ANALYSIS OF HUMAN POLYSOMNOGRAPHY (PSG) FOR AUTOMATIC SLEEP

EVENT DETECTION USING HIDDEN MARKOV MODEL

by

Alakh Sudhir Biniwale, MS

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Computer Science
August 2016

Committee Members:

Vangelis Metsis, Chair

Mina Guirguis

Byron Gao

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

The Copyright Laws of the United States (Public Law-553, section 107) protects this work. Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgment. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work, I, Alakh Sudhir Biniwale, authorize duplication of this work, in completely or in part, for educational or scholarly purposes only.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Vangelis Metsis, for his patience, support, and encouragement throughout my graduate studies. He always suggested what was good for my career and me. He has been an adviser who listens to problems and gives solutions, no matter how small or big they were. This thesis would not have been possible without him. He was very passionate in every step of the thesis. I would like to thank especially him for his relentless support and guidance.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

ABSTRACT

This thesis work evaluates our proposed methodology for automated detection of sleep events from Polysomnographic (PSG) data. The sleep data was collected during real sleep studies using Profusion PSG3. The event detection tasks used a Hidden Markov Model (HMM) to achieve signal classification for sleep event detection. The Hilbert transform (envelope) was used to extract features for input to the HMM. HMM was selected as our classification method of choice, due to the fact that it was able to capture the temporal variations of the biosignals collected through PSG. In this work, we detected sleep motion events, such as rapid eye movements (REM) and leg movements, and breathing events like obstructive apnea, hypopnea and snore. The task of detecting events of interest was achieved using a sliding window approach, and classifying each signal segment as containing an event or not, hence, leading to a binary classification problem for each type of event. Our experimental results show that our proposed approach can be successfully used for sleep event detection, to assist experts in sleep quality assessment, however, the big imbalance between the number of segments that contain a positive event and the ones that do not, often negatively affects the performance of our classification method.

I. INTRODUCTION

The American Academy of Sleep Medicine estimates that 22 million Americans suffer from sleep disorders, the vast majority of which remain undiagnosed due to inconvenience and high cost associated with sleep studies using PSG. With an estimated 25-30% of the general adult population and a comparable percentage of adolescents and children in the US experiencing fragmented and inefficient sleep, the incidence of disability, morbidity, and mortality is on the rise (NHLBI National Center on Sleep Disorders Research, 2011). Most of the adults suffer from most common sleep disorders such as insomnia, REM behavior disorder, restless legs syndrome (RLS) or periodic limb movements in sleep (PLMS), and obstructive sleep apnea. Sleep medicine is still at its early stages of development, and new technologies promise to enable new research and further understanding of sleep disorders and their connection with other medical conditions.

PSG is considered the diagnostic standard for diagnosis of sleep disorders; there are drawbacks to its use. The PSG is uncomfortable for the patient and involves a considerable investment for the healthcare system requiring equipment, bed space and specialized technical support (Baraglia et al., 2005). Sleep disorder detection is more complex and tedious because of the involvement of multiple bio-signals such as six Electroencephalograms (EEG), four Electromyograms (EMG), two Electrooculograms (EOG), Oxymeters (SpO2), etc. Moreover, sleep disorder detection is highly dependent on interpretation of data (event detection), which is quite inconvenient process requiring a qualified sleep technician and significant amount of manual analysis. Thus, automation of task involved in sleep studies is highly desirable goal. The advantages of an automated system for diagnosis include speed, reliability, economic saving and improved reliability of diagnosis (Baraglia et al., 2005). In this thesis, we tried to detect sleep events employing sequence-labeling algorithm Hidden Markov Model (HMM). We used the envelope of a signal as a feature for our analysis, which goes through segmentation process and is converted into sequences of observations. We attempted to classify sleep events: rapid eye movements, limb movements, and breathing events, like obstructive apnea, hypopnea and snoring.

Sleep events classification was achieved with an average accuracy of around 80% with some events being accurately scored and others partially scored. The Limb movements and rapid eye movements were detected with 81% and 68% accuracy respectively. The two breathing events hypopnea and apnea were scored with closely varying degrees of accuracy with the highest scores being around 78% and 75%. While snore event detection was more precise with the precision of 95% and accuracy of around 88%.

The main contribution of this thesis is to develop and evaluate a methodology that can be used to detect a variety of sleep events associated with different human bio signals. The proposed method is independent of the specific hardware used to collect those signals, and it is robust to noise, artifacts, and sensor setup variations.

II. BACKGROUND

Before we head towards implementation and methodology, let us first have a look at how Polysomnography (PSG) is used for detection of sleep disorders. We will be further discussing related work which inspired this thesis.

Understanding Polysomnography (PSG)

PSG is a multi-parametric test used in the study of sleep as a diagnostic tool in sleep medicine. PSG records brain waves, the oxygen level in blood, heart rate and breathing, as well as eye and leg movements during the study. PSG capturers EEG, EMG and EOG data. In addition, other signals related to respirations are also captured. These signals are time series waveforms, and detection of a disorder can be associated with the occurrence of an event in a particular signal. Let us have a look on how disorders can be visually seen in a signal as an event.

1. Limb movements can be detected by observing the EMG signal, produced by muscle activity, from electrodes placed on the right and left legs. Figure 1: shows leg electrode placement for capturing the EMG signal, and a sample visualization of a signal snapshot.

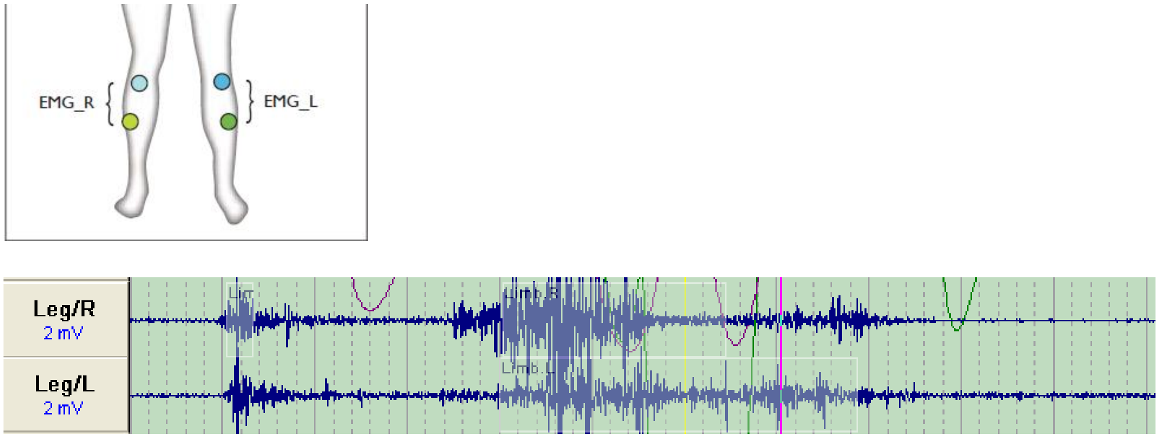

Figure 1: An event (muscle activity) of leg movement observed at EMG signals.

2. Rapid eye movements can be detected by observing the EOG signal captured from muscle activity, using electrodes placed on the upper and lower parts of someone's eyes. Figure 2: shows eye electrode placement for capturing the EOG signal, and a snapshot of a signal sample.

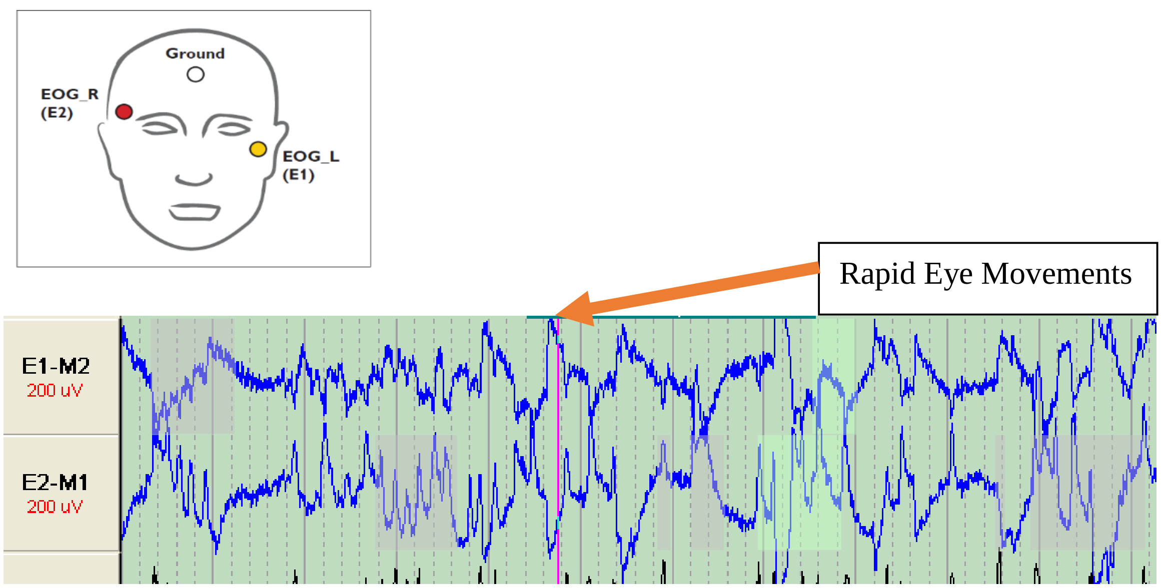

Figure 2: Rapid eye movements observed in EOG signal of both eyes.

3. Obstructive sleep apnea and Hypopnea are repertory events, In order to detect these, the following three signals are observed:

a) Continuous Positive Airway Pressure (CPAP Flow),

b) Thoracic (chest) belt pressure

c) Abdomen belt pressure

Figure 3 shows obstructive apnea detected at CPAP flow and hypopnea detected at chest and abdomen belt. You can observe that there is an abrupt change in signals.

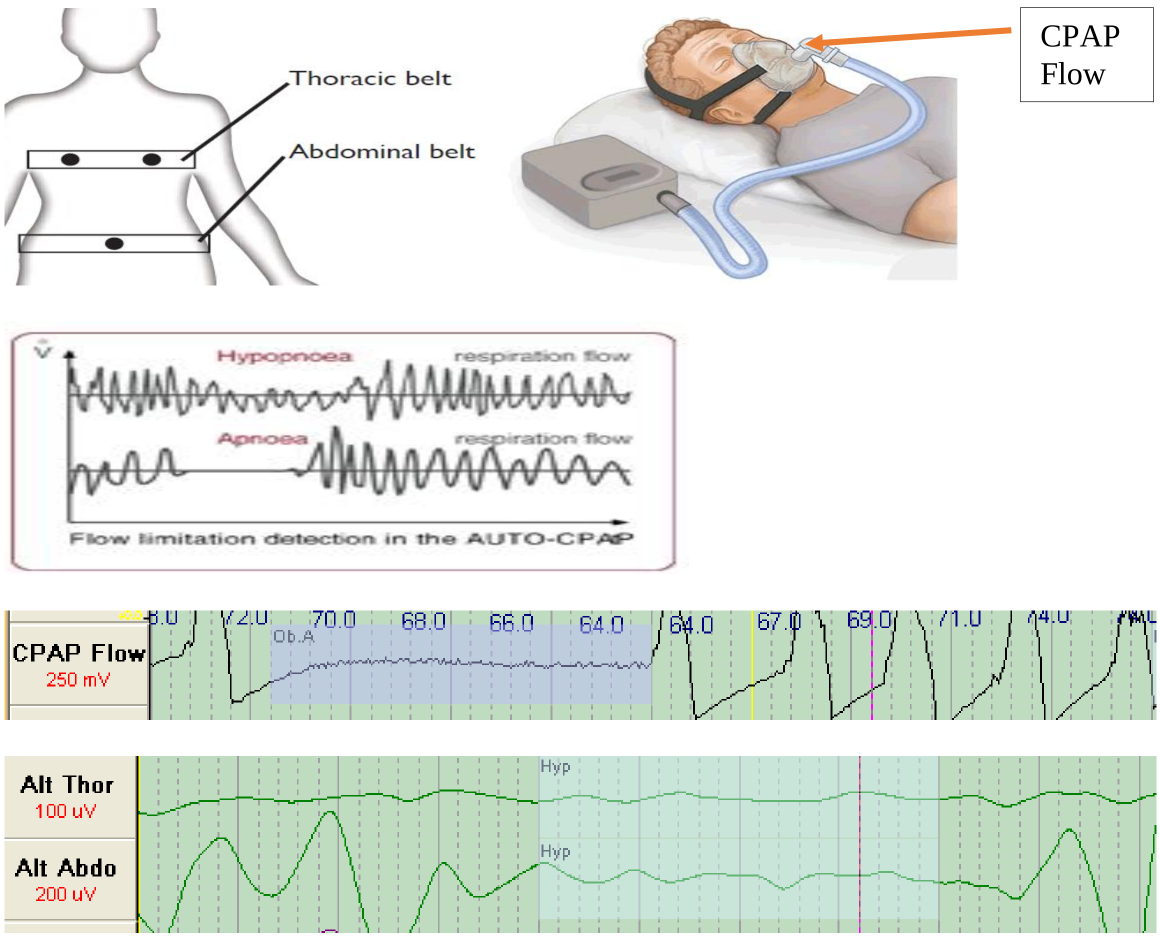

Figure 3: Observation that CPAP flow, chest and abdomen belt exhibits change in a signal waveform.

4. Snoring is the low frequency sound produced by vibration of the upper airway during sleep. A Microphone can simply record the sound produced when patient is snoring shown in Figure 4.

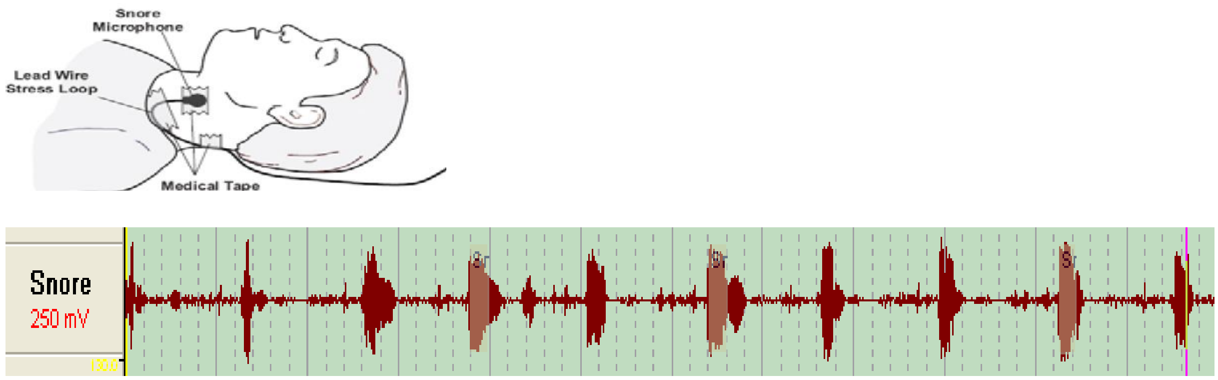

Figure 4: Shows sound wave as intervals if snores in sleep study.

<u>Why HMM</u>

Hidden Markov models have been used for at least three decades in signal-processing applications, especially in the context of automatic speech recognition (Rabiner and Juang, 1986), but interest in their theory and application has expanded to other fields, e.g.:

• All kinds of recognition: face, gesture, handwriting, signature;

• Bioinformatics: biological sequence analysis;

• Environment: wind direction, rainfall, earthquakes;

• Finance: series of daily returns;

• Biophysics: ion channel modelling.

Attractive features of HMMs include their simplicity, their general mathematical tractability, and specifically the fact that the likelihood is relatively straightforward to compute (MacDonald and Zucchini, 2009). It is a traditional statistical tool for modeling a generative sequence, which can learn from input pairs, each consisting of a sequence of observations and a sequence of labels (Nguyen and Edu, 2004). HMM is a type of pattern matching technique, which is widely practiced especially in time series analysis. Weigend and Shi have previously applied HMMs to financial time series analysis (Shi and Weigend, 1997). There have been other fields of interest where HMMs have been utilized for computational biology, biomedical signal interpretation (Wu and Xie, 2009). Their application has also been extended to EEG, EMG and EOG for particular event detection (Huang et al., 1996). The data collected from Profusion PSG in this research are physiologic signals from overnight sleep studies, which is a time series measurement of certain physiological indicators. That makes HMM suitable for our task.

Related Work

When it comes to sleep event detection or sleep disorder detection, the majority of research works focus on the detection of sleep apnea. Other important events such as limb movement and eye movements are less considered. In (Baraglia et al., 2005), the author investigates the automated detection of the patients' breathing rate and heart rate from their skin conductivity as well as sleep stage, scoring and breathing event detection from their EEG. The sleep scoring and breathing event detection tasks used neural networks to achieve signal classification. The Fourier transform and the Higuchi fractal dimension were used to extract features for input to the neural network. Sleep stage classification was achieved with accuracy of around 65% with some stages accurately scored, and others poorly scored. The two breathing events hypopnea and apnea were scored with varying degrees of accuracy with the highest scores being around 75% and 30%. However, the skin conductivity experiment only used filtering on the skin conductivity. It is possible that more advanced signal processing techniques could produce better results.

In (Nakano et al., 2007), single-channel airflow monitors developed for screening of sleep-disordered breathing (SDB) have conflicting results for accuracy. Three hundred ninety nine polysomnography (PSG) records were employed, including a thermal sensor signal. The algorithm was designed to obtain a time series (flow-power) using power spectral analysis, which expressed fluctuation in the airflow signal amplitude. From the time series the algorithm detected transient falls of the flow-power and calculated flow-respiratory disturbance index (RDI), defined as the number of falls per hour. The diagnostic sensitivity/specificity ratios of the flow-RDI were 96/76, 88/80 and 97/77 percent, respectively. The presented results suggested that a single-channel airflow monitor could be used to detect sleep-disordered breathing automatically if the analytic algorithm was optimized. By reading above related work, we realized that CPAP flow would be a good choice to detect breathing events.

For limb movements, in (Alessandria and Provini, 2013) author mentioned that event started when the EMG amplitude exceeded 8 mV above baseline and ended when the amplitude remained below 2 mV above baseline for at least 0.5 s. The movements must be 0.5–10 s in duration. A sequence of four or more such movements during any sleep stage separated by an interval of at least 5 s and not more than 90 was considered as PLMS. For rapid eye movements, in (Kempfner et al., 2011), the authors found that REM sleep detection without the use of chin electromyography (EMG) was useful. This was addressed by analyzing the classification performance when implementing two automatic REM sleep detectors. The first detector used the electroencephalography (EEG), electrooculography (EOG) and EMG to detect REM sleep, while the second detector only used the EEG and EOG. By referring (Kempfner et al., 2011) we considered EOG for REM detection.

Other previous works which were helpful during our research include non-invasive analysis of sleep patterns via multimodal sensor input (Metsis et al., 2012) and recognition of sleep patterns using a bed pressure mat (Metsis et al., 2011). The related research focused on the importance of sleep patterns for detection and treatments of sleep disorders. The author used data collected from FSA bed pressure mat and Kinect for motion sensing. The author classified some of the basic motions and body postures such as changing body posture, moving arms or legs, getting in bed or out of bed, making bed, left side and right side. The classification of the sequences of motion frames are performed using Template Matching (TM), k- Nearest Neighbors (KNN), Support Vector Machines (SVM) and HMM. They evaluated their classification algorithms for body posture recognition and motion recognition combining pressure data (pressure mat) and depth sensing data (Kinect). Their experiment resulted showing, HMM achieved accuracy of 97.87% for motion recognition. The classification accuracy results are promising and the proposed methods could be used for detection of sleep disorders (Metsis et al., 2011).

A previous work, which attempted to detect sleep events in PSG data (Espiritu and Metsis, 2015) used EEG signals and extracted features from each segments such as power spectral density estimate peaks, energy of a discrete-time signal and zero-crossing rate. The supervised learning algorithms used for classification are Naive Bayes, Logistic Regression and Decision Trees. Results of (Espiritu and Metsis, 2015) shows that the highest accuracy obtained from leg movement event detection was achieved by the decision tree around 88.39%. However, this results were achieved using static classifier and can be used to compare with our results achieved using continues HMM in this research.

III. METHODOLOGY

We followed a supervised learning approach for detecting events of interests from PSG data. Following standard procedures, anonymous Polysomnography (PSG) data were acquired during sleep studies at the Texas State Sleep Lab, using a PSG system called Profusion, build by Compumedics. This raw data needed some pre-processing before we could use it for experimental purposes with MATLAB. The pre-processed data was then imported to MATLAB, and we selected a number of signals, such as EEG, EOG and EMG to continue our research. We produced features of these signals using the traditional signal processing techniques. These features were further reduced to a manageable size using down sampling procedures. We used the Gesture Recognition Toolkit (GRT) a cross-platform, open-source, C++ machine learning library for building our event predictor HMM model. The dataset was divided into train and test samples using a K-fold cross validation approach. Finally, we trained a HMM model with a labeled training dataset and validated the classification performance on the labeled test dataset. We generated confusion matries to understand accuracy and precision of the model. We have extensively developed MATLAB scripts for importing data, feature extraction, data transformation and performing statistical analysis. Figure 5 shows the schematic representation of the methodology.

Data Collection and Pre-processing

Our data source are prerecorded sleep studies; we selected 10 subjects' sleep study to carry out our research. As I mentioned earlier, these sleep studies were recorded using a PSG system. The Profusion PSG system consisted of a set of hardware sensors attached to human subjects during sleep studies, as well as a specialized software package, fine-tuned for these specific set of sensors. The software package provided some basic visualization of the acquired signals and could also automatically analyze the data to provide information to sleep experts, such as sleep staging and detection of sleep events related to certain sleep disorders. We installed the Profusion PSG3 to access the sleep studies and export them. Figure 6 shows a working screenshot of a PSG. Each sleep study consisted of 28 channels; we chose Leg EMG, EEG, and EOG to perform experiments.

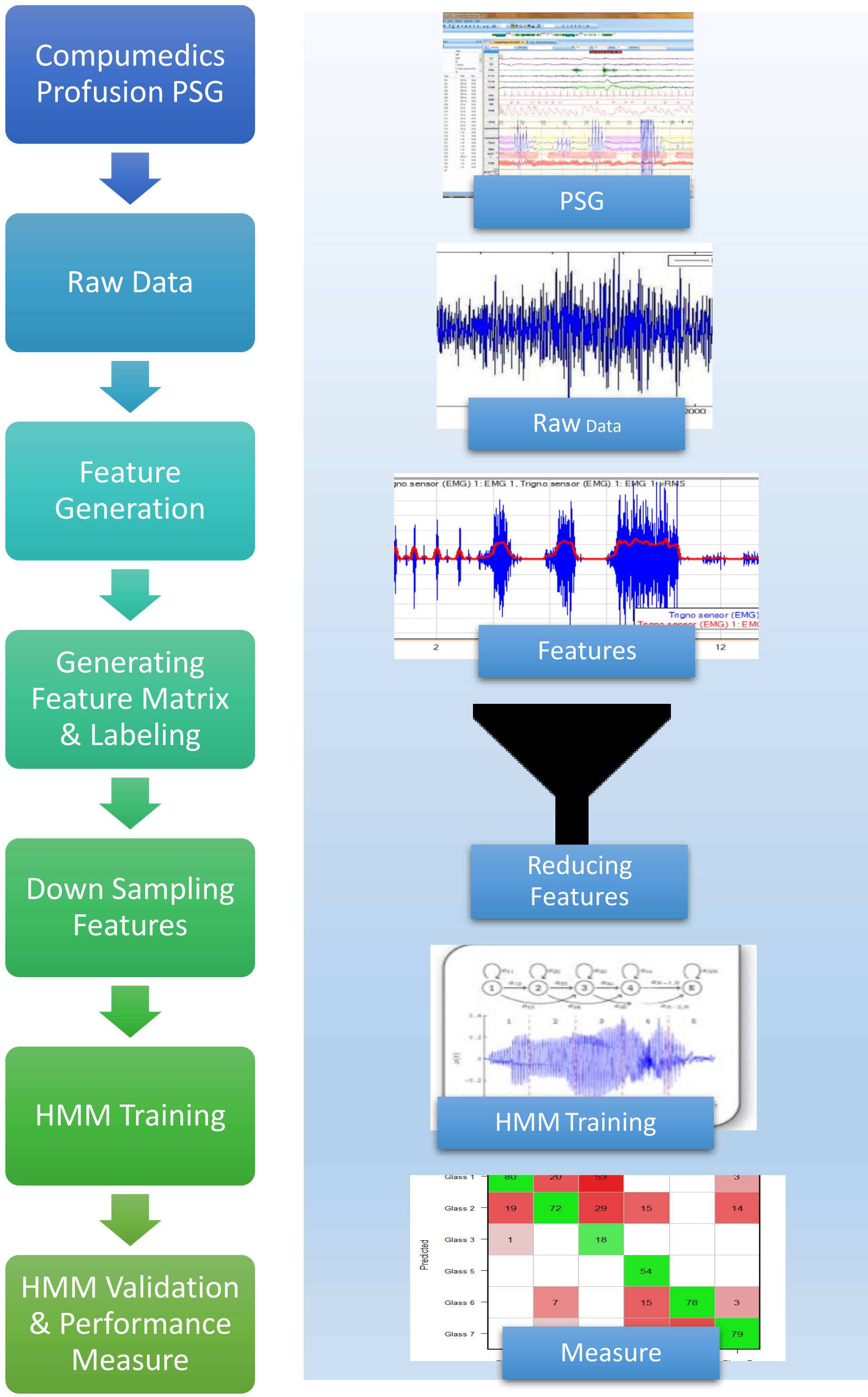

Figure 5: Schematic representation of the methodology.

A Raw data from the selected channels was exported, which was in EDF format and which could only process through an EDF browser (Freeware). After installing EDF browser, we were able to view the above mentioned signals and were able to convert them to an ASCII format. We built scripts to load data in the desired format on MATLAB. At the same time, we exported event labels for our raw data from ProFusion PSG3.

The labels provided automatically by Profusion PSG, generally do not achieve manual ground truth levels of accuracy. However due to the extremely tedious process of manually examining the entire duration of the signal for events of interest, therefore, sleep experts often trust these labels during sleep quality assessments. In the absence of manually annotated PSG datasets, we also used the labels provided by Profusion PSG as ground truth for our experiments.

Feature Generation

Sleep disorders are associated with a set of physiological events occurring during sleep time. Often, the order in which events occur can be even more important than the events themselves. Therefore, we were interested in a technique that would provide the ability to recognize sequence of events. When extracting features from a signal, our primary goal was to detect event in a signal using temporal pattern recognition (Sung and Priebe, 1988). While working with time series data, Hilbert transform (envelope) was useful in calculating instantaneous attributes of a signal (Luo, Fang, and Ertas, 2009). The instantaneous envelope is the amplitude of the complex Hilbert transform; the instantaneous frequency is the rate of change of the phase angle. These properties were applied to identify dynamic characteristics of acquired signals.

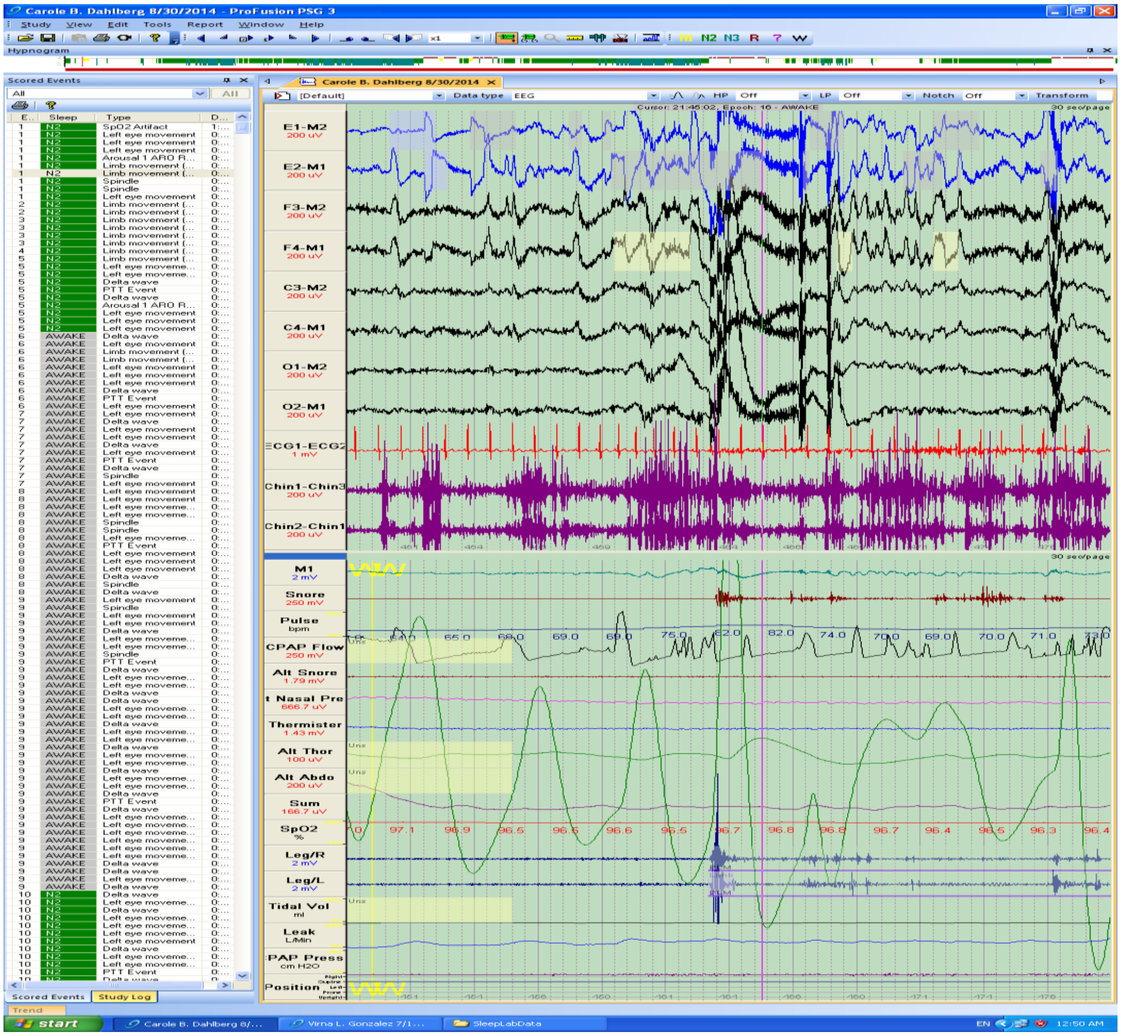

Figure 6: Working screen shot of a profusion PSG3. Right side of the column shows all 28 channels and left side is list of events on each of them.

The conventional Fourier transform, has shown the lack of accuracy for time-derivative calculation (Luo, Fang, and Ertas, 2009). (Zhang et al., 1991) provides a useful information that linear envelope was used successfully to classify the EMG signals. The envelope of a signal is the appearance of a signal in the time domain. Qualitatively, the envelope of a signal is that boundary within which the signal is contained when viewed in the time domain. This boundary has an upper and lower part. In practice, when speaking of the envelope, it is customary to consider only one of them as ‘the envelope’ (typically the upper boundary).

The classical amplitude envelope technique, root-mean square (RMS) is the most popular method for estimating the temporal evolution of the signal energy (Caetano and Rodet, 2011). It can be easily used to obtain an estimate of the amplitude envelope by simply applying it with a sliding window.

Figure 7: Shows an example of an upper envelope in a raw data EMG.

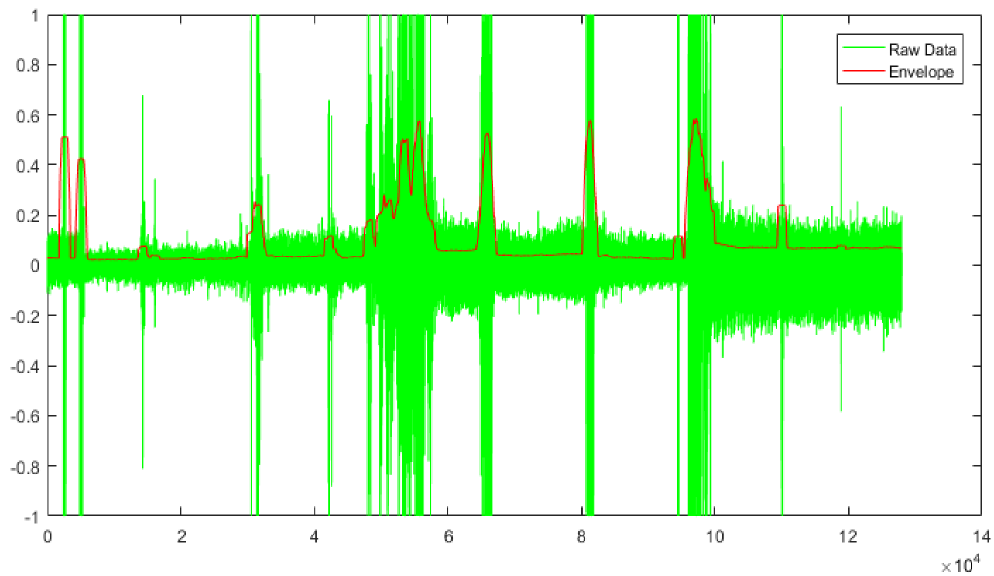

Figure 7: Green signal is the raw data leg EMG and red signal is the generated upper envelope of the signal.

Generating Feature Matrix

Once we had the envelope of a signal, we transformed it to a 2-Dimenstional matrix of sequences. In order, to get a 2D Matrix we employed a segmentation process, in which the envelope was divided into small segments of vectors. Deciding the length of a segment was a complex part. We initially started with length of 10-second window but that was not a good choice and we ended up with lots of misclassification. We were trying out with different window sizes and number of features per segments based on random choice to achieve better results, but constant misclassification persisted. After thorough study and analysis, we found out that most of the events in a signal lasted for between 0.5 seconds to 4 seconds (figure 8). Therefore, we decided to keep the segment size of 3 seconds with an overlapping of 0.25 seconds. Figure 9 Illustration of the process of feature matrix generation.

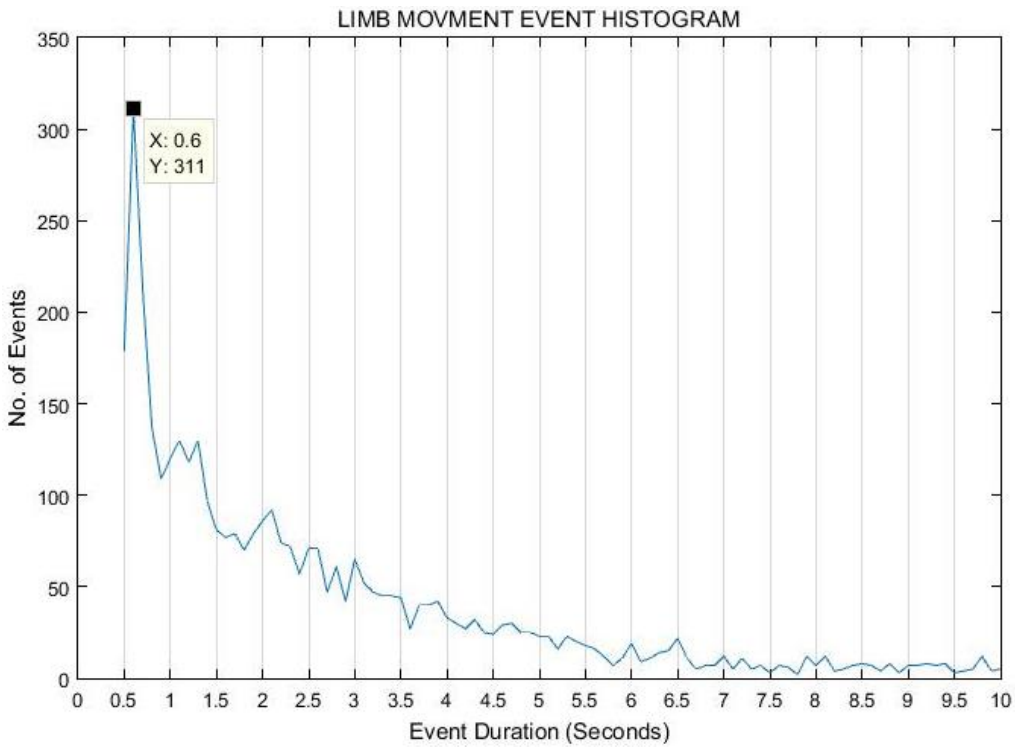

Figure 8: Histogram of a limb movement durations. X - axis represents duration in seconds and Y - axis No. of events. The maximum number of event is 311 with duration of 0.6 seconds.

Labeling Feature Matrix

After converting the envelope to a 2D matrix where each row was a single sequence and each column was a single feature, we progressed further to a labeling process. By the time we started labeling we were aware that the duration of most of the events was mostly 0.5 seconds, Figure 8 shows Limb movement events duration histogram. It was a clear indication to label a sequence with 1 (Positive) if it has an event of duration 0.5 seconds or more. the rest of the sequences were labelled as 0 (Negative).

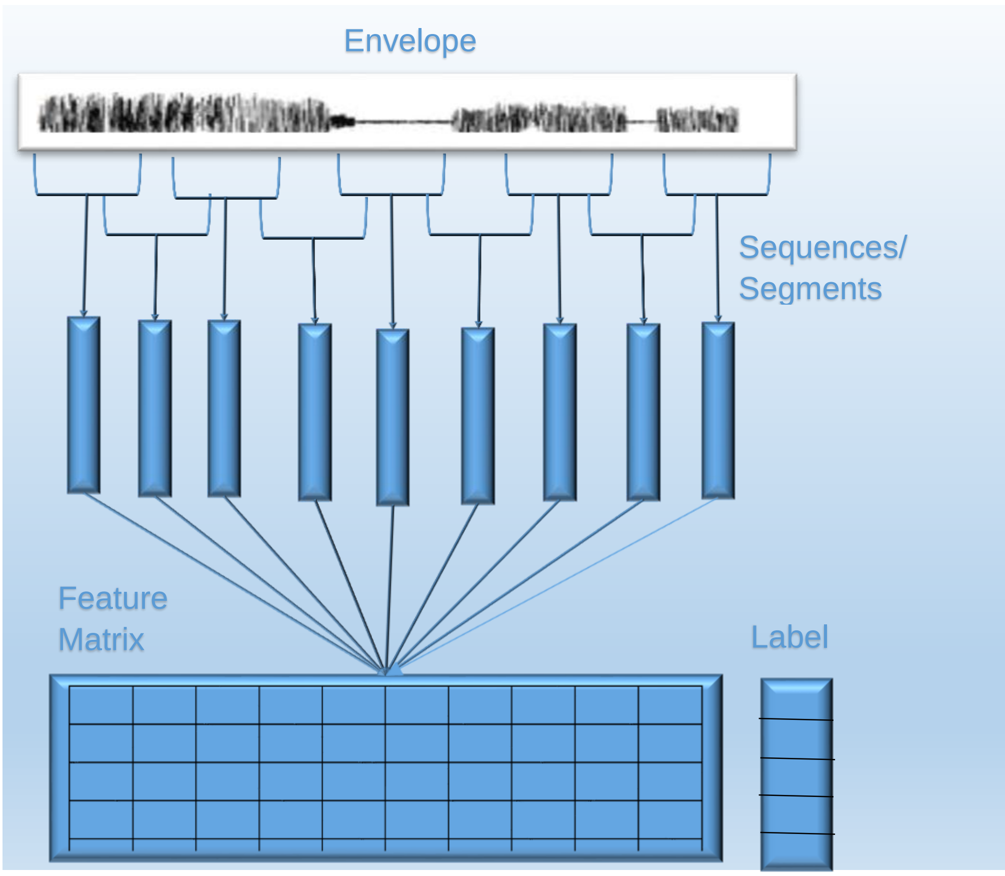

Figure 9: Illustration of converting generated envelope of a signal to a sequence/feature matrix using fixed length segmentation.

<u>Down Sampling the Feature Matrix</u>

We used Down Sampling method to avoid a biased distribution of positive and negative samples for training. As per the observation, a normal sleep study is a 7-8 hours of human bio-signal acquisition at the sampling rate of 128 Hz. The occurrences of a particular event signals are very few in compared to the total duration of the signal.

Therefore, most of the time a negative labeled data is in higher proportion. While experimenting we experienced a biased result. So, we decided randomly trimming down the negative labeled data to twice the size of positive labeled data, keeping 2:1 ratio respectively.

Hidden Markov Model Training and Validation

We started working with Hidden Markov Model (HMM) Toolbox for MATLAB written by Kevin Murphy, because it had libraries for continuous HMM. There were a number of other available libraries and even MATLAB had one, but they all were for discrete HMM. We were interested in continuous HMM because our data was not a discrete observation and we were trying to employ temporal pattern recognition technique. Unfortunately, HMM Toolbox was not appropriate in building HMM state transition matrix and produced erroneous model. We made many attempts to change dataset format and libraries in Toolbox, but did not get desired results. At the same time, we started looking for alternatives and found Gesture Recognition Toolkit (GRT), open-source, C++ machine learning library that had been specifically designed for real-time gesture recognition. GRT had libraries for continuous HMM, we extracted the core C++ HMM implementation and started working with our dataset available on MATLAB. GRT libraries did work with our training and testing dataset and resulting appropriate HMM model. We continued training and validating with rest of the samples. Figure 10 shows a training and testing process of a HMM.

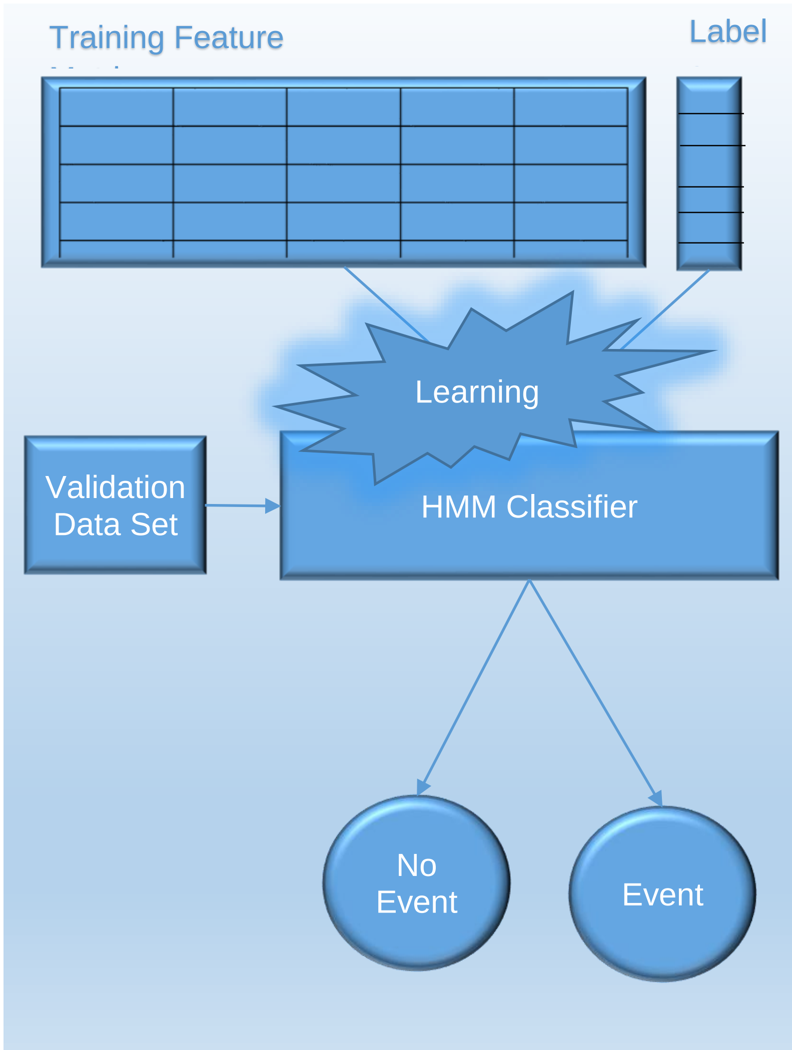

Figure 10: Representation of training and validation process of HMM built from GRT libraries. After learning phase model accurately desired events given by validation dataset.

<u>Performance Measure</u>

We performed 10 fold – cross validation on our samples. For each event detection, we performed 10 trials and in each round, we selected nine datasets for training and excluded one dataset for testing. We generated a confusion matrix for each trial and calculated the accuracy, recall, specificity and precision. Accuracy measures the overall correctness of the classifier. Recall measures true positive rate, also known as sensitivity that is when it's actually an event, how often does the classifier predicts an event.

Specificity measures the true negative rate that is when it's actually a non-event, how often does the classifier predict a non-event. Finally, precision is the measure of closeness that is when the classifier predicts an event, how often is it correct. Find tabulated results below for comparison.

IV. RESULTS

This section shows the results of our experiments. We have classified five sleep events and their results are in Tabular, chart and confusion matrix form.

<u>Limb Movements</u>

Table 1: Limb movement results.

Subjects	Accuracy	Recall	Precision	Specificity
User1	83.89	61.29	86.43	83.1
User2	84.3	58.75	90.96	82.48
User3	82.17	55.23	86.36	81.03
User4	73.41	34.32	70.9	73.89
User5	86.71	72.18	85.71	87.1
User6	77.6	53.12	72.34	79.31
User7	85.21	78.93	77.21	89.35
User8	81.45	88.67	66.68	93.21
User9	83.5	58.92	87.5	82.34
User10	80.8	65.81	73.77	83.78
Average	**81.90**	**62.72**	**79.78**	**83.55**

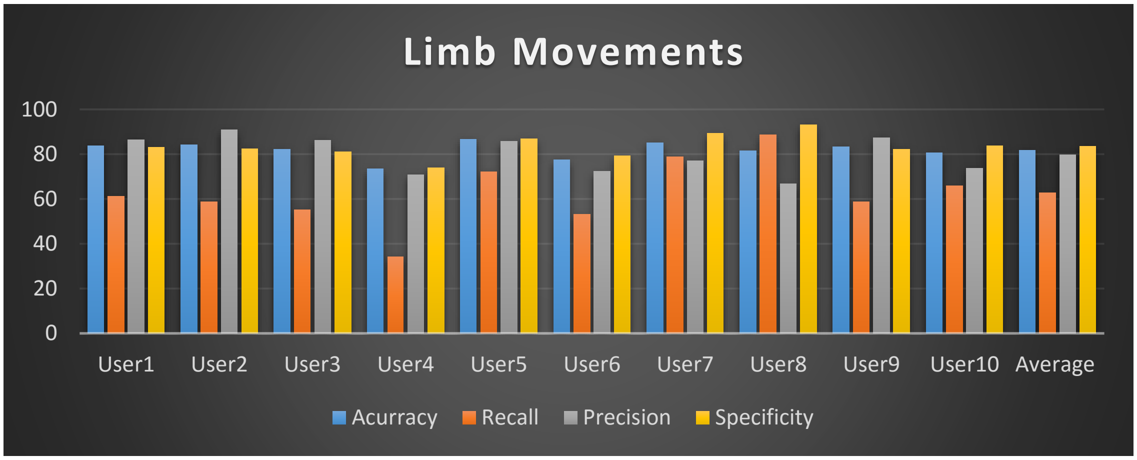

Figure 11: Limb movement representation of accuracy, recall, precision and specificity.

	Class 1	Class 2	Classification overall
Class 1	11148	1474	12622
Class 2	2024	4287	6311
Truth overall	13172	5761	18933

Overall accuracy (OA):	81.524%
Kappa[1]:	0.575

Figure 12: In the confusion matrix Class-1 represents as a no movement and Class-2 represent as movement.

Obstructive Apnea

Table 2: Obstructive apnea results.

Subjects	Accuracy	Recall	Precision	Specificity
User1	76.72	55.17	68.81	79.6
User2	80.51	72.3	70.14	85.93
User3	78.57	64.28	69.23	82.75
User4	76.51	52.27	69.69	78.78
User5	78.33	60	70.58	81.39
User6	74.5	64.7	61.11	81.81
User7	81.81	68.18	75	84.78
User8	80.95	64.28	75	83.33
User9	79.48	61.53	72.72	82.14
User10	76.66	60	66.66	80.95
Average	**78.40**	**62.27**	**69.89**	**82.14**

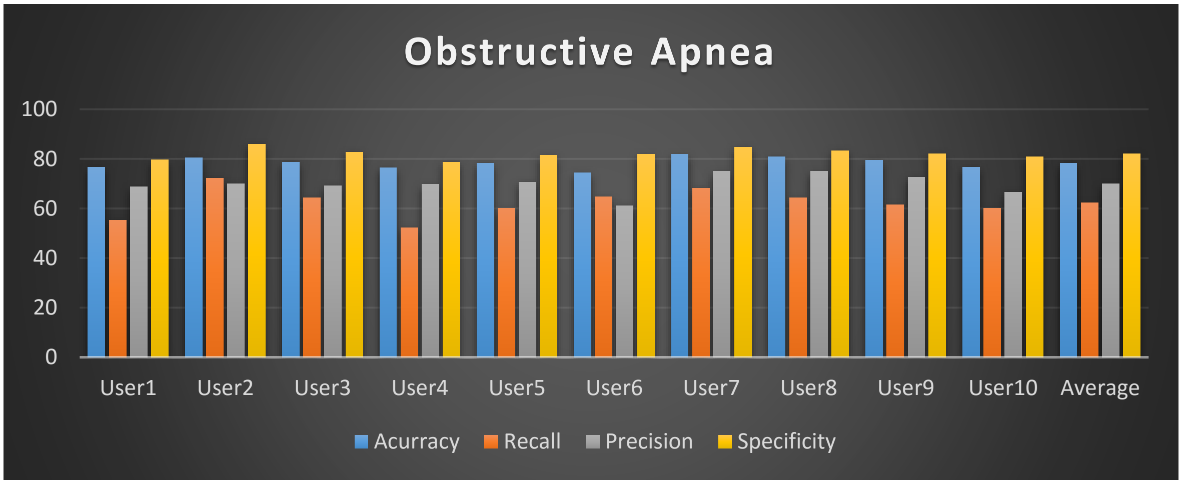

Figure 13: Obstructive apnea representation of accuracy, recall, precision and specificity.

	Class 1	Class 2	Classification overall
Class 1	581	89	670
Class 2	131	204	335
Truth overall	712	293	1005

Overall accuracy (OA):	78.109%
Kappa[1]:	0.492

Figure 14: In the confusion matrix Class-1 represent as a no apnea event and Class-2 represent as apnea event.

Hypopnea

Table 3: Hypopnea results.

Subjects	Accuracy	Recall	Precision	Specificity
User1	80	87.5	64.81	92.42
User2	75	59.78	77.78	80.72
User3	76.51	60.45	62.89	78.96
User4	75.21	52.9	55.87	78
User5	77.88	63.44	74.64	78.18
User6	74.89	62.11	62.72	76.66
User7	73.33	71.85	70.13	73.96
User8	74.5	73.06	77.05	76.08
User9	68.48	66.87	53.488	73.77
User10	75	68.21	64.86	78.94
Average	**75.08**	**66.62**	**66.42**	**78.77**

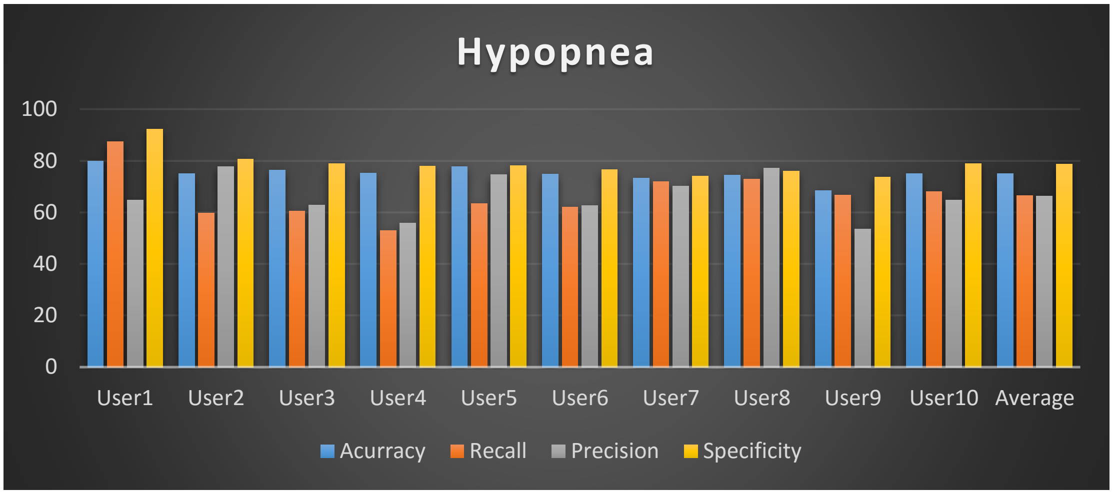

Figure 15: Hypopnea representation of accuracy, recall, precision, and specificity.

	Class 1	Class 2	Classification overall
Class 1	974	158	1132
Class 2	221	345	566
Truth overall	1195	503	1698

Overall accuracy (OA): 77.68%

Kappa[1]: 0.483

Figure 16: In the confusion matrix Class-1 represent as a no hypopnea event and Class-2 represent as hypopnea event.

Rapid Eye Movements (REM)

Table 4: Rapid eye movements results.

Subjects	Accuracy	Recall	Precision	Specificity
User1	74.33	63.77	61.05	81.44
User2	61.95	44.63	47.51	69.44
User3	72.22	74.58	78.57	66.3
User4	65.25	63.38	87.49	37.95
User5	50.77	85.17	39.077	81.9
User6	77.13	63.92	66.3	82.69
User7	78.43	71.88	66.27	85.32
User8	64.56	57.64	65.78	63.64
User9	68.33	55.5	53.11	77.23
User10	76.51	67.24	71.91	79.23
Average	**68.94**	**64.77**	**63.70**	**72.51**

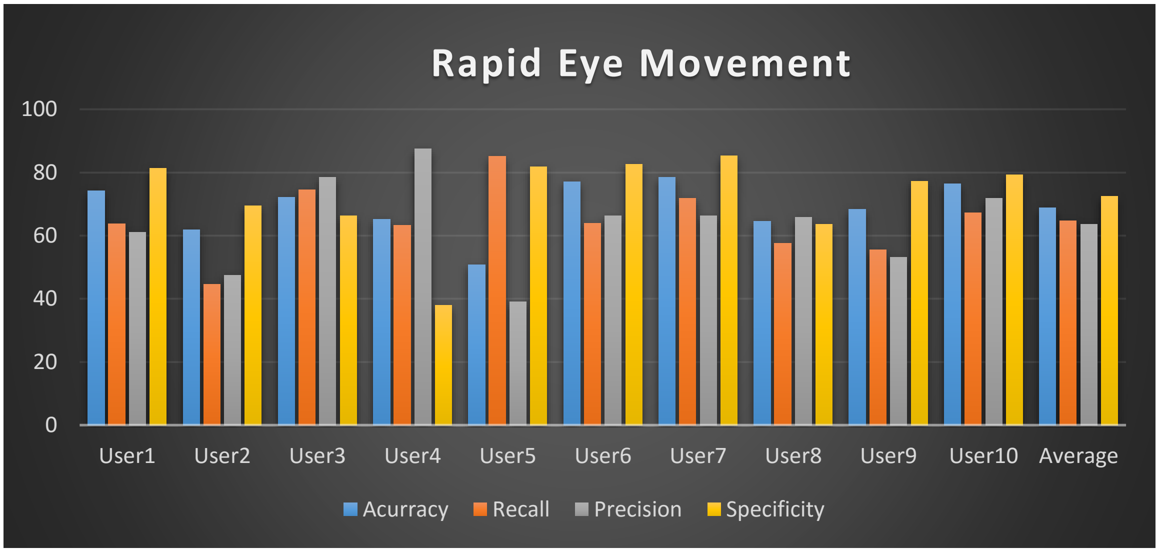

Figure 17: REM representation of accuracy, recall, precision and specificity.

	Class 1	Class 2	Classification overall
Class 1	8924	3272	12196
Class 2	2221	3877	6098
Truth overall	11145	7149	18294

Overall accuracy (OA):	69.974%
Kappa[1]:	0.352

Figure 18: In the confusion matrix Class-1 represent as a no REM and Class-2 represent as RME

Snore Detection

Table 5: Snore detection results.

Subjects	Accuracy	Recall	Precision	Specificity
User1	80.39	41.17	100	77.27
User2	77.77	33.33	100	75
User3	90.19	72.05	98	87.66
User4	96.92	95.23	95.25	97.72
User5	91.11	85.18	87.78	92.7
User6	87.93	67.24	95.12	85.71
User7	88	74	88.09	87.96
User8	91.02	82.69	89.58	91.66
User9	90.88	72.41	100	87.87
User10	90.9	75	97.05	88.77
Average	**88.51**	**69.83**	**95.09**	**87.23**

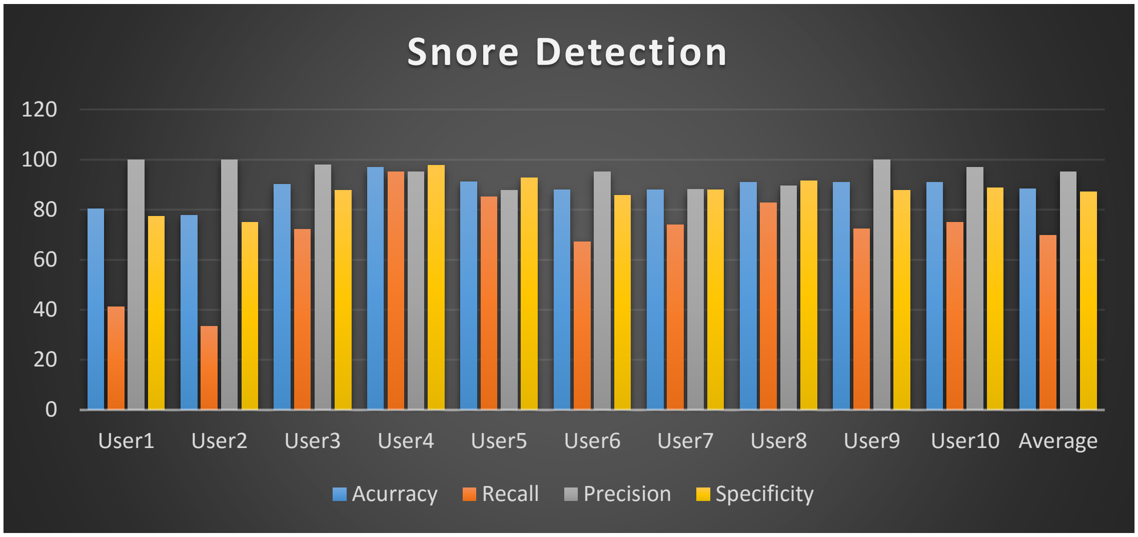

Figure 19: Snore representation of accuracy, recall, precision and specificity.

	Class 1	Class 2	Classification overall
Class 1	1534	68	1602
Class 2	168	633	801
Truth overall	1702	701	2403

Overall accuracy (OA):	90.179%
Kappa[1]:	0.772

Figure 20: In the confusion matrix Class-1 represent as a no snore event and Class-2 represent as snore event.

Comparison

Table 6: Comparing results.

	Avg. Accuracy	Avg. Recall	Avg. Precision	Avg. Specificity
Limb Movement	**81.90**	**62.72**	**79.78**	**83.55**
Rapid Eye Movement	**68.94**	**64.77**	**63.70**	**72.51**
Obstructive Apnea	**78.40**	**62.27**	**69.89**	**82.14**
Hypopnea	**75.08**	**66.62**	**66.42**	**78.77**
Snore	**88.51**	**69.83**	**95.09**	**87.23**

Left Leg Movement Event

Fixed Window Segmentation			
Classifier	Accuracy	Recall	Precision
DT	**74.86%**	**90.38%**	79.94%
LR	73.75%	83.85%	82.67%
NB	68.21%	67.58%	**88.91%**
Adaptive Segmentation			
Classifier	Accuracy	Recall	Precision
DT	**88.39%**	**98.03%**	89.96%
LR	86.99%	96.11%	90.09%
NB	57.97%	57.65%	**92.92%**

Right Leg Movement Event

Fixed Window Segmentation			
Classifier	Accuracy	Recall	Precision
DT	**74.86%**	**90.19%**	80.42%
LR	73.84%	86.57%	81.52%
NB	67.10%	66.82%	**88.82%**
Adaptive Segmentation			
Classifier	Accuracy	Recall	Precision
DT	**89.79%**	**98.55%**	90.94%
LR	87.44%	95.46%	91.11%
NB	57.35%	57.20%	**93.05%**

Figure 21: From (Espiritu and Metsis, 2105), shows the leg movement classification using Decision Tree (DT), Linear regression (LR) and Naïve Bayes (NB).

Discussion

Results shown in Table 6 Indicate that the highest Avg. Accuracy 88% was achieved in snore detection, along with highest Avg. Recall, Avg. Precision, and Avg. Specificity is 69%, 95%, and 87% respectively. Snore signal is a sound wave (audio signal) and provides clear variation in its amplitude, which can be easily detected in envelope. Considering Obstructive Apnea and Limb movements, CPAP air flow is widely used in apnea detection and EMG signal for leg movements. However, 78% of accuracy was achieved in Obstructive apnea and 81% in Limb movements. These results could have been improved by introducing more accurately labeled data.

For Hypopnea detection, let us have a look at the table 3, which clearly indicates that recall is low and precision is higher. That means model considered some of the Hypopnea events as normal breathing and most of the time wrongly classified to no events. Looking at figure 18 REM confusion matrix, we can see that out of the total 6908 events, 2221 have been classified as non-REM events. The reason for the low recall is that the REM pattern is slightly different from eye blinks pattern. Hence, most of the eye blinks are classified to REM events. We also got a chance to compare our results with the results of (Espiritu and Metsis, 2015), which have also been obtained using the same data source Profusion PSG. The result of the above mentioned comparison indicated that the accuracy of fixed window segmentation for leg movements is around 74%, which is 81% in our experiment. We did not apply adaptive segmentation but it is clear from figure 21 that accuracy could have been improved if adaptive segmentation method had been applied.

Conclusion

To conclude my observation, the results of the study show that the HMM is capable of detecting the different sleep events with multi-channeled features. However, the performance in Rapid eye movements and Hypopnea shows that more work is required to produce reliable event scoring. Due to higher variation in the EOG signal, we found that envelope might not be a good choice for classifying rapid eye movements, Normal eye blinks are more often falsely classified to REM. While classifying breathing events there is little difference between the performances. It was found that snores were consistently easier to detect than apneas. We even found that Hypopneas are more often falsely classified to normal breathing. It is possible that more advanced signal processing techniques could produce better results. Another issue is the labeling of the signals, we used labels provided from Profusion PSG3, which is already an automated system and could provide inaccurate labels. The results of this experiment may have turned out more accurate if the labels used for training were manually scored (hand-labelled) and quality of the signals was superior.

BIBLIOGRAPHY

Alessandria, M. and Provini, F. (2013) ‘Periodic limb movements during sleep: A new sleep-related cardiovascular risk factor?’, *Frontiers in Neurology*, 4. doi: 10.3389/fneur.2013.00116.

Baraglia, D.P., Berryman, M.J., Coussens, S.W., Pamula, Y., Kennedy, D., Martin, A.J. and Abbott, D. (2005) ‘<title> Automated sleep scoring and sleep apnea detection in children</title>’, *Complex Systems*, doi: 10.1117/12.638867.

Huang, R.S., Kuo, C.J., Tsai, L.-L. & Chen, O.T.C. (1996) ‘EEG pattern recognition-arousal states detection and classification’, *Proceedings of International Conference on Neural Networks (ICNN’96)*, doi: 10.1109/icnn.1996.548971.

Kempfner, J., Sorensen, G.L., Sorensen, H.B.D. and Jennum, P. (2011) ‘Automatic REM sleep detection associated with idiopathic rem sleep behavior disorder’, *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, . doi: 10.1109/iembs.2011.6091498.

Luo, H., Fang, X. and Ertas, B. (2009) ‘Hilbert transform and its engineering applications’, *AIAA Journal*, 47(4), pp. 923–932. doi: 10.2514/1.37649.

MacDonald, I.L. and Zucchini, W. (2009) *Hidden Markov models for time series: A practical introduction using R*. United States: Chapman & Hall/CRC.

Metsis, V., Galatas, G., Papangelis, A., Kosmopoulos, D. and Makedon, F. (2011) ‘Recognition of sleep patterns using a bed pressure mat’, *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA ’11*, . doi: 10.1145/2141622.2141633.

Metsis, V., Kosmopoulos, D., Athitsos, V. and Makedon, F. (2012) ‘Non-invasive analysis of sleep patterns via multimodal sensor input’, *Personal and Ubiquitous Computing*, 18(1), pp. 19–26. doi: 10.1007/s00779-012-0623-1.

Nakano, H., Tanigawa, T., Furukawa, T. and Nishima, S. (2007) ‘Automatic detection of sleep-disordered breathing from a single-channel airflow record’, *European Respiratory Journal*, 29(4), pp. 728–736. doi: 10.1183/09031936.00091206.

Nguyen, N. and Edu, G.C. (no date) *Comparisons of sequence labeling Algorithms and extensions Yunsong Guo*. Available at: http://www.machinelearning.org/proceedings/icml2007/papers/206.pdf (Accessed: 1 July 2016).

NHLBI National Center on Sleep Disorders Research (2011) *2011 national institutes of health sleep disorders research plan*. Available at: http://www.nhlbi.nih.gov/files/docs/resources/sleep/201101011NationalSleepDisordersResearchPlanDHHSPublication11-7820.pdf (Accessed: 22 June 2016).

Rabiner, L. and Juang, B. (1986) ‘An introduction to hidden Markov models’, *IEEE ASSP Magazine*, 3(1), pp. 4–16. doi: 10.1109/massp.1986.1165342.

Shi, S. and Weigend, A.S. (1997) ‘Taking time seriously: Hidden Markov experts applied to financial engineering’, *Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr)*, . doi: 10.1109/cifer.1997.618944.

Sung, C.-H. and Priebe (1988) ‘Temporal pattern recognition’, *IEEE International Conference on Neural Networks*, . doi: 10.1109/icnn.1988.23907.

Wu, J. and Xie, J. (2009) 'Hidden Markov model and its applications in motif findings', *Statistical Methods in Molecular Biology*, , pp. 405–416. doi: 10.1007/978-1-60761-580-4_13.