereof). Here we have completed that assignment by finding the best matching state within the complete hypnogram, which is not feasible in practice. For instance, in Problem 1 the rater may identify only N, R, and

W, but not stages of *N*. We have fitted the incompletely scored data to a five-state model on the assumption that the two excess states will emerge from *N* as a product of the E-M algorithm. While this was always the case in the recordings analyzed here, it need not always be so. Consider a sample from a different cohort—for instance a more elderly one—in which deep sleep (N3) is absent or poorly represented [40]. A five state model of this data may have support for *N1* and *N2*, but the remaining state may be carved out of the distribution of *X* under *R* or *W* rather than *N3*. More investigation is necessary for defining objective criteria for labeling model states that are better aligned with human-recognized vigilance states. A graphical analysis of the linkage between the states on the basis of the ordering of common spectral measures (e.g., delta/theta power ratio, EMG amplitude) may help resolve this problem.

Secondly, while the algorithm appears to match the rater's opinion for those states that were scored in the training data, the remaining states that are identified must still appeal to the end user by some yardstick. This is not a straightforward concern to address. We speculate however that the use of quasi-supervised classifiers could, over time, help resolve discrepancies between data-driven definitions and human perceptions of vigilance. The framework proposed here for sleep scoring provides a fresh perspective on human-computer interaction that calls for further investigation.

Finally, although the quasi-supervised algorithm was applied here to data from healthy subjects, the methods do not rely on the assumption of normal sleep patterns. They are likely to apply to disordered sleep as well—for instance, the algorithm performed equally well on the ST database, in which patients reported mild difficulty falling asleep. Performance on other conditions in which sleep quality is compromised, such as in epilepsy or REM sleep behavior disorder, remains to be seen and is deferred to a future investigation.

References

1. Addison C, Jenkins B, White M, LaVigne DA. Sleep duration and mortality risk. Sleep 2014;37(8):1279-1280.

2. Suglia SF, Kara S, Robinson WR. Sleep duration and obesity among adolescents transitioning to adulthood: do results differ by sex? J Pediatr. 2014. doi: 10.1016/j.jpeds.2014.06.052. [Epub ahead of print]

3. Thorpy MJ. Classification of sleep disorders. Neurotherapeutics 2012;9(4):687-701.

4. Eriksson SH. Epilepsy and sleep. Curr Opin Neurol. 2011;24(2):171-6.

5. Cochen de Cock V. Recent data on rapid eye movement sleep behavior disorder in patients with Parkinson disease: analysis of behaviors, movements, and periodic limb movements. Sleep Med. 2013;14(8):749-53.

6. Gerstner JR, Perron IJ, Pack AI. The nexus of Aβ, aging, and sleep. Sci Transl Med. 2012;4(150):150fs34.

7. Iber C, Ancoli-Israel S, Chesson A, and Quan SF. The AASM manual for the scoring of sleep and associated events. American Academy of Sleep Medicine, 2007.

8. Zhang X, Dong X, Kantelhardt JW, Li J, Zhao L, Garcia C, Glos M, Penzel T, Han F. Process and outcome for international reliability in sleep scoring. Sleep Breath. 2014 May 7. [Epub ahead of print]

9. Haustein W, Pilcher J, Klink J, Schulz H. Automatic analysis overcomes limitations of sleep scoring. Electroenceph Clin Neurophysiol. 1986;64:364-74.

10. Flexer A, Grubera G, Dorffner G. A reliable probabilistic sleep stager based on a single EEG signal. Artif Intell Med. 2005;33:199-207.

11. Koupparis AM, Kokkinos V, Kostopoulos GK. Semi-automatic sleep EEG scoring based on the hypnospectrogram. J Neurosci Methods 2014;221:189-95.

12. Sinha RK. Artificial neural network and wavelet based automated detection of sleep spindles, REM sleep and wake states. J Med Syst. 2008;32:291-9.

13. Chapotot F, Becq G. Automated sleep-wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules. Int J Adapt Control Signal Process. 2010;24:409-23.

14. Flury B. A First Course in Multivariate Statistics. New York: Springer. 1997.

15. Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77(2):257-86.

16. Kemp B, Zwinderman AH, Tuk B, Kamphuisen HAC, Oberyé JJL. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE Trans Biomed Eng. 2000;47(9):1185-94.

17. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101:215-20.

18. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B Stat Methodol. 1977;39(1):1-38.

19. Bilmes JA. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. International Computer Science Institute TR-97-021, April 1998.

20. Doroshenkov LG, Konyshev VA, Selishchev SV. Classification of Human Sleep Stages Based on EEG Processing Using Hidden Markov Models. Biomed Eng 2006;41:25-28.

21. Pan S, Kuo C, Zeng J, Liang S. A transition-constrained discrete hidden Markov model for automatic sleep staging. BioMedical Engineering OnLine 2012;11:52.

22. Langrock R, Swihart BJ, Caffo BS, Punjabi NM, Crainiceanu CM. Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms. Statist Med. 2013;32:3342-56.

23. Fraser AM. Hidden Markov models and dynamical systems. SIAM, 2008.

24. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20: 37-46.

25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977:159-74.

26. Frost JD. An automatic sleep analyzer. Electroenceph Clin Neurophysiol. 1970;29:85-92.

27. Smith JR, Karacan I. EEG sleep stage scoring by an automatic hybrid system. Electroenceph Clin Neurophysiol. 1971;31:231-7.

28. Zung WW, Naylor TH, Gianturco DT, Wilson WP. Computer simulation of sleep EEG patterns with a Markov chain model. Recent Adv Biol Psychiatry 1965;8:335-55.

29. Kim JW, Lee JS, Robinson PA, Jeong DU. Markov analysis of sleep dynamics. Phys Rev Lett. 2009; 102(17):178104.

30. Bizzotto R, Zamuner S, De Nicalao G, Karlsson MO, Gomeni R. Multinomial logistic estimation of Markov-chain models for modeling sleep architecture in primary insomnia patients. J Pharmacokinet Pharmacodyn. 2010;37(2):137-55.

31. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. UCLA, 1968.

32. Silber MH. Staging sleep. Sleep Medicine Clinics 2012;7(3):487-96.

33. Magalang UJ, Chen NH, Cistulli PA, Fedson AC, Gislason T, Hillman D, Penzel T, Tamisier R, Tufik S, Phillips G, Pack AI; SAGIC Investigators. Agreement in the scoring of respiratory events and sleep among international sleep centers. Sleep 2013;36(4):591-6.

34. Larsen LE and Walter DO. On automatic methods of sleep staging by EEG spectra. Electroenceph Clin Neurophysiol. 1970;28:459-67.

35. Itil TM, Shapiro DM, Fink M, Kassebaum D. Digital computer classifications of EEG sleep stages. Electroenceph Clin Neurophysiol. 1969;27:76-83.

36. Fraiwan L, Lweesy K, Khasawneh N, Fraiwan M, Wenz H, Dickhaus H. Classification of sleep stages using multi-wavelet time frequency entropy and LDA. Method Inform Med. 2010;49(3):230-7.

37. Tagluk ME, Sezgin N, Akin M. Estimation of sleep stages by an artificial neural network employing EEG, EMG and EOG. J Med Syst. 2010;34:717-25.

38. Khalighi S, Sousa T, Pires G, Nunes U. Automatic Sleep Staging: A Computer Assisted Approach for Optimal Combination of Features and Polysomnographic Channels. Expert Syst Appl. 2013;40:7046–59.

39. Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H. Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. Comput Meth Prog Bio. 2012;108(1):10-19.

40. Rajput V, Bromley SM. Chronic insomnia: a practical review. Am Fam Physician 1999;60(5):1431-8.

41. Chang, B. S., D. L. Schomer, and E. Niedermeyer. "Normal EEG and sleep: adults and elderly." Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, ed 6 (2010): 183-214.

REFERENCES

Abbas A., Yaghouby F., Schildt C., Ajwad A., Donohue KD., O'Hara BF. "Effect of Ambient Temperature on Sleep in Mice." 31st Annual BGSFN Spring Neuroscience Research Day, Lexington, KY, Mar 2015.

Badawy R.A ,Freestone D.R, Lai A., Cook M. J. "Epilepsy: ever-changing states of cortical excitability". Neuroscience 222 (2012) 89–99

Baldy-Moulinier M. "Inter-relationships between sleep and epilepsy". Recent advances in epilepsy, 3. Edinburgh: Churchill Livingstone, 1986. pp. 37±55.

Bastianini, S., Berteotti, C., Gabrielli, A., Del Vecchio, F., Amici, R., Alexandre, C., ... & Zoccoli, G. SCOPRISM: A new algorithm for automatic sleep scoring in mice. Journal of neuroscience methods, 235, 277-284. (2014).

Bazil C.W, Castro L.H, Walczak T.S." Reduction of rapid eye movement sleep by diurnal and nocturnal seizures in temporal lobe epilepsy". Arch Neurol. 2000;57(3):363-368

Bazil, C. W., & Walczak, T. S. (1997). Effects of sleep and sleep stage on epileptic and nonepileptic seizures. Epilepsia, 38(1), 56-62.

Bizzotto R, Zamuner S, De Nicalao G, Karlsson MO, Gomeni R. Multinomial logistic estimation of Markov-chain models for modeling sleep architecture in primary insomnia patients. J Pharmacokinet Pharmacodyn. 2010;37(2):137-55.

Cavalheiro, E. A., Santos, N. F., & Priel, M. R. (1996). The pilocarpine model of epilepsy in mice. Epilepsia, 37(10), 1015-1019.

Crespel A, Coubes P, Baldy-Moulinier M. "Sleep influence on seizures and epilepsy effects on sleep in partial frontal and temporal lobe epilepsies". Clinical Neurophysiology 111, Suppl. 2 (2000) S54±S59

Curia G, Longo D, Biagini G, Jones RS, Avoli M. The pilocarpine model of temporal lobe epilepsy. J Neurosci Methods 2008;172(2):143–57.

Dingledine R.,Engel J.,"Epilepsy Research Benchmarks Progress Report ", 2007.

Donohue KD, Medonza DC, Crane ER, O'Hara BF. Assessment of non-invasive high-throughput classifier for behaviors associated with sleep and wake in mice. Biomed Eng Online. 11;7-14. 2008

Doroshenkov LG, Konyshev VA, Selishchev SV. Classification of Human Sleep Stages Based on EEG Processing Using Hidden Markov Models. Biomed Eng 2006;41:25-28.

Endo T., Schewierin B.,Borbely A., Tobler I. Selective and total sleep deprivation: effect on the sleep EEG in the rat. Psychiatry Research 66 (1997) 97 110

Esteller, R., Echauz, J., Tcheng, T., Litt, B., & Pless, B. (2001). Line length: an efficient feature for seizure onset detection. In Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE (Vol. 2, pp. 1707-1710). IEEE.

Fisher NI 1993. Statistical Analysis of Circular Data. Cambridge University Press.

Flexer A, Grubera G, Dorffner G. A reliable probabilistic sleep stager based on a single EEG signal. Artif Intell Med. 2005;33:199-207.

Flores AE, Flores JE, Deshpande H, Picazo JA, Xie XS, Franken P, Heller HC, Grahn DA, O'Hara BF 2007. Pattern recognition of sleep in rodents using piezoelectric signals generated by gross body movements. IEEE Trans Biomed Eng. 54(2):225-33.

Friedman L, Haines A, Klann K, Gallaugher L, Salibra L, Han F, Strohl KP 2004. Ventilatory behavior during sleep among A/J and C57BL/6J mouse strains. J Appl Physiol. 97(5):1787-95.

Frost JD. An automatic sleep analyzer. Electroenceph Clin Neurophysiol. 1970;29:85-92. Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: a review. Journal of Experimental Psychology: Applied, 6(3), 236.

Herman ST, Walczak TS, Bazil CW. "Distribution of partial seizures during the sleep--wake cycle: differences by seizure onset site". Neurology (2011).

Holland, K., "Epilepsy by the Numbers: Facts, Statistics, and You." *Healthline*,2014.

Huffman D., Hargis K., Yaghouby F., Blalock E., Sunderam S. "Feasibility of Selective Deep Sleep Restriction in Rats Using Mild Somatosensory Stimulation." 31st Annual BGSFN Spring Neuroscience Research Day, Lexington, KY, Mar 2015

Kaiser JF 1990. On a simple algorithm to calculate the 'energy' of a signal. ICASSP 381-384.

Kim JW, Lee JS, Robinson PA, Jeong DU. Markov analysis of sleep dynamics. Phys Rev Lett. 2009; 102(17):178104.

Kushida C.A., Bergmann B.M. and Rechtschaffen A., "Sleep deprivation in the rat: IV. Paradoxical sleep deprivation," Sleep, vol. 12, pp. 22-30, 1989.

Langrock R, Swihart BJ, Caffo BS, Punjabi NM, Crainiceanu CM. Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms. Statist Med. 2013;32:3342-56.

Libourel, P. A., Corneyllie, A., Luppi, P. H., Chouvet, G., & Gervasoni, D. Unsupervised online classifier in sleep scoring for sleep deprivation studies. Sleep. (2014).

Long, X., Yang, J., Weysen, T., Haakma, R., Foussier, J., Fonseca, P., & Aarts, R. M. (2014). Measuring dissimilarity between respiratory effort signals based on uniform

scaling for sleep staging. Physiological measurement, 35(12), 2529.

Louis, R. P., Lee, J., & Stephenson, R.. Design and validation of a computer-based sleep-scoring algorithm. Journal of neuroscience methods, 133(1), 71-80. (2004)

Mallick B. Singh S., Singh A. Mechanism of noradrenaline-induced stimulation of Na–K ATPase activity in the rat brain: implications on REM sleep deprivation-induced increase in brain excitability. Mol Cell Biochem (2010) 336:3–16

Mang, G. M., Nicod, J., Emmenegger, Y., Donohue, K. D., O'Hara, B. F., & Franken, P. (2013). Evaluation of a piezoelectric system as an alternative to electroencephalogram/electromyogram recordings in mouse sleep studies. Sleep, 37(8), 1383-1392.

Matos G, Andersen ML, Tufik S." The relationship between sleep and epilepsy: Evidence from clinical trials and animal models". Journal of the Neurological Sciences 295 (2010b) 1–7

McCoy J.G and Strecker R.E , The cognitive cost of sleep lost, Neurobiology of Learning and Memory 96 (2011) 564–582

McShane, B. B., Jensen, S. T., Pack, A. I., & Wyner, A. J.. Statistical learning with time series dependence: an application to scoring sleep in mice. Journal of the American Statistical Association, 108(504), 1147-1162. (2013)

Mendelson W.B., Guthrie R.D., Frederick G., Wyatt R.J., "The flower pot technique of rapid eye movement (REM) sleep deprivation," Pharmacol Biochem Behav, vol. 2, pp. 553-556, 1974.

Mistlberger RE, Belcourt J, Antle MC. Circadian clock resetting by sleep deprivation without exercise in Syrian hamsters: dark pulses revisited. J Biol Rhythms. 2002 Jun;17(3):227-37

Niedermeyer E. Petit mal," primary generalized epilepsy and sleep". Sleep and epilepsy, New York: Academic Press, 1982. pp. 191±208.

Orff, Henry J., Liat Ayalon, and Sean PA Drummond. "Traumatic brain injury and sleep disturbance: a review of current research." The Journal of head trauma rehabilitation 24.3 (2009): 155-165.

Ortinski, P., & Meador, K. J. (2004). Cognitive side effects of antiepileptic drugs. *Epilepsy & Behavior*, 5, 60-65.

Pack Allan I. "Changes in Respiratory Motor Activity During Rapid Eye Movement Sleep". Regulation of Breathing , Ed. Jerome A. Dempsey Ed. Allan I. Pack. Marcel Dekkler, 1995.

Pan S, Kuo C, Zeng J, Liang S. A transition-constrained discrete hidden Markov model for automatic sleep staging. BioMedical Engineering OnLine 2012;11:52.

Pitkanen A, Sutula TP. Is epilepsy a progressive disorder? Prospects for new therapeutic approaches in temporal-lobe epilepsy. Lancet Neurol 2002;1:173–81.

Placidi F, Zannino S, Albanese M, Romigi A, Izzi F, Marciani MG, Palmieri MG. Increased cortical excitability after selective REM sleep deprivation in healthy humans: a transcranial magnetic stimulation study. Sleep Med. 2013 Mar;14(3):288-92.

Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77(2):257-86.

Racine R. J. (1979). Modification of seizure activity by electrical stimulation: II. Motor seizures. Electroencephalogr. Clin. Neurophysiol. 32 281–294 10.1016/0013-4694(72)90177-0

Rechtschaffen, A., Hauri, P., & Zeitlin, M. (1966). Auditory awakening thresholds in REM and NREM sleep stages. Perceptual and motor skills, 22(3), 927-942.

Rytkönen, K. M., Zitting, J., & Porkka-Heiskanen, T. Automated sleep scoring in rats and mice using the naive Bayes classifier. Journal of neuroscience methods, 202(1), 60-64. (2011)

Sahu S., Kauser, Ray K., Kishore, S. Kumar, U. Panjwani, ” Caffeine and modafinil promote adult neuronal cell proliferation during 48 h of total sleep deprivation in rat dentate gyrus,” Exp Neurol, vol. 248, pp. 470-481, 2013.

Sato, M., Sagawa, Y., Hirai, N., Sato, S., Okuro, M., Kumar, S., ... & Nishino, S. (2014). Noninvasive detection of sleep/wake changes and cataplexy-like behaviors in orexin/ataxin-3 transgenic narcoleptic mice across the disease onset. Experimental neurology, 261, 744-751.

Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Timmer, J., & Schulze-Bonhage, A. (2006). Do False Predictions of Seizures Depend on the State of Vigilance? A Report from Two Seizure-Prediction Methods and Proposed Remedies. Epilepsia, 47(12), 2058-2070.

Schwierin B., Borbély A., Tobler I., ”Prolonged effects of 24-h total sleep deprivation on sleep and sleep EEG in the rat,” Neurosci Lett, vol.261, pp. 61–64, 1999.

Shibley, H., & Smith, B. N. (2002). Pilocarpine-induced status epilepticus results in mossy fiber sprouting and spontaneous seizures in C57BL/6 and CD-1 mice. Epilepsy research, 49(2), 109-120.

Sunagawa, G. A., Séi, H., Shimba, S., Urade, Y., & Ueda, H. R. FASTER: an unsupervised fully automated sleep staging method for mice. Genes to Cells, 18(6), 502-518. (2013).

Swanson, L. M., Arnedt, J., Rosekind, M. R., Belenky, G., Balkin, T. J., & Drake, C. (2011). Sleep disorders and work performance: findings from the 2008 National Sleep Foundation Sleep in America poll. Journal of sleep research, 20(3), 487-494.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. Proceedings of the IEEE, 70(9), 1055-1096.

Wisor J.P., Clegern W.C., Schmidt M.A., "Toll-Like Receptor 4 is a Regulator of Monocyte and electroencephalographic Responses to Sleep Loss, " Sleep, vol. 34, pp. 1335-1345, 2011.

Yaghouby F, Salmon E., Donohue K., O'Hara B., Sunderam S. "Noninvasive Scoring of Mouse Sleep and Behavior Using a Piezoelectric Motion Sensor." Society for Neuroscience Annual Meeting, Washington, DC, Nov 2011.

Yaghouby F, Schildt C, Donohue KD, O'Hara BF, Sunderam S. Development of a Technique for Assessing the Effects Selective Sleep Restriction in Mice. American Epilepsy Society 68th Annual Meeting, Seattle, WA, Dec 2014c

Yaghouby F., Schildt C., Ajwad A., Donohue KD., O'Hara BF., Sunderam S. "Determination of Vigilance State in Mice Using a Noninvasive Motion Sensor." 31st Annual CCTS/BGSFN Spring Neuroscience Research Day, Lexington, KY, 2015b.

Yaghouby F., Zhang T., Schildt C., Striz M., Donohue K., O'Hara B., Sunderam S. "Markov Modeling of Sleep-Wake Dynamics Following Acute Neural Injury." American Epilepsy Society 67th Annual Meeting, Washington, DC, Dec 2013.

Yaghouby F., Zhang T., Striz M., Donohue K., O'Hara B., Sunderam S. "Markov modeling of sleep-wake dynamics for applications inneurotherapy." Society for Neuroscience meeting (SFN 2012), New Orleans, LA, Oct 2012

Yaghouby, F., & Sunderam, S. Quasi-supervised scoring of human sleep in polysomnograms using augmented input variables. Computers in biology and medicine, 59, 54-63, 2015a.

Yaghouby, F., Modur, P., & Sunderam, S. Naive scoring of human sleep based on a hidden Markov model of the electroencephalogram. In Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE (pp. 5028-5031). 2014a.

Yaghouby, F., Schildt, C. J., Donohue, K. D., O'Hara, B. F., & Sunderam, S. Validation of a closed-loop sensory stimulation technique for selective sleep restriction in mice. In Engineering in Medicine and Biology Society (EMBC), (pp. 3771-3774),2014b

Zung WW, Naylor TH, Gianturco DT, Wilson WP. Computer simulation of sleep EEG patterns with a Markov chain model. Recent Adv Biol Psychiatry 1965;8:335-55.

VITA

Author's Name: Farid Yaghouby

Education: M.S. in Bioelectrical Engineering, Iran University of Science and Technology, Iran 2009.

Publications:

1. Yaghouby F, Sunderam S. "Quasi-Supervised Scoring of Human Sleep in Polysomnograms Using Augmented Input variables". Computers in Biology and Medicine, Elsevier, 59 (1): 54-63 , 2015.

2. Yaghouby F, Madahian B, Mirinejad H, Sunderam S. "Feasibility of Seizure Risk Prediction Using Intracranial EEG Measurements in Dogs".41st Annual Northeast Bioengineering Conference, Rensselaer, NY, Apr 2015.

3. Yaghouby F, Modur P, Sunderam S. "Naive scoring of human sleep based on a hidden Markov model of the electroencephalogram". IEEE Engineering in Medicine and Biology Society meeting , Chicago, IL, Aug 2014.

4. Yaghouby F, Schildt C, Donohue KD, O'Hara BF, Sunderam S. "Validation of a closed-loop sensory stimulation technique for selective sleep restriction in mice". IEEE Engineering in Medicine and Biology Society meeting, Chicago, IL, Aug 2014.

5. Yaghouby F., Ayatollahi A., Yaghouby M., Alavi A.H. "Towards Automatic Detection of Atrial Fibrillation: A Hybrid Computational Approach. "Computers in Biology and Medicine, Elsevier, 40(11-12):919-930, 2010.

6. Yaghouby F., Ayatollahi A., Yaghouby M., Alavi A.H., "Robust Genetic Programming-Based Detection of Atrial Fibrillation Using RR Intervals." Expert Systems, Wiley-Blackwell, DOI: 10.1111/j.1468-0394, 2010.