SLEEP STAGE CLASSIFICATION USING HYDRAULIC BED SENSOR

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Ruhan Yi

James Keller, Thesis Supervisor

DECEMBER 2018

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

SLEEP STAGE CLASSIFICATION USING HYDRAULIC BED SENSOR

presented by Ruhan Yi,

candidate for the degree of master of science

and hereby certify that, in their opinion, it is worthy of acceptance.

James M. Keller

Marjorie Skubic

Mihail Popescu

DEDICATION

This thesis is dedicated to my parents for their endless support and encouragement.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. James Keller for his patience, motivation, and continuous support of my master's study. His guidance helped me in all the time of research and writing this thesis.

My sincere thanks also goes to Dr. Marjorie Skubic and Dr. Mihail Popescu. Thanks for providing me an opportunity to join the sleep study group and for their insightful comments.

I also want to thank my colleague, Moein Enayati, who gave me a lot of suggestions and was always willing to help.

TABEL OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Sleep monitoring can help physicians diagnose and treat sleep disorders. Polysomnography(PSG) system is the most accurate and comprehensive method widely used in sleep labs to monitor sleep. However, it is expensive and not comfortable, patients have to wear numerous devices on their body surface. So a non-invasive hydraulic bed sensor has been developed to monitor sleep at home.

In this thesis, the sleep stage classification problem using hydraulic bed sensor was proposed. The sleep process divided into three classes, awake, rapid eye movement (REM) and non-rapid eye movement (NREM). The ground truth sleep stage came from regularly scheduled PSG studies conducted by a sleep-credentialed physician at the Sleep Center at the Boone Hospital Center (BHC) in Columbia, Missouri. And we were allowed to install our hydraulic bed sensors to their study protocol for consenting patients. The heart rate variability (HRV) features, respiratory rate (RV) features, and linear frequency cepstral coefficient(LFCC) were extracted from the bed sensors' signals. In this study, two scenarios were applied, put all subjects together and leave one subject out. In each scenario, two types of classification structures were implemented, a single classifier and a multi-layered hierarchical method. The results show both potential benefits and limitations for using the hydraulic bed sensors to classify sleep stages.

CHAPTER 1. INTRODUCTION

Sleep occupies a considerable part of a person's life. Good sleep quality contributes to the repair of the human physiological and neurological systems. Insufficient sleep may increase the incidence of chronic diseases and may also threaten public safety. Sleep disorders, such as insomnia and sleep apnea, have become factors that affect people's daily life. Studies show that the prevalence of obstructive sleep apnea (OSA) in the general population ranges from 9% to 38% and is higher in men [1]. Davies et al. [2] noted that 92% of women and 82% of men with moderate to severe OSA are undiagnosed. In addition, sleep disturbance is associated with many neurological and psychiatric diseases, such as Alzheimer's Disease [2, 3] and Parkinson's disease [4]. Therefore, monitoring sleep and studying sleep structure is especially important.

Polysomnography(PSG) is a multi-parametric system widely used in sleep labs to diagnose and treat sleep disorders. By placing many electrodes, sensors, tubes, and masks on a patient's body surface, the system is able to simultaneously monitor multiple biological signals, including those signals normally recorded by an electroencephalogram (EEG), electrocardiography (EKG), electrooculography (EOG), and electromyography (EMG). The conditions monitored include airflow, respiratory effort, leg movements, and oxygen saturation. Besides, a technician monitors the patient throughout the night and annotates the sleep stages based on 30-second epochs. The sleep scoring follows the American Academy of Sleep Medicine (AASM) Manual [5]. The normal sleep structure is the alternate of rapid eye movement (REM) sleep and

non-rapid eye movement (NREM) sleep. Furthermore, the NREM sleep from light to deep can be divided into stages NREM1, NREM2, NREM3, and NREM4. Although the PSG system has great advantages in accuracy and comprehensiveness, inevitably, it is expensive, and it can only be done by a sleep-credentialed physician in a sleep lab. In addition, wearing all the devices on the body surfaces and sleeping in a different bed in a new environment will change the patients' sleep pattern and affect the sleep study results to some extent. In this case, widely applicable non-invasive sleep monitoring systems can greatly facilitate in-home sleep monitoring [6, 7].

A non-invasive hydraulic bed sensor has been developed to monitor sleep at home [8, 9]. No electrodes are placed on the body surface. The sensor is installed under the mattress. The system starts collecting signals as soon as a patient lies on the bed. Compared with the PSG system, it is convenient to operate and will not affect a person's normal sleep pattern. Previous research has proposed reliable approaches to detect the heart rate and respiration rate from the hydraulic bed sensor signals [10-12]. Further, variabilities in the heart rate and respiration are also extracted. Many studies have shown high classification accuracy with these features [7, 13-15]. Therefore, the sensors' ability to provide the features of our heart rate and respiration feedback are applied in this study.

For this thesis, the data collection took place during the regularly scheduled PSG studies conducted by a sleep-credentialed physician at the Sleep Center at the Boone Hospital Center (BHC) in Columbia, Missouri. We were allowed to install our hydraulic bed sensors to their study protocol for consenting patients. Furthermore, we

shared the de-identified PSG data.

In this study, two scenarios were applied—one where all the sleep study participants are together and the other where all the participants are together except for one subject, who is left out. In each scenario, two types of classification structures were implemented, a single classifier and a multi-layered hierarchical method. The support vector machine (SVM) with different types of kernels and k nearest neighbors (k-NN) with a varying number of neighbors and different distance metrics were applied. The heart rate variability (HRV) features, respiratory rate (RV) features, and linear frequency cepstral coefficient(LFCC) were extracted from the bed sensors' signals. The results show both potential benefits and limitations for using the hydraulic bed sensors to classify sleep stages. The performance where all subjects were together was much better than the leave-one-out performance.

CHAPTER 2. BACKGROUND

2.1 Sleep

The first terminology referencing sleep stages appeared in the 1930s. The official scoring system for staging the sleep of humans was first released in 1968 [16]. Since 2007, the American Academy of Sleep Medicine Manual (AASM, 2007) [17] has become the standard for scoring sleep stages and it also provides rules for associated events during sleep. The latest version of the manual was released in April 2018 (AASM 2018)[18] .

The AASM protocol requires sleep stage scoring in 30-second epochs. In different stages of sleep, human brainwaves, eye movements, and muscle activities show different patterns. In each 30-second epoch, after integrating all this information, the technician can classify sleep into different stages.

Normal human sleep can be divided into two main categories, REM sleep and NREM sleep. Further, from shallow to deep, the NREM sleep is divided into NREM1 (N1), NREM2 (N2), and NREM3 (N3) sleep. A normal night's sleep cycle consists of these two types of sleep stages alternating and accompanied with awake. A hypnogram is a visual representation of the sleep process, as shown in Figure1, which shows the sleep cycle of a normal night's sleep [19]. The sleep cycle starts with N1 and then gradually transitions to N2 and N3 before finally settling into REM. The new cycle begins by following the short duration of REM sleep. In a normal night sleep, there will be 4 to 5 cycles. As sleep progresses, the duration of REM sleep increases and N2 accounts for the majority of NREM sleep [19]. Overall, for adults REM sleep accounts

for about 20 to 25 percent of total sleep time and NREM accounts for 75 to 80 percent [20].

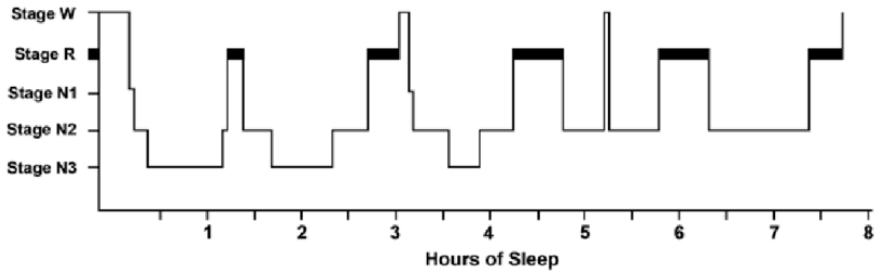


Figure 1. Example of hypnogram for the entire night

REM sleep is the stage in which our eyes move rapidly and most dreams happen. The characteristics of NREM sleep are different from N1 to N3. N1 is a transition from wake to sleep. Most sleep begins with N1, and this sleep can be easily interrupted by external noises. N2 accounts for 45 to 55 percent of NREM sleep. The duration of N2 gradually increases in the successive sleep cycle. As sleep progresses deeper, more external stimuli are needed to wake up from sleep. The N3 stage is known as slow wave sleep, and it only appears in the first few cycles.

There are many body system changes from one stage to another, such as sympathetic nerve activity, respiratory, and cardiovascular system changes [20]. All of these changes form the basis for extracting the features in the following method.

2.2 Ballistocardiography (BCG)

Ballistocardiography (BCG) is a non-invasive method that can measure body motion generated by the blood that flows out of lower heart chambers (left to right ventricles) with each heart contraction [21]. The modern BCG waveform was proposed by Isaac Starr [22] when he constructed a bed BCG measurement device. Since then, many different types of BCG measurement devices have become available, BCG

waveforms are similar. Figure 2 shows the theoretical BCG waveform, which represents the different phases of the heartbeat. The fiducial peak point F-G-H complex represents pre-ejection, I-J-K complex represents the ejection process, L-M-N represents the diastolic process. Peak J in BCG corresponds to the R peak in the electrocardiogram (ECG) waveform. The time interval of RJ peaks represents the time differences in the response to the electrical activation in the left and right ventricles as well as body motion caused by this activation. Figure 3 shows the RJ interval. The detection of the J peak in the BCG waveform is an important foundation for extracting the features related to the heartbeat.

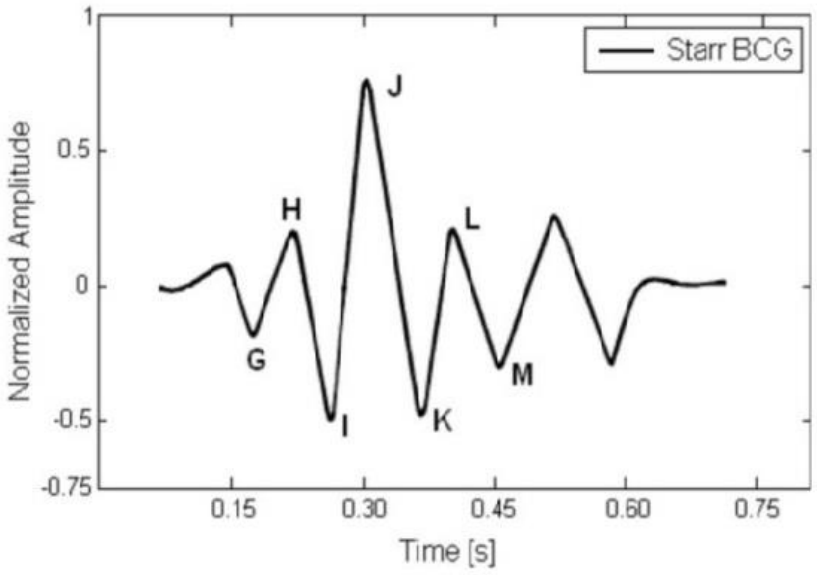


Figure 2. Theoretical BCG waveform.

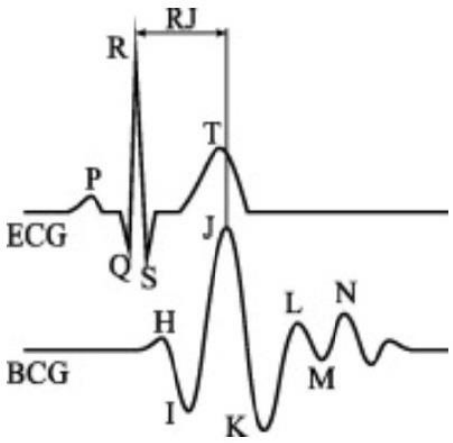


Figure 3. RJ interval of ECG and BCG signal

2.3 Polysomnography(PSG)

Polysomnography(PSG) is a multi-parametric recording method applied in sleep labs to monitor physiological changes during sleep. It is a reliable tool for diagnosing sleep disorders, and it can also help adjust the treatment. In this study, we collected 77 subjects' PSG data in cooperation with the BHC sleep lab. A total of 21 channel signals were collected by attaching different types of devices to the subjects. For instance, electrodes were placed on the subjects' head measuring EEG and eye movements. Belts embedded with sensors were secured around the subjects' chest and abdomen to monitor respiration. Snoring was recorded by a mini microphone placed close to the chin.

Sleep apnea is a sleep disorder, which can cause a person to repeatedly stop breathing during their sleep. Those diagnosed with sleep apnea are prescribed a continuous positive airway pressure (CPAP) mask or a bilevel positive airway pressure (BiPAP) mask. These two types of masks can deliver the flow of pressure to keep the wearer's airway open during sleep. In the BHC sleep lab, a software called Natus

SleepWorks (Natus Medical Inc., San Carlos, CA, USA) is used to help the staff technician monitor a patient's sleep during the night. It not only collects the PSG data but also performs a video recording. If the technicians have any uncertainty about the data, they can view the patient's sleep video. It can provide a preliminary analysis of the collected data and can generate a report, which assists those physicians who make treatment recommendations. Figure 4 shows the 21 channels of PSG signals visualized in the Natus SleepWorks interface. The occurrence of obstructive apnea was also annotated by a technician and can be seen as the orange bar in Figure 4.

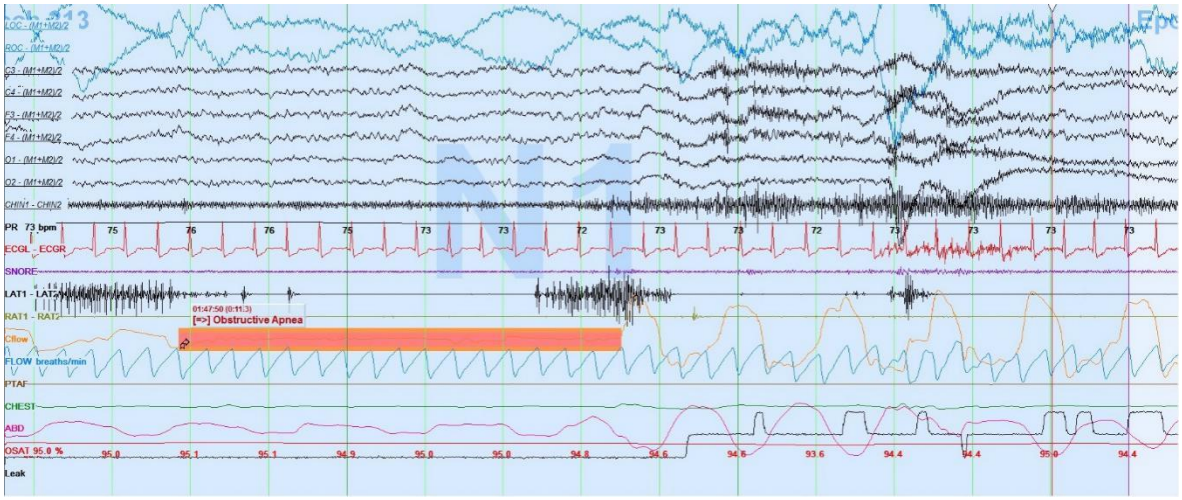


Figure 4. Natus SleepWorks visualization

2.4 Heart Rate Variability (HRV)

Numerous body systems change based on the sleep stages, which means physicians can use the monitored cardiovascular system, autonomic nervous system, or respiratory system information to determine sleep stages. Variations in heart rate controlled by the autonomic nervous system [23], and heart rate variability (HRV) can be easily acquired; thus, HRV is considered one of the characteristics which can give insight into what happens to the human body during sleep stages.

Two branches of the autonomic nervous system are related to sleep, the sympathetic nervous system and parasympathetic nervous system. In many circumstances, the two systems have an opposite reaction. Previous studies have demonstrated that when subjects have a deeper sleep in the NREM sleep stages, the heart rate slows down, the sympathetic nervous activity decreases from wakefulness, and parasympathetic activity gradually increases. During REM sleep, the heart rate increases, and the sympathetic nervous activity increases comparable to wakefulness [24].

Measurement of HRV in time and frequency domains are discussed in this study. The premise of calculating HRV is to detect the heart beat and successive heartbeat intervals. Details of time domain variables derived from the heartbeat intervals are discussed in Section 3.3.1.1. In the frequency domain, the power spectral density of the heartbeat interval are decomposed into specific frequency bands. Several studies demonstrate the relation between frequency bands and the autonomic nervous system [23, 25]. The low frequency band (0.04–0.15 Hz) is influenced by both sympathetic and parasympathetic systems. The high frequency band (0.15–0.40 Hz) is related to the parasympathetic nervous system. The ratio of the low frequency band and high frequency band represents the balance of the sympathetic and parasympathetic system. The HRV frequency domain's derived process is discussed in Section 3.3.1.2.

2.5 Respiratory Variability (RV)

Numerous studies have shown that breathing regulation is significantly different between wakefulness and sleep [26-28]. Sleep is a dynamic physiologic state. At the

beginning of sleep, significant changes happen in the respiration control processes, particularly in REM sleep and NREM sleep. Minute ventilation represents the volume of air a person inhales or exhales in a minute. It starts to fall at the onset of sleep. During the NREM sleep, the minute ventilation shows increased regularity. The lowest level minute ventilation occurs during slow wave sleep. In comparison to NREM sleep, REM sleep is characterized by a reduced regularity, which is further reduced in minute ventilation. Therefore, RV can be considered as a characteristic that distinguishes different sleep stages.

2.6 Performance measurements

2.6.1 Classifier performance measurements

In a common classification problem, accuracy is one of the widely used performance measurements. It is expressed as the proportion of correct prediction. However, as shown in Table 2, the sleep stage data is imbalanced. The percentage of each sleep stage varies greatly. The NREM stage accounts for the largest proportion; thus, the average percentage for 33 preliminarily selected subjects in this research is 62%. The awake and REM only accounted for 21% and 17%, respectively. The problem of using accuracy in imbalanced data is that high accuracy means we classify the majority class. For sleep studies, class imbalance is a common problem. Typically, both accuracy and Cohen's kappa coefficient are used to measure the performance of the classification. The kappa value (κ) measures the inter-rater agreement and is considered to be a more robust way to measure the agreement [29, 30]. In this thesis, the agreement of ground truth and predicted results were calculated in each experiment.

The formulas of accuracy and kappa value are described as follows:

The confusion matrix is defined as:

		Predicted Class	
		Class1	Class2
Actual Class	Class1	True Positive(TP)	False Negative(FN)
	Class2	False Positive(FP)	True Negative(TN)

The accuracy is defined as:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \tag{2-1}$$

The kappa value is defined as:

$$\kappa = \frac{p_0-p_e}{1-p_e} \tag{2-2}$$

where

$$p_0 = \frac{TP+TN}{TP+FN+FP+TN} \tag{2-3}$$

$$p_e = \frac{(TP+FN)*(TP+FP)+(FP+TN)*(FN+TN)}{(TP+FN+FP+TN)^2} \tag{2-4}$$

where $p_0 = Accuracy$ and represents the observed agreement between two classes.

p_e represents the probability of chance agreement.

The following table shows the agreement guidelines proposed in [31]. The strength of the agreement assigned to the specific range of kappa value.

Table 1. Strength of agreement corresponding to kappa value

Kappa	Strength of Agreement
< 0.00	Poor
0.01-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

2.6.2 Sleep quality measurements

In order to measure sleep quality, we may use [32]:

- Sleep efficiency (SE): The ratio of total sleep time and total time in bed.
- Percentage of REM sleep (SR): The ratio of time spent in REM sleep and total sleep time.

2.7 Validation

Here, we used two cross-validation scenarios to validate our classifiers. The first scenario is referred to as the 10-fold cross-validation. We put all the epochs of selected subjects together and partitioned the data into 10 equal sized subsets. Each time we used this validation method, one of the subsets was selected as the validation set for testing the model, and the remaining nine subsets were used for training the model. We repeated this approach 10 times and determined the accuracy by averaging the results. The benefit of this approach lies in the accuracy of its ability to evaluate the performance of the model's features.

The second scenario is called the leave-one-subject-out cross validation. Since

the purpose of designing a sleep stage classification system is to recognize the sleep stage of a new individual using existing data, this method has a more practical meaning. Each time we used this validation method, we selected one subject's data to use for testing. The remaining subjects' data were used for training the classifier. The system was designed to have robust performance results on the chosen subjects.

2.8 Literature review

Previous studies have reported sleep stage classification results under different conditions. The data can be obtained through a PSG system, bed sensor, radar system, and a watch-based device. The quality of the signal is quite different for these systems. Another factor is the subjects who participated in these studies. Most studies recruit healthy young adults; thus, when comparing the healthy participants to those subjects with a sleep disorder, the signal quality is much better. In this section, we compare classification results in literature.

Jialei Yang presented a sleep-stage classification method using bed sensors in her thesis [33]. Her data were collected from same healthy young subject for eight entire nights. Two scenarios were implemented for validation in the experiments. The first scenario put all the recordings together and used the 10-fold cross-validation. The best results for detecting REM were a 93% accuracy and a 0.78 kappa value obtained by using smoothed linear frequency cepstral coefficients (LFCC). For the three-stage classification, the accuracy was 81% with a 0.70 for kappa value. The three-stage classification results were better than most results for this type of test in the literature. The second scenario, the leave-one-night out method, removed one entire night's

testing data; the remaining nights were used for training. The REM detection accuracy and kappa value degraded significantly, at 62% and 0.09, respectively. Such a kappa value can be regarded as a random guess for classification. For the REM and awake stages together versus NREM, the results were worse. The accuracy decreased by 50% as did the negative kappa value, which means the classification results were worse than random guessing.

Park et al. [7] showed three threshold comparison experiments detecting sleep stages by using heart rate variability parameters derived from the BCG signal. The experiment detecting the awake epoch was based on the heart rate variation threshold. The result for normal subjects was 97.4% in accuracy with a 0.83 kappa coefficient. The result for the subjects with obstructive sleep apnea was 96% in accuracy with a 0.81 kappa coefficient. The experiment for detecting REM sleep compared three heart rate parameters with the threshold. If all of the parameters were higher than their corresponding thresholds, the epoch was labeled as REM. The accuracy of 92% with a 0.72 kappa coefficient for five normal subjects was reported. The third experiment was set up to detect deep sleep. Four HRV parameters from the time domain and frequency domain were selected; thus, all of the parameters lower than the threshold in the epoch regarded as deep sleep. The results showed 89.4% in accuracy with a 0.48 kappa coefficient. After detecting the awakening, REM, and deep sleep, the remaining tests were classified as light sleep epochs. The accuracy was 76.2%, and the kappa value was 0.53.

Redmond and Heneghan [34] proposed an automatic classification method

based on EEG signals. The subjects in this paper were separated into two group. The subjects in one group experienced low apnea events during their sleep. The subjects in another group experienced high apnea events. A number of heart rate and respiratory rate features were extracted. A subject-specified classifier yielded 79% accuracy. When a similar subject-independent classifier was trained, the accuracy dropped to 67%. For a comparison trained classifier with EEG features, the accuracy for a subject-specified classifier was 87%. The accuracy for the subject-independent classifier was 84%.

Kortelainen et al. [6] proposed a three-stage classification system using the Emfit bed sensors (Emfit Ltd., Vaajakoski, Finland). The bed sensors obtained heart beat intervals and movements during the user's sleep. A time-variant autoregressive model (TVAM) was used for extracting the features. A hidden Markov model (HMM) was used for training. Eighteen healthy subjects participated in the experiment. The system obtained 79% accuracy with a kappa value of 0.44.

Huang et al. [35] presented an automatic four-stage classification system with a hierarchical structure using forehead EEG signals. The hierarchical structure consists of five layers—the preliminary wake detection layer, three SVM classifier layers, and a final fifth layer with two adaptive adjustment schemes. Ten healthy subjects were involved in this study. The accuracy was about 77% with a kappa value of 0.67. A similar hierarchical structure was implemented in one of the experiments in this study.

Three out of five methods mentioned above used BCG signal, which described

similar purpose of using non-invasive method to classify the sleep stages in this thesis. Two of them had health subjects participated in the studies. The accuracy range from 76% to 81% were obtained in three-stage classification. The same features in Jialei's work were applied in this thesis, and obtained 85% accuracy. The accuracy for detecting REM was around 92% for healthy subjects were reported in Jialei's and Parks's works. Since the sleep data is imbalanced, NREM and wake together accounted for majority of the sleep stage. The high accuracy in detecting NREM and wake may contribute to accuracy of REM detection. In Jialei's thesis, 86% of true positive rate of REM was reported, which indicates REM was correctly classified as REM. However, Park did not show the confusion matrix or true positive rate of REM detection, so it is impossible to know the details. All of the studies show the potential of applying non-invasive sensor to build an automatic sleep stage classification system.

Only in Jialei's work the leave-one-night-out strategy was mentioned. The data were collected from same person for 8 nights. Although they are eight separate nights, it is still different from collecting the data from eight different people. The accuracy of 63% and 45% of true positive rate in detecting REM indicated that even the data collected from same person have different characteristic from night to night. In this thesis, the results from using this strategy were not satisfactory. So, using a leave-one-subject-out strategy in sleep studies, remains a challenging problem.

CHAPTER 3. METHODS

3.1 Data collection

3.1.1 Hydraulic bed sensors

The noninvasive hydraulic bed sensor system consists of four hydraulic bed transducers made of a flat hose partially filled with water. An integrated pressure sensor is connected to one end of each transducer. The pressure sensors can convert the weight placed on the surface into the voltage signal. These sensors are also sensitive to low amplitude variations, which makes it possible to detect a heartbeat [36].

The position of the four transducers is shown in Figure 5. To capture the heartbeat and respiration, the transducers were placed under the mattress of each bed with the same separation distance between each transducer and parallel to the direction in which the patient lies on the bed. The incline degree of each bed in the Boone Hospital Center's sleep lab is adjustable by the patient.

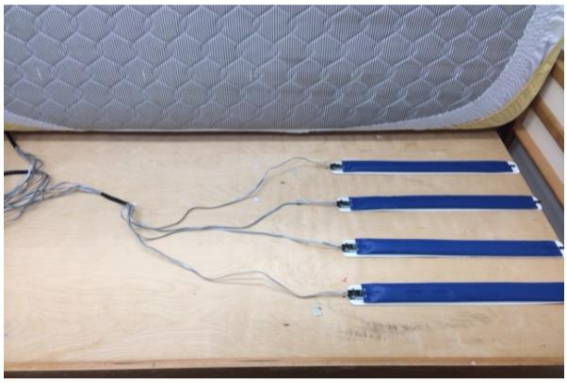


Figure 5. Placement of four transducers under the mattress.

The transducers were placed above the bending part of the bed to avoid folding when the patient changes the incline degree. Each transducer has two output channels: one for raw data and another for filtered signal. An 741 op-amp amplifier was applied to serve the hardware filtering circuit, and an 8th-order integrated Bessel filter was used

for filtering the noise [9]. Finally, the filtered signal was sampled at 100 Hz. The details of the bed sensor construction, hardware filtering, and refinement are described in previous work [9].

Figure 6 shows a 30-second timelines for the four channels in response to the raw signals and the filtered signal of the bed sensors with one patient lying on the bed.

In the feature extraction process, for each epoch only one transducer was selected. The selection criteria in this study is based on maximum DC value. We assumed that large DC value represents more weight on the transducer and a better connection with subjects.

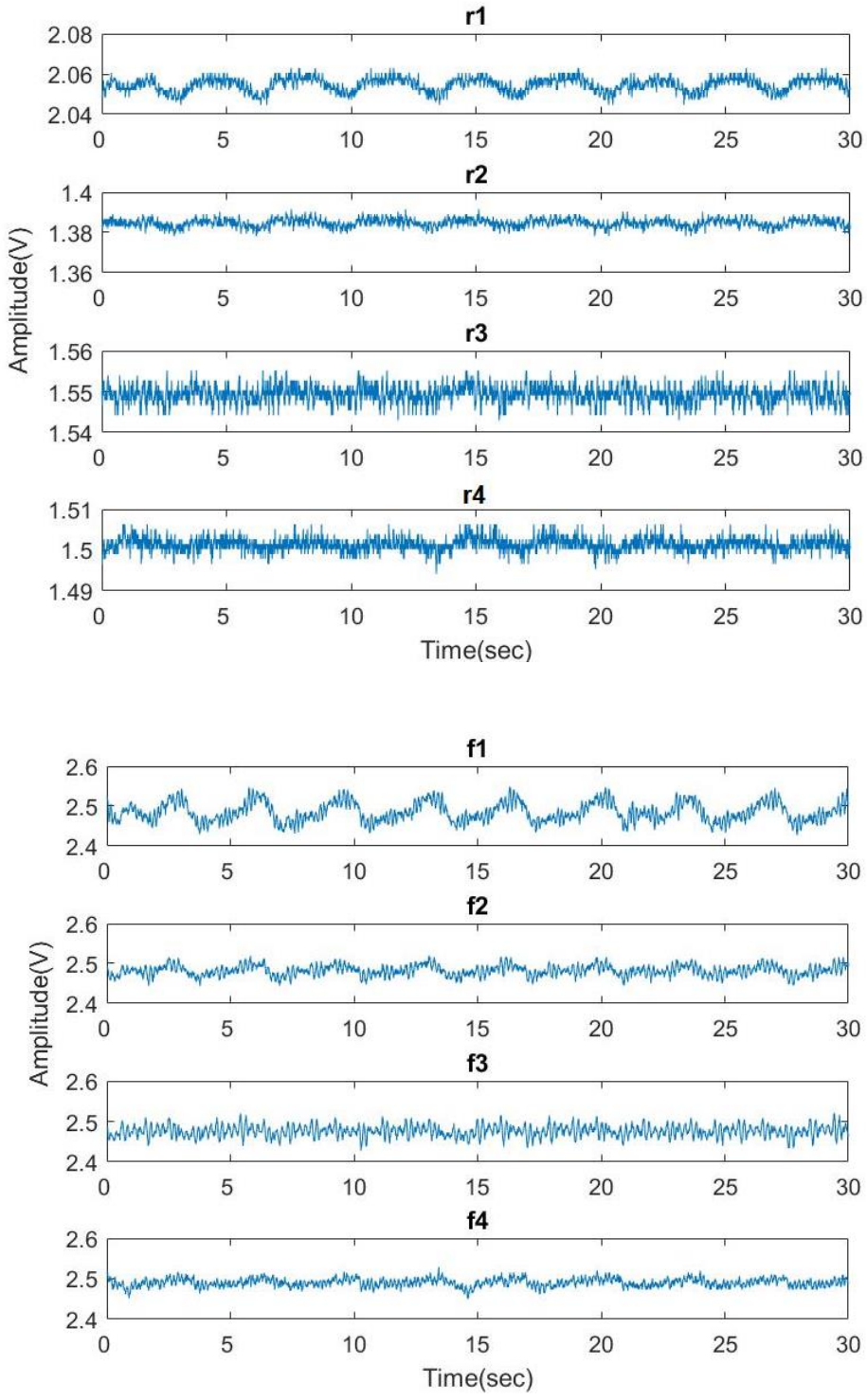


Figure 6. Thirty-second timelines for raw signals and filtered signals of four transducers. The top four channel timelines show the raw signals; the bottom four channel timelines show the filtered signals.

3.1.2 Boone Hospital Center (BHC) sleep lab data

As described in Section 2.3, the BHC sleep lab provided de-identified polysomnography (PSG) data. Different from bed sensor data, the sampling rate of all the PSG signal is 256 Hz. In addition, the annotations of sleep technician-recorded valuable information is essential to the success of this study. Aside from setting the tests up and preparing the patient for the tests, a sleep technician scores each patient's clinical events (e.g., respiratory and cardiac events, limb movements, and arousals) using the American Academy of Sleep Medicine (AASM) standards [37].

Hypnogram is a form of PSG; it shows the sleep stages as a function of time. Figure 7 is a hypnogram presented in the NatusNeuroWorks®/SleepWorks™ interface [38]. Each sleep stage is annotated in 30-second epochs. From top to bottom, the sleep stages are the wake, REM, NREM1, NREM2, and NREM3. For this patient, 742 epochs were monitored during sleep. REM sleep happened between epoch 392 and epoch 485.

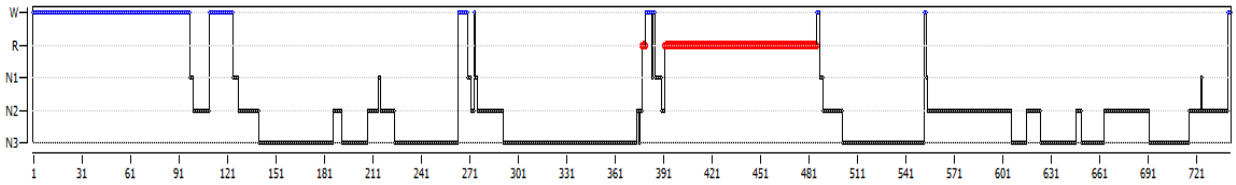


Figure 7. Hypnogram of an entire night exported from a PSG system

Figure 8 shows the example of an annotation displayed on the interface, which recorded all of the sleeper's events during the sleep study. These events include the time, title, and duration of an event, and corresponding sleep stage and epoch. In this example we can see that the patient changed body position to supine, and then changed to left. There will be some movement artifacts along with the change of body position. Movement artifact is generated by movement of cables or electrodes when patients

move normally in sleep. It usually solves itself when the patient stops moving. The technician should ensure that the electrodes are not displaced after the patient stops moving [39]. If there is still an artifact, the technician must repair the electrodes.

Time	Title	Epoch	Stage	Duration
00:12:49	Started Analyzer - Sleep Events	1	W	
00:12:54	Body Position: Supine	1	W	
00:13:54	Bed activated, remove EKG	3	W	
00:14:11	Artifact	4	W	0:07.8
00:15:06	Artifact	6	W	0:32.1
00:15:43	Body Position: Supine	7	W	
00:16:02	Artifact	8	W	0:16.6
00:16:27	O2: 3 L/min	9	W	
00:16:28	IP:15 cmH2O, EP:9 cmH2O	9	W	
00:28:18	Artifact	32	W	1:17.9
00:29:10	Body Position: Left	34	W	
00:30:03	Artifact	36	W	0:04.2
00:30:10	Artifact	36	W	0:13.5
00:30:52	Artifact	37	W	0:13.4
00:52:43	Artifact	81	W	0:15.6
00:59:30	Oxygen Desaturation (-4.0%)	95	W	1:36.0
01:01:01	Oxygen Desaturation (-5.0%)	98	N1	1:36.0

Figure 8. Annotation of sleep study

By exporting all the signals and annotations, we were able to obtain the comprehensive information and ground truth needed for our sleep study.

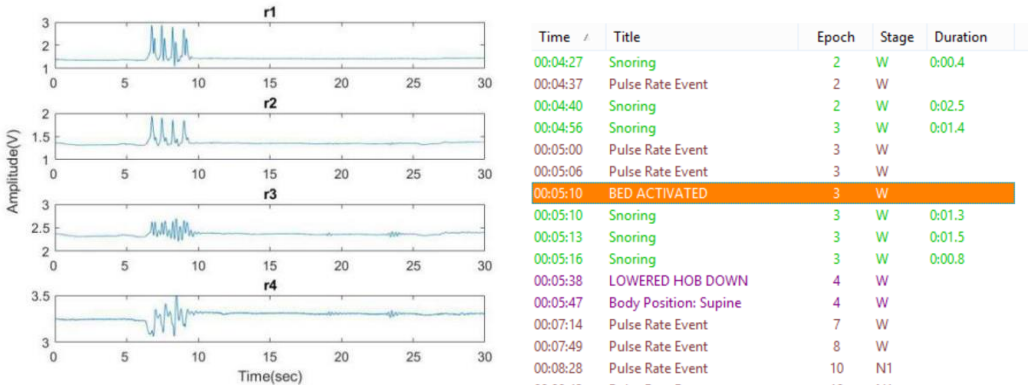
3.1.3 Data synchronization

3.1.3.1 Start time synchronization

As previously mentioned, to protect the privacy of patients participating in this experiment, all the data collected from the sleep lab was de-identified. We are not able to identify the subjects by their name. Furthermore, no matter when the data collection begins, the time stamp for the annotation starts at midnight (00:00:00). Since we collected the data using two independent systems simultaneously, we know the accurate

time stamp of the bed sensor system. Synchronizing the two system became an issue. We used two approaches to solve this issue. One approach is based on sudden variations in the signal of the two systems. This procedure was done with help of technicians in the sleep lab. After every test, all the devices were working properly. If the technician taps on a mattress three times, this action will be recorded in the annotation as an "active bed sensor."

The synchronization approach is to find the time stamp when the "three taps" occurred in the bed sensor signal, and then synchronize it with the time in which the "active bed sensor" became a part of the sleep lab annotation. Figure 9 is an example of the procedure where (a) shows the tapping action occurring at 22:22:25 in the bed sensor signal, and (b) shows the tapping action occurring at 00:05:10 in the de-identified annotation. In fact, the two time stamps represent the same moment in the two systems. This approach depends on signal variation at the beginning of the sleep study. At this time, the patient had just laid down on the bed, generating many movements and noise. Thus, it was difficult to distinguish the tapping action from the other noisy signals.



Time	Title	Epoch	Stage	Duration
00:04:27	Snoring	2	W	0:00.4
00:04:37	Pulse Rate Event	2	W	
00:04:40	Snoring	2	W	0:02.5
00:04:56	Snoring	3	W	0:01.4
00:05:00	Pulse Rate Event	3	W	
00:05:06	Pulse Rate Event	3	W	
00:05:10	BED ACTIVATED	3	W	
00:05:10	Snoring	3	W	0:01.3
00:05:13	Snoring	3	W	0:01.5
00:05:16	Snoring	3	W	0:00.8
00:05:38	LOWERED HOB DOWN	4	W	
00:05:47	Body Position: Supine	4	W	
00:07:14	Pulse Rate Event	7	W	
00:07:49	Pulse Rate Event	8	W	
00:08:28	Pulse Rate Event	10	N1	

Figure 9. Tapping motion in bed sensor signal and corresponding annotation

The second approach depends on the entire night's signals, aligning the motion annotation to the corresponding events occurring in the bed sensor signals. Figure 10 shows one of the alignment approaches based on "bathroom visit" annotation with the rapid decrease average DC value of the bed sensor data. Obviously, the bed sensor signals generated by weight pressure on the sensors when the patient gets out of the bed caused the sensors to capture the weight changes and decrease rapidly.

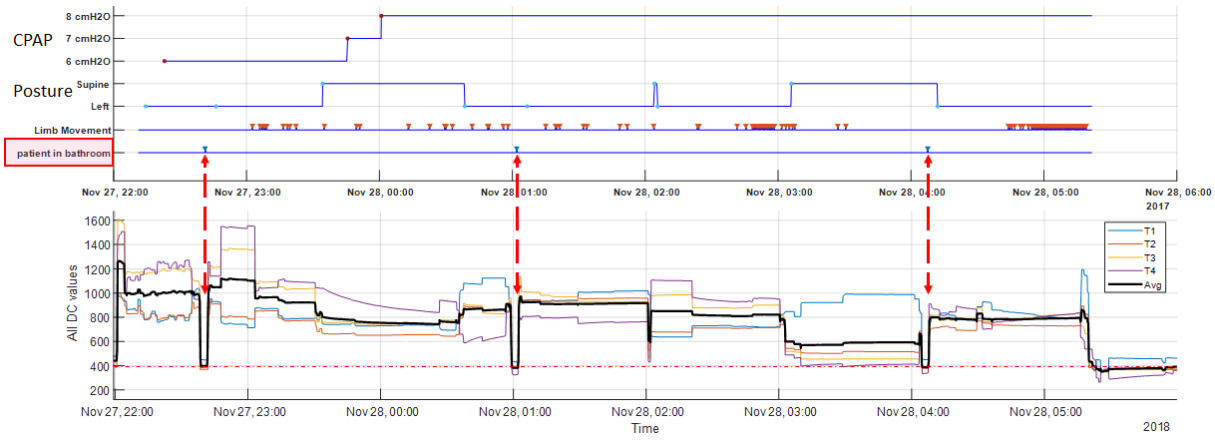


Figure 10. Time synchronization by aligning the "Bathroom visit" annotations to the "Low DC value" of the bed sensor.

This approach is not applicable in all situations. For example, Figure 11 shows the

subject that did not get out of the bed for the entire night. In this case, we use posture changes to align the data. The changes in the bed sensor signal does not always match the posture changes. For example, the first DC value change happened around 01:20 am, the posture also changed around that time. However, the third DC value change happened around 01:40 am, there is a delay in posture change.

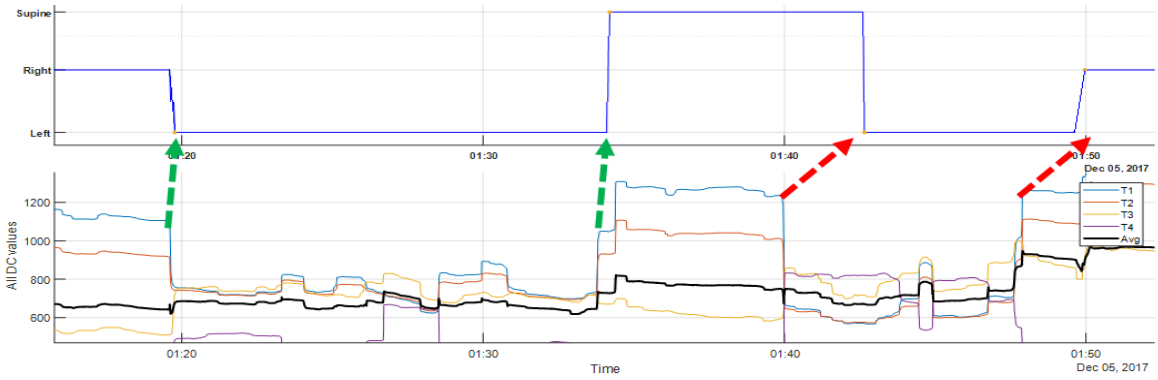


Figure 11. Aligning the data using posture changes

Since both approaches are not applicable in some situations, we alternately applied two approaches to improve the synchronization accuracy.

3.1.3.2 Break time synchronization

Although we synchronized the starting time of two systems, there were some interruptions during the sleep. Forty-seven out of 77 patients went to the restroom during the study. The number of restroom visits varied from one to five times. When a patient gets out of the bed, the technician has to disconnect all the devices worn on that person's body surface and interrupt the PSG data collection. At the same time, the bed sensors are supposed to stop collecting the data.

The first interruption to the flow of signals was when the subjects got out of bed but the bed sensors kept on collecting data. Figure 12 shows an annotation example of

the PSG system with the epoch number labeled at the top left of the beginning of each epoch. As can be seen in the figure, during epoch 55 the system was disconnected, and then reconnected during epoch 62.

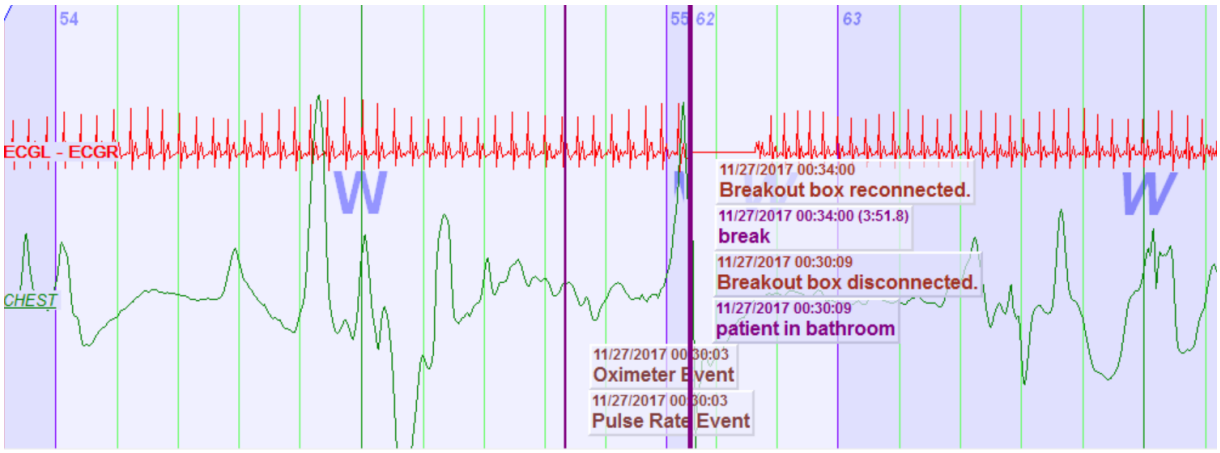


Figure 12. Annotation example of PSG system with break

According to the results of data synchronization, the restroom visit timestamp in the bed sensor system is from 22:38:19 to 22:42:10. However, due to a hardware malfunction, from Figure 13 we can see that the bed sensors kept on collecting data while the patient was away from the bed.

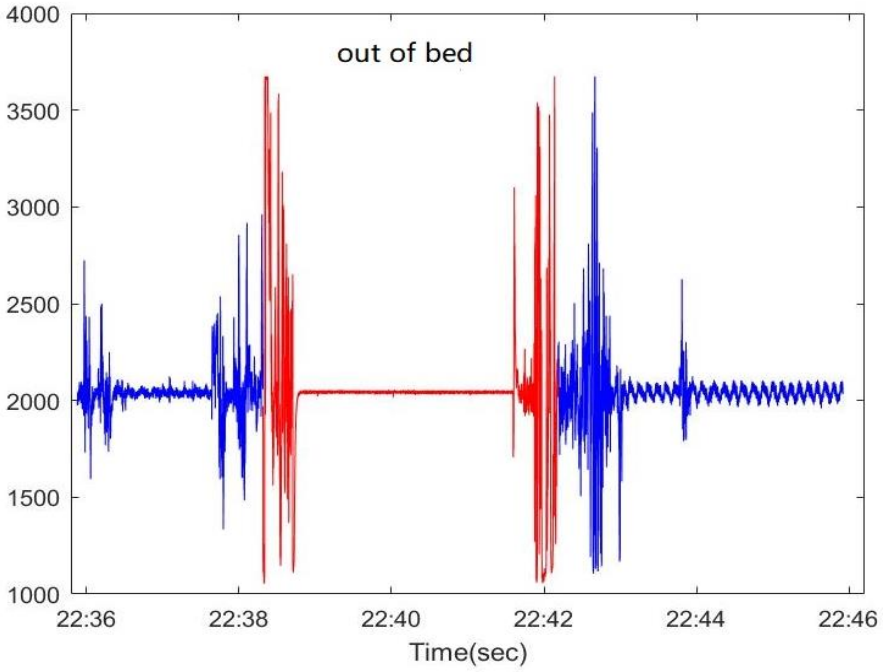


Figure 13. Example of bed sensor collecting data with subject out of bed

Since sleep stage is based on 30-second epochs, we don't want to keep incomplete sleep stages. In the PSG system we look for complete epochs before the disconnected time point. In this example, the complete epoch before disconnection is epoch 54, and the complete epoch after the reconnected time point is epoch 63. The incomplete epochs with the yellow background in Figure 14 were removed. For the bed sensor system, we removed all the data between epoch 54 and epoch 63 as shown in Figure 15. The remaining signals are the integral multiple segments of 30-second epochs. We then concatenated all the signals together.

Figure 14. Incomplete epoch removed from PSG system

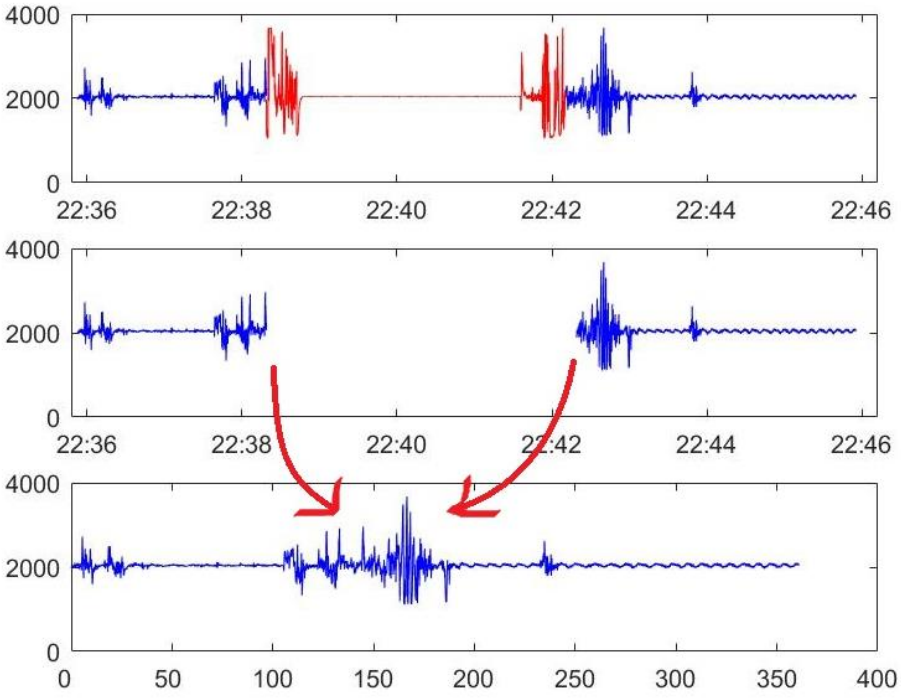


Figure 15. Incomplete epoch removed from bed sensor signal

The second exception was caused by a hardware malfunction. The bed sensor missed some part of the data, even when the patient was on his or her bed. Figure 16 is an example of this problem. According to this annotation, a patient was out of bed for about six minutes; however, there was an interval of about 13 minutes between the patient's out of bed time and the time signals could be converted to data. For some

hardware reason, the bed sensors experienced a seven-minute delay before they could start collecting data again.

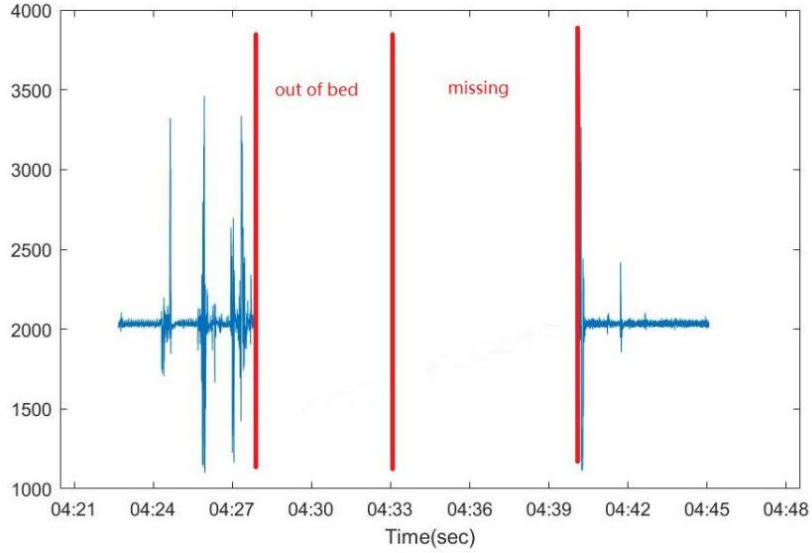


Figure 16. Bed sensor signals with missing data

However, in the PSG system, sleep stages were labeled during this time period. The synchronization process was much like the previous one, and the complete epoch was found to be coincident with the timestamp data. Then, the sleep stage labels were removed, which matched the missing data. Figure 17 shows the corresponding sleep stage data. Epoch 820 to epoch 845 had to be removed from the sleep stage label.

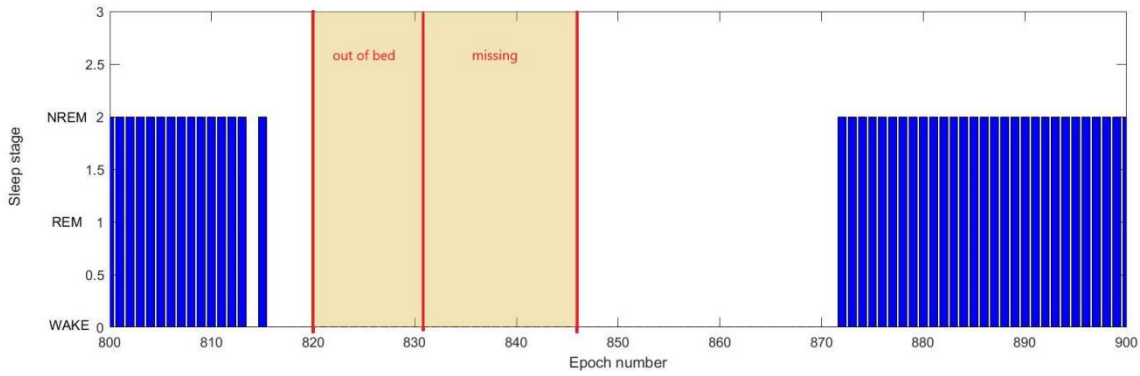


Figure 17. Corresponding data to the missing epoch in sleep stage

3.1.4 Data collection

Seventy-seven subjects (50 males, 27 females; mean age 62.8 ± 11.4 years) participated in the sleep study. Thirty-seven of the records for these patients were removed due to poor sleep quality, such as no REM sleep or staying awake for the entire night. Some were removed because the patients wore a pacemaker which can affect heartbeat detection. Other records were removed due to patients sleeping in the wrong position, e.g., sleeping with their head on the other side of the bed.

Those who participated in the sleep study had been diagnosed with a possible sleep disorder. The complete collapse and partial collapse of the airway are called apnea and hypopnea, respectively. The absence of airflow will affect breathing patterns and then influence sleep stage classification. It is necessary to remove patients with serious symptoms. The Apnea-hypopnea index (AHI) represents the number of apnea and hypopnea events per hour during the sleep. According to the AASM, the mild sleep apnea AHI is between 5 and 15. Based on the AHI provided by the PSG diagnostic report exported from NatusNeuroWorks®/SleepWorks™, we eliminated seven other subjects that had an AHI higher than 15. Table 2 shows 33 subjects after preliminary selection. An epoch number is given for each sleep stage representing each subject.

Table 2. Subjects with proportion of each sleep stage after preliminary selection

Stage Subject	WAKE	REM	NREM 1	NREM 2	NREM 3	Total	WAKE(%)	REM(%)	NREM(%)
20171127	162	83	53	252	260	810	20.00%	10.25%	69.75%
20171128	123	139	61	280	190	793	15.51%	17.53%	66.96%
20171205	64	81	88	327	239	799	8.01%	10.14%	81.85%
20171214	141	121	105	317	125	809	17.43%	14.96%	67.61%
20171218	272	69	66	184	168	759	35.84%	9.09%	55.07%
20171227	166	88	39	339	88	720	23.06%	12.22%	64.72%
20180109	130	121	37	326	137	751	17.31%	16.11%	66.58%
20180110	296	146	34	192	197	865	34.22%	16.88%	48.90%
20180131	333	142	38	212	152	877	37.97%	16.19%	45.84%
20180208	117	96	18	184	306	721	16.23%	13.31%	70.46%
20180213	80	175	37	470	82	844	9.48%	20.73%	69.79%
20180228	298	54	63	262	92	769	38.75%	7.02%	54.23%
20180312	72	115	51	368	240	846	8.51%	13.59%	77.90%
20180313	60	185	51	289	120	705	8.51%	26.24%	65.25%
20180314	88	230	53	331	218	920	9.57%	25.00%	65.43%
20180319	203	133	44	295	266	941	21.57%	14.13%	64.29%
20180320	135	150	69	437	112	903	14.95%	16.61%	68.44%
20180327	130	149	41	181	215	716	18.16%	20.81%	61.03%
20180418	291	106	75	222	168	862	33.76%	12.30%	53.94%
20180419	90	139	59	378	158	824	10.92%	16.87%	72.21%
20180423	366	57	32	160	164	779	46.98%	7.32%	45.70%
20180425	214	140	59	195	109	717	29.85%	19.53%	50.63%
20180426	155	183	98	183	214	833	18.61%	21.97%	59.42%
20180430	240	83	36	315	131	805	29.81%	10.31%	59.88%
20180501	134	134	130	363	120	881	15.21%	15.21%	69.58%
20180521	234	80	48	347	42	751	31.16%	10.65%	58.19%
20180522	94	153	64	446	24	781	12.04%	19.59%	68.37%
20180523	133	136	42	341	239	891	14.93%	15.26%	69.81%
20180530	187	102	35	163	243	730	25.62%	13.97%	60.41%
20180605	236	79	72	160	267	814	28.99%	9.71%	61.30%
20180607	219	133	49	231	136	768	28.52%	17.32%	54.17%
20180612	148	131	64	283	109	735	20.14%	17.82%	62.04%
20180618	226	116	28	151	251	772	29.27%	15.03%	55.70%

3.2 Signal processing

Although the bed sensor has built-in hardware filtering, the filtered signal is still far from being ready to use in our experiment. In this section, a further filtering method

is proposed to separate the heart rate and the respiration signal.

3.2.1 Bed sensor signal filtering

As shown in Section 3.1.1, each channel of hardware filtered BCG consists of heartbeats, respiration, and noise. For the purpose of extracting the HRV and RV features, further filtering was applied to the hardware-filtered BCG signal to separately obtain a clean heart rate and respiration signal.

The Butterworth bandpass filter of the 6th order was implemented in this work [10]. The normal respiration rate was lower than 0.5 Hz. The filter with a cutoff range from 0.7 Hz to 10 Hz removed most of the low-frequency respiration components. So, the higher frequency heart rate component was retained.

In order to remove the high-frequency heart rate component, a low pass 6th order Butterworth filter with a cutoff frequency of 0.7 Hz was run on hardware filtered BCG [12]. Figure 18 is an example of a filtered heart rate and respiration signal in 30-second epochs.

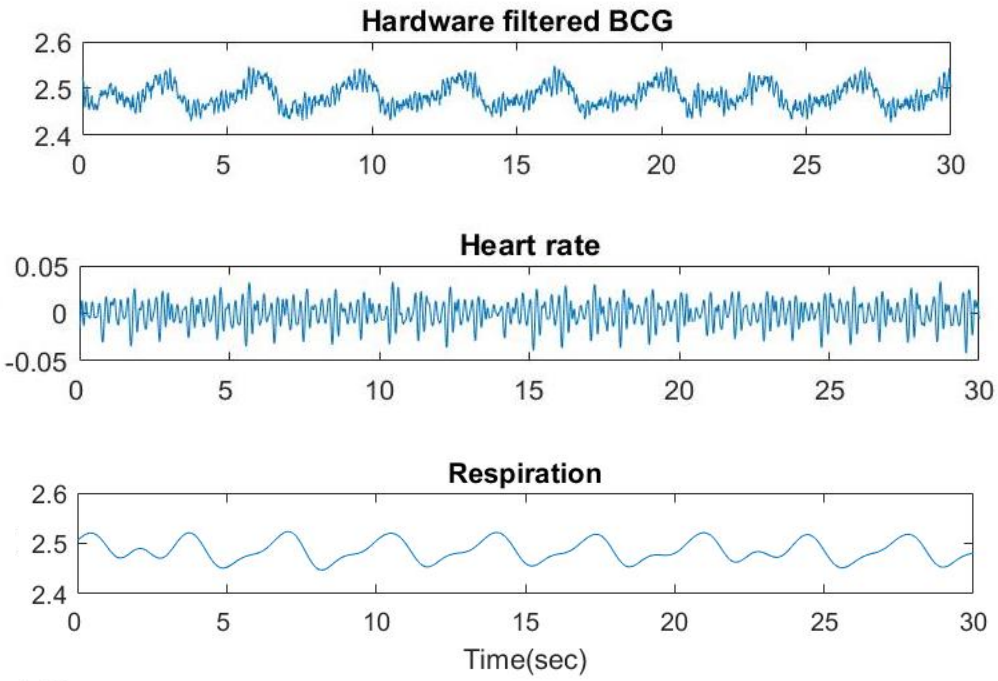


Figure 18. Filtered heart rate and respiration signal in 30-second epochs

3.2.2 Heart beat detection

In Section 3.2.1, the heart rate signal was separated from the BCG signal. An algorithm proposed in [10] was implemented in this experiment to detect the heart beat. For each 30-second epoch, the algorithm detects the heart beat based on an energy of 0.3 seconds in a moving window. The red circles above the signal in Figure 19 are the heartbeat detection results of a 30-second epoch. For this subject, 36 heart beats were detected in one epoch.

The normal heart rate range is between 60 and 100 bpm. Since the bed sensor is sensitive to body movement, much noise overlaps the BCG signal. This results in an inaccurate detection of beats. Therefore, any epoch with heart rates less than 40 bpm or larger than 150 bpm was removed.

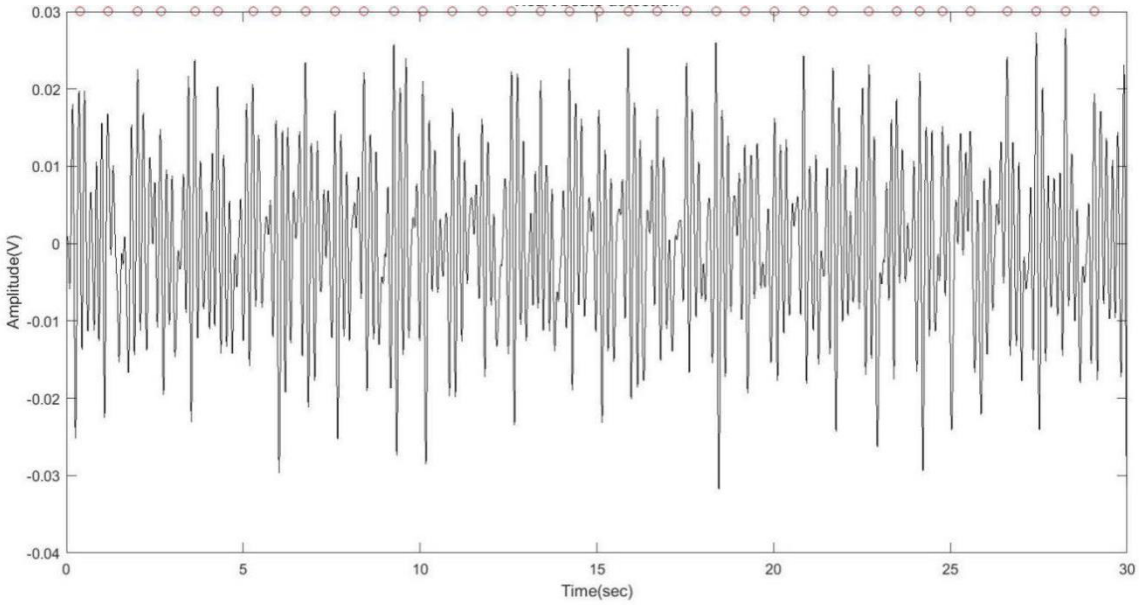


Figure 19. Heart beat detection for one epoch using algorithm from[10]. The red circles at the top are the detected heartbeats

If we define x(n) as the time stamp of heart beats in a 30-second epoch, and n is the number of beats, Figure 20 shows the beat locations with respect to beat number. A nearly straight line indicates the subject's heart beat in this epoch is regular.

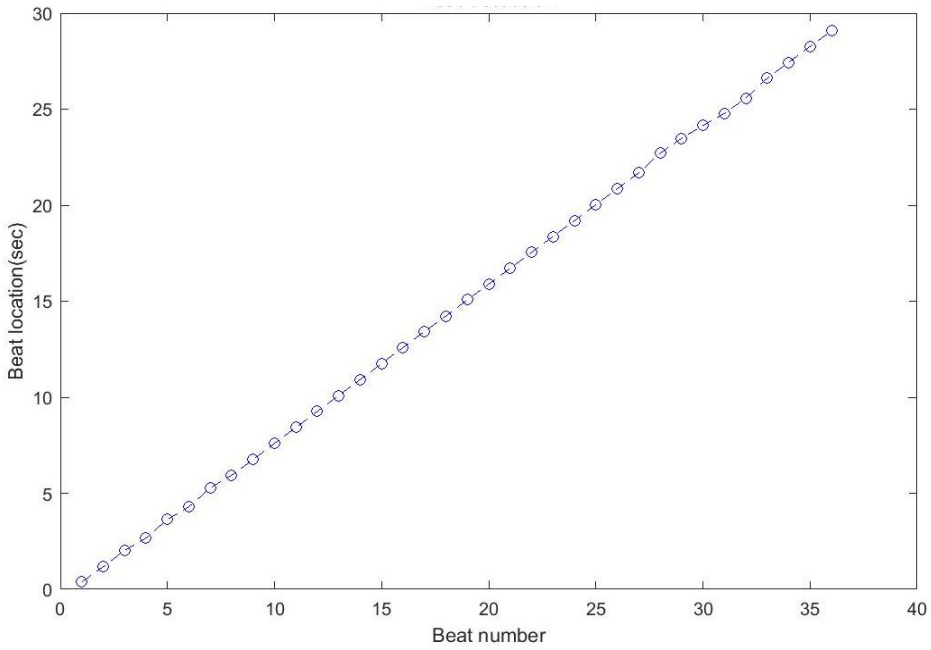


Figure 20. Heart beat location with respect to beat number for the epoch shown in Figure 19

3.2.3 Respiration peak detection

In Section 3.2.1, the respiratory signal is separated from the heart beat signal. The peaks and troughs of the respiratory signal can be detected easily. Respiratory peaks and troughs detection for one epoch signal is shown in Figure 21. The red and blue circles are detected peaks and troughs respectively.

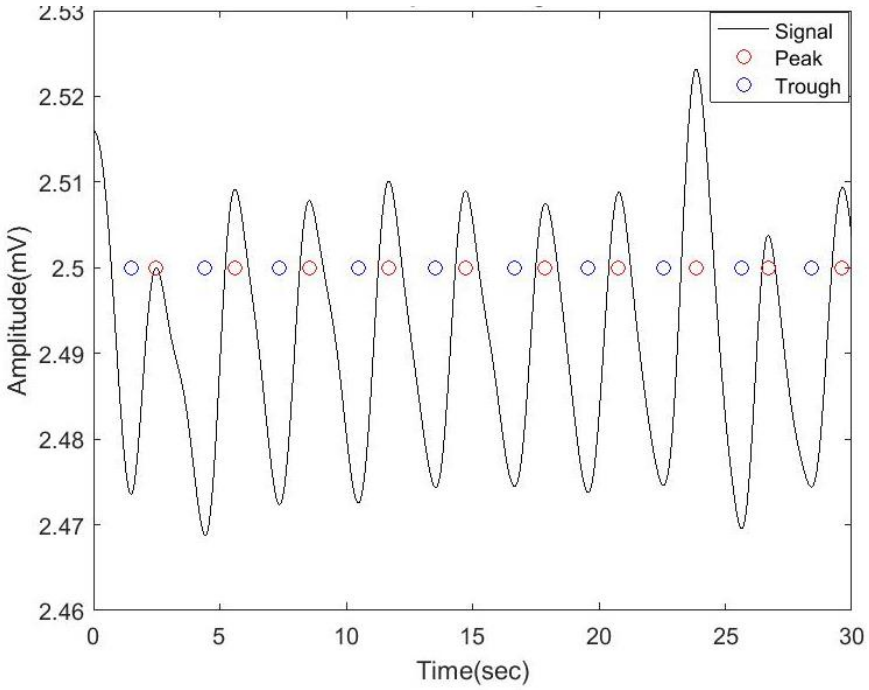


Figure 21. Respiratory peaks and troughs detection for the epoch shown in Figure 19. The red and blue circles are detected peaks and troughs

For the respiration signal, default peaks and troughs appear between each other which means exhalation and inhalation appear between each other. There are different types of exceptions in the respiration signal. Figure 22 shows one of the exceptions where two successive troughs are detected due to the shallow breath between two normal breaths. So, we find a peak between two troughs and insert the peak between them.

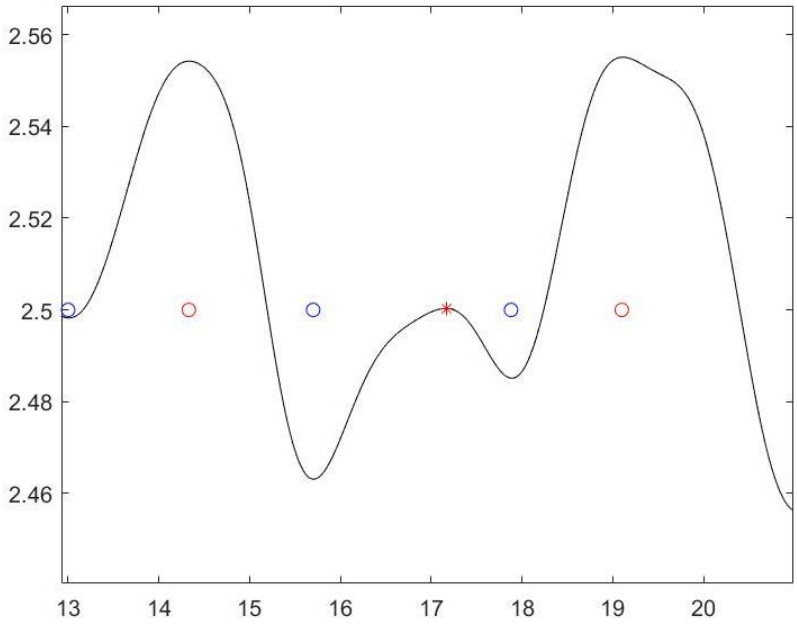


Figure 22. Insert one peak between two successive troughs

3.3 Feature extraction

3.3.1 Heart rate variability (HRV) features

After the signal processing step, heart beat locations were obtained to further determine the HRV feature and heart beat intervals (HBI). The heart beat interval is defined as the difference in successive heart beats from their location:

$$I(n) = x(n+1) - x(n) \;\; n = 1, 2, 3 \ldots \qquad (3\text{-}1)$$

Figure 23 plots the heart beat intervals corresponding to the heart beat detection shown in Figure 19.

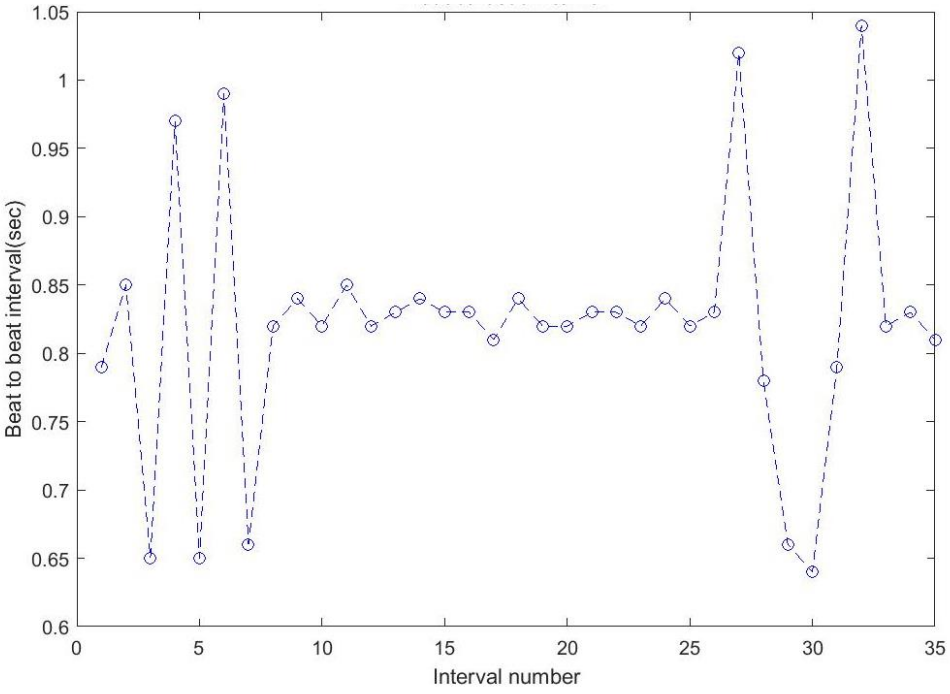


Figure 23. Heart beat intervals for the epoch shown in Figure 19

The difference in successive beat-to-beat intervals is defined as:

$$D(n) = I(n+1) - I(n) \;\; n = 1, 2, 3 \ldots \qquad (3\text{-}2)$$

Figure 24 plots the differences in successive beat-to-beat intervals corresponding to the heart beat intervals shown in Figure 23. Point number 8 to number 25 are around zero, which mean heart rate in this time period is stable.

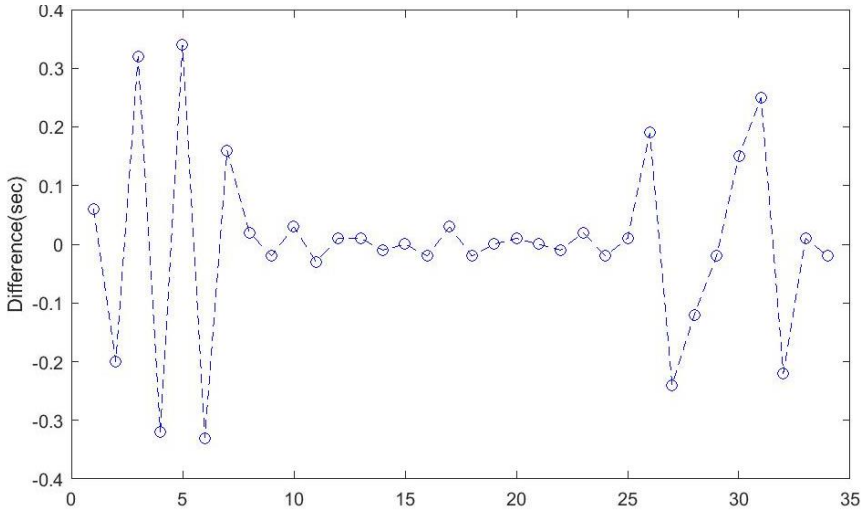


Figure 24. Successive difference of heart beat intervals for the epoch shown in Figure 19

HRV features can be derived from two different domains, the time domain and frequency domain.

3.3.1.1 Time domain HRV features

Time domain HRV features are the statistical measures derived from the heart beat intervals and differences in successive heart beats intervals. The definition and formula are as follow:

- Square root of the mean of the squares for successive difference in interval can be calculated as

$$\text{RMSSD} = \sqrt{\frac{\sum_{n=1}^{N} D(n)^2}{N}} \tag{3-3}$$

- The percentage of successive difference in interval > 50 ms is given as

$$\text{PNN50} = \text{P}(|\text{D(n)}| > 50\text{ms}) \tag{3-4}$$

  Where P means percentage.

- Mean of heart beat intervals is expressed as

$$\text{mHBI} = \frac{\sum_{n=1}^{N} I(n)}{N} \tag{3-5}$$

- Standard deviation of successive difference of beat-to-beat intervals is written as

$$\text{SDSD} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}|D(n)-\mu|^2} \quad (3\text{-}6)$$

$$\mu = \frac{\sum_{n=1}^{N} D(n)}{N} \quad (3\text{-}7)$$

- Standard deviation of beat-to-beat intervals is shown as

$$\text{SDNN} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}|I(n)-mHBI|^2} \quad (3\text{-}8)$$

- Maximum of beat-to-beat intervals is written as

$$\text{maxHBI} = \underset{1\le n\le N}{max} I(n) \quad (3\text{-}9)$$

- Minimum of beat-to-beat intervals is given as

$$\text{minHBI} = \underset{1\le n\le N}{mIN} I(n) \quad (3\text{-}10)$$

- Differences in maximum of beat-to-beat interval and minimum of beat to beat interval are shown as

$$\text{max\_minHBI} = \underset{1\le n\le N}{max} I(n) - \underset{1\le n\le N}{min} I(n) \quad (3\text{-}11)$$

- Coefficient of variance is expressed as

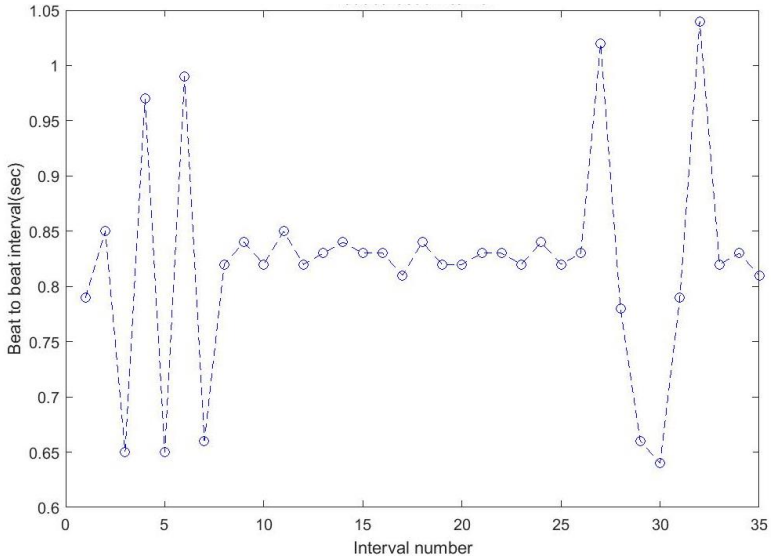
$$\text{CV} = \frac{SDNN}{mHBI} \quad (3\text{-}12)$$

3.3.1.2 Frequency domain HRV features

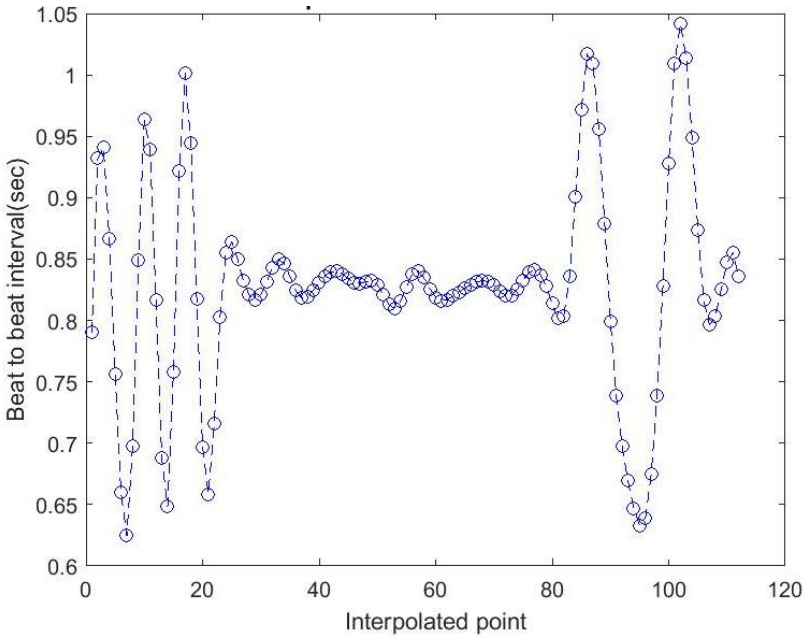
As mentioned before, the different frequency bands are related to parasympathetic and sympathetic nervous system activity. Moreover, the nervous system's active levels change during the different sleep stages. The frequency measures of the HBI series can be used to indicate the sleep stages.

Before calculating the power spectral density (PSD) of HBI, spline interpolation

with 4 Hz was applied on the unevenly-spaced HBI. Figure 25.(a) shows the original beat to beat interval from the result in (3-1). Figure 25.(b) shows the interpolated HBI and the data points, which are evenly distributed in 30-second epochs.



(a)



(b)

Figure 25. Heart beat intervals based on applied interpolation with 4 Hz for the epoch shown in Figure 19. (a) Before applied interpolation. (b) After applied interpolation

The next step is to compute the PSD of HBI for each epoch using discrete Fourier transform (DFT) with 256 sampling points. Figure 26 is the corresponding PSD of HBI in Figure 23.

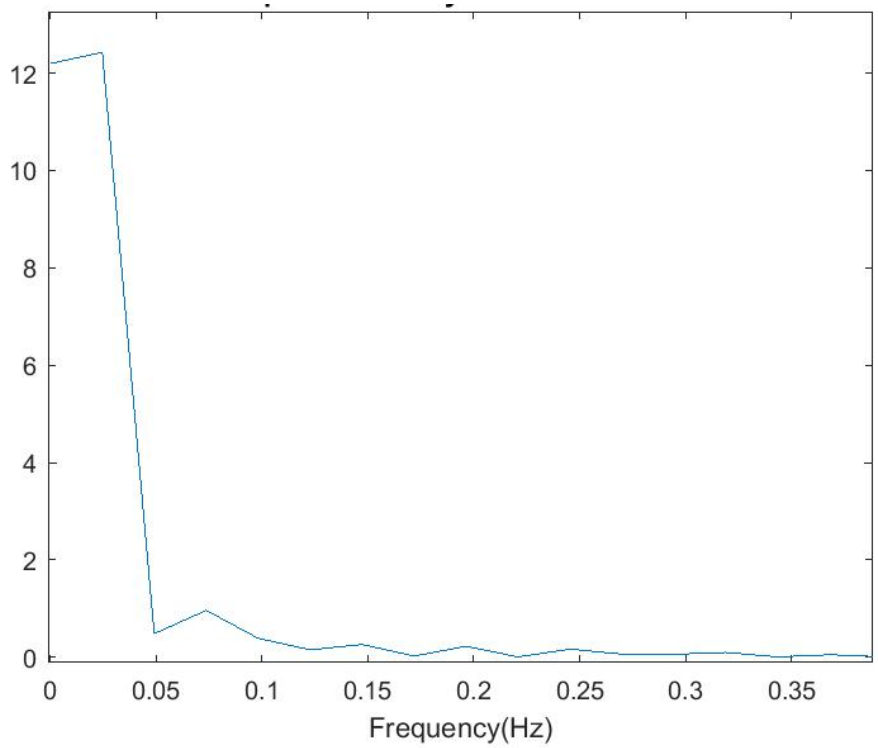


Figure 26. Power spectral density of heart beat interval for the epoch shown in Figure 19

The frequency band definition and frequency range are given in Table 3.

Table 3. Frequency band definition and frequency range

Frequency measures	**Frequency range**
Low frequency band(LF)	0.04-0.15Hz
High frequency band(HF)	0.15-0.4Hz
Total power(TF)	0-0.4Hz
Ratio of low frequency and high frequency: LF/HF	

3.3.2 Respiratory Variability (RV) Features

In Section 3.2.3 the respiration signal peaks and troughs detection were mentioned. Generally, the same procedures were followed for HRV features extraction

as the RV features were extracted from the breath-to-breath interval and from the successive differences of breath-to-breath intervals. The breath-to-breath intervals were defined as peak-to-peak intervals in this experiment. The time stamp of peak is x(n) and the breath-to-breath intervals are defined as:

$$I(n) = x_p(n+1) - x_p(n) \quad n = 1, 2, 3 \ldots \tag{3-13}$$

and the successive differences of breath-to-breath intervals are defined as:

$$D(n) = I(n+1) - I(n) \quad n = 1, 2, 3 \ldots \tag{3-14}$$

Different from HRV features, the respiration process can be divided into two parts, expiration and inspiration. The expiration is defined as trough-to-peak and the inspiration is defined as peak-to-trough. $I_{\mathrm{i}}(n)$ and $I_e(n)$ represent the interval of inspiration and expiration. $A_{\mathrm{i}}(n)$ and $A_e(n)$ represent the differences between the inspiration and expiration amplitude. The signal of each 30-second epoch can be divided into two types, the signal beginning with inspiration and the signal beginning with expiration.

If the signal begins with inspiration as shown in Figure 27(a), the inspiration and expiration intervals are defined as:

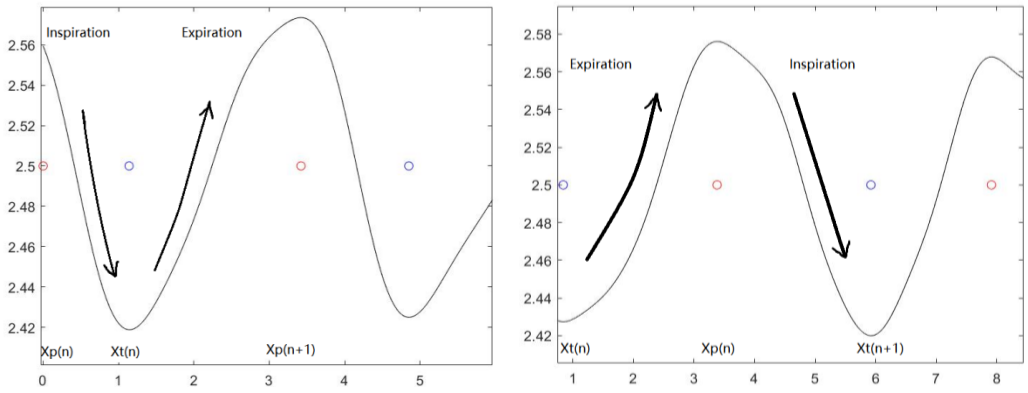
$$I_i(n) = x_t(n) - x_p(n) \quad n = 1, 2, 3 \ldots \tag{3-15}$$

$$I_e(n) = x_p(n) - x_t(n) \quad n = 1, 2, 3 \ldots \tag{3-16}$$

The difference of the amplitudes are defined as:

$$A_e(n) = A_p(n+1) - A_t(n) \quad n = 1, 2, 3 \ldots \tag{3-17}$$

$$A_i(n) = A_p(n) - A_t(n) \quad n = 1, 2, 3 \ldots \tag{3-18}$$



(a) (b)

Figure 27. Inspiration and expiration of two different signal types in one epoch

If the signal begins with expiration as shown in Figure 27(b), the inspiration and expiration intervals are defined as:

$$I_e(n) = x_p(n) - x_t(n) \quad n = 1, 2, 3 \ldots \tag{3-19}$$

$$I_i(n) = x_t(n+1) - x_p(n) \quad n = 1, 2, 3 \ldots \tag{3-20}$$

The differences of the amplitudes are defined as:

$$A_e(n) = A_p(n) - A_t(n) \quad n = 1, 2, 3 \ldots \tag{3-21}$$

$$A_i(n) = A_p(n) - A_t(n+1) \quad n = 1, 2, 3 \ldots \tag{3-22}$$

The RV features extracted from breath-to-breath intervals, successive differences of intervals, and intervals of inspiration and expiration are listed as follows:

- The square root of the mean of the squares of successive differences of breath-to-breath intervals is calculated as

$$\text{RMSSD} = \sqrt{\frac{\sum_{n=1}^{N} D(n)^2}{N}} \tag{3-23}$$

- The mean of successive differences of breath-to-breath intervals is expressed as

$$mDI = \frac{\sum_{n=1}^{N} D(n)}{N} \tag{3-24}$$

- The max of the absolute differences of breath-to-breath intervals is defined as

$$MADI = \max_{1 \le n \le N} |D(n)| \tag{3-25}$$

- The mean of respiratory rate can be found by

$$mRR = \frac{60}{N} \sum_{n=1}^{N} \frac{1}{I(n)} \tag{3-26}$$

- The standard deviation of respiratory rates is given as

$$\text{SDRR} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} |\frac{60}{I(n)} - mRR|^2} \tag{3-27}$$

- The coefficient of variance is defined as

$$\text{CV} = \frac{SDRR}{mRR} \tag{3-28}$$

- The median of respiratory rate is written as

$$\text{MedianRR} = Q_2(\frac{60}{I(n)}) \tag{3-29}$$

Where Q_2 means median

- The inter quartile range of respiratory rate can be found by the following calculation:

$$\text{IQR} = Q_3\left(\frac{60}{I(n)}\right) - Q_1\left(\frac{60}{I(n)}\right) \tag{3-30}$$

Where Q_1 and Q_3 are the first quartile and the third quartile respectively.

- The mean of the absolute deviation value of respiratory rates is expressed as

$$\text{MAD} = \frac{1}{N} \sum_{n=1}^{N} |\frac{60}{I(n)} - mRR| \tag{3-31}$$

- The ratio of the mean of differences between the amplitudes of expiration and

inspiration

$$\frac{\sum_{n=1}^{N} A_e(n)}{\sum_{n=1}^{N} A_i(n)} \quad (3\text{-}32)$$

- The ratio of mean of expiration intervals and inspiration intervals are defined as

$$\frac{\sum_{n=1}^{N} I_e(n)}{\sum_{n=1}^{N} I_i(n)} \quad (3\text{-}32)$$

3.3.3 Linear Frequency Cepstrum Coefficients (LFCC) features

The cepstrum is the result of an inverse Fourier transform of the logarithm of the signal spectrum. The mel frequency cepstral coefficient (MFCC) is defined as the real cepstrum of a windowed signal derived from FFT of the signal [40], and it has been widely used in voice recognition algorithms [41]. In some sleep studies, linear frequency cepstrum coefficients (LFCC) have been used to solve sleep stage classification and obstructive sleep apnea detection problems. In this study, the LFCC features of heart rate signals were extracted. The implementation steps are described as follows [42]:

- Frame 30-second each epoch signal into short time frames with a sliding window size of 2s (200 points) with an 80% overlap. For one epoch, there will be 71 short time frames.
- For each time frame, calculate the power spectrum. Figure 28 displays the power spectrum of the first time frame of one epoch.

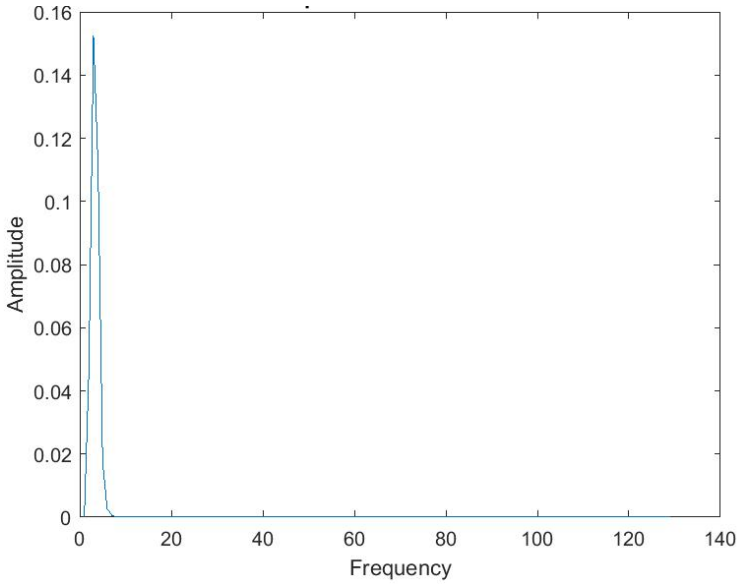


Figure 28. Power spectrum of the first time frame for the epoch shown in Figure 19

- Generate a linear filter bank. Generally, a filter bank consists of 20-40 (26 is standard) filters. The linear filter bank consists of standard 26 filters was applied in this thesis. Since the heart rate frequency ranges from 0.7 Hz to 10 Hz, to generate 26 filters, 28 points were linearly spaced between this frequency range and rounded to the nearest FFT bins calculated in the previous step. The filter bank with 26 triangular filters is shown in Figure 29.

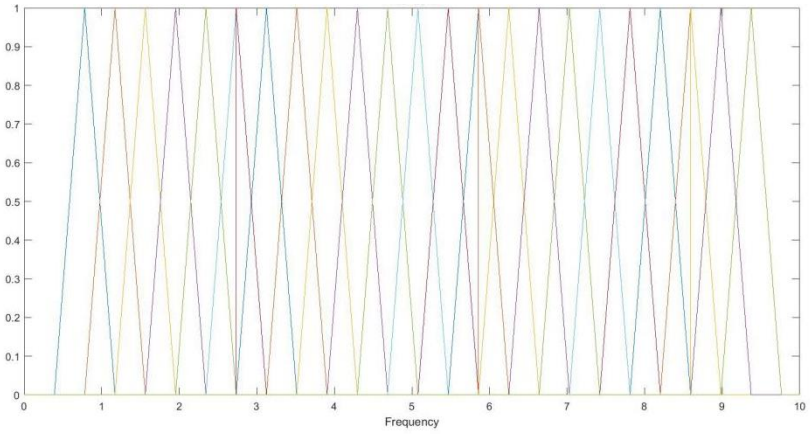


Figure 29. Linear filter bank with 26 triangular filters between frequency range 0.7Hz to 10Hz

- Apply the filter bank to the power spectra. For each time frame 26 filter bank energies were generated in total.
- Calculate all the logarithms of the filter bank energies. Figure 30 shows the 26 log filter bank energies in the first time frame.

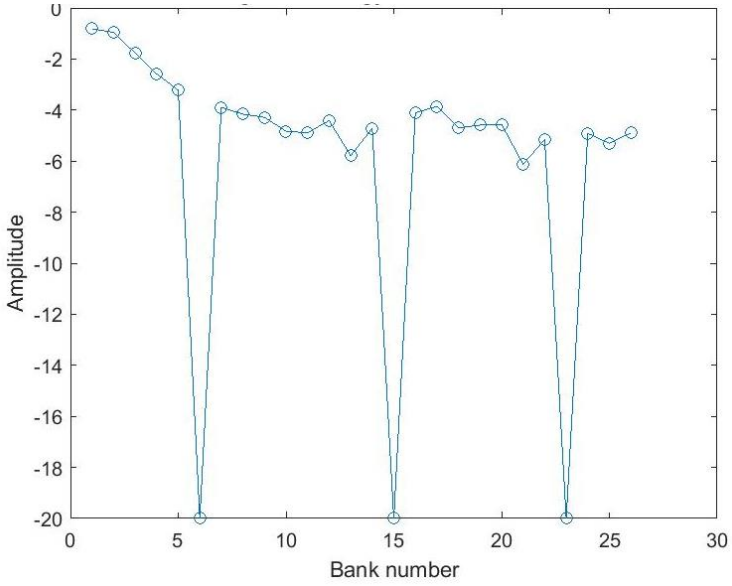


Figure 30. Twenty-six log filter bank energies of the first time frame for the epoch shown in Figure 19

- The discrete cosine transform (DCT) was applied on 26 log filter bank energies to generate 26 cepstral coefficients. To discard the first DC term for each time frame leaves 25 cepstral coefficients.
- For one epoch, with a window size of 2s and 80% overlapping, there were 71 short time frames. Compute the mean and standard deviation of the coefficient along the time frame. We obtained a total of 25 mean LFCCs and 25 standard deviation LFCCs for each epoch. Figure 31 and Figure 32 represent the mean LFCC and standard deviation LFCC, respectively.

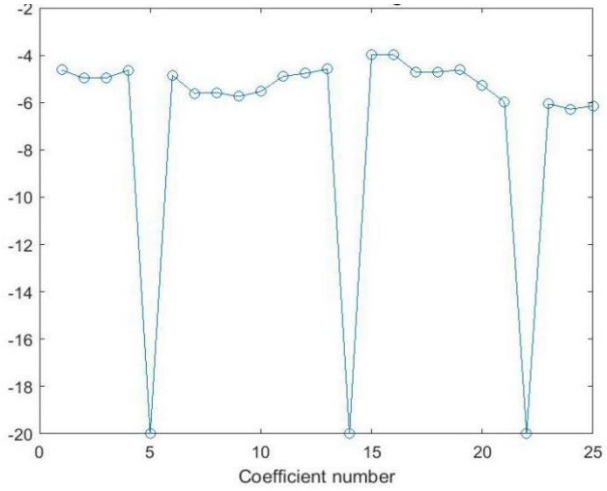


Figure 31. Twenty- five mean LFCCs along time frame for the epoch shown in Figure 19

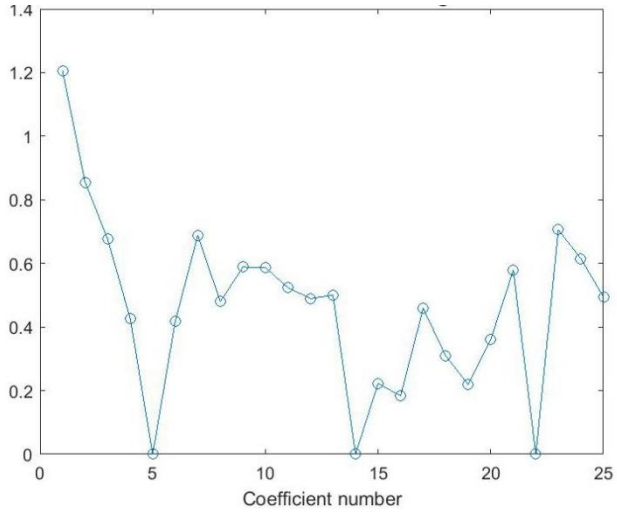


Figure 32. Twenty- five standard deviation LFCCs along time frame for the epoch shown in Figure 19

3.4 Support vector machine (SVM)

A support vector machine (SVM) is one of the discriminative classifiers implemented in this research. The mechanism of classification is finding a hyperplane that can maximize the margin between two classes.

Given data points:

$$\{(\mathbf{x}_i, d_i); i = 1 \dots N, d_i \in \{-1, +1\}\} \tag{3-33}$$

the hyperplane consists of data point **x**, which can be defined as:

$$\mathrm{H} = \{\boldsymbol{x}|g(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b = 0\} \tag{3-34}$$

where $\boldsymbol{w}$ is a normal vector perpendicular to the hyperplane.

Assuming that two classes are separable, the data points satisfy:

$$\boldsymbol{w}^T\boldsymbol{x}_i + b \geq 1 \quad d_i = 1 \tag{3-35}$$

$$\boldsymbol{w}^T\boldsymbol{x}_i + b \leq -1 \quad d_i = -1 \tag{3-36}$$

The margin width is the projection of difference between the two support vectors on the normal vector:

$$\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}(\boldsymbol{x}_2 - \boldsymbol{x}_1) = \frac{2}{\|\boldsymbol{w}\|} \tag{3-37}$$

In order to maximize the margin width:

$$\max \frac{2}{\|\boldsymbol{w}\|} \quad d_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 \quad i = 1, 2, \ldots N \tag{3-38}$$

The problem can be solved by using quadratic programming; thus,

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 \quad d_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 \quad i = 1, 2, \ldots N \tag{3-39}$$

The problem converts into solving:

$$\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j d_i d_j \boldsymbol{x}_i \boldsymbol{x}_j \tag{3-40}$$

$$\sum_{i=1}^{N} a_i d_i = 0 \qquad a_i \geq 0, \qquad i = 1,2 \ldots N$$

The hyperplane separates the data points into two non-overlapping classes. However, perfect separation is not common. In some situations, a linear hyperplane is not able to separate the classes. So, a kernel trick was introduced mapping the input space into a higher dimension linear separable feature space, where make linear classifiers are still applicable [43]. If $\Phi(\boldsymbol{x})$ is mapping function, in higher dimension

feature space the hyperplane can be expressed as:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \Phi(\boldsymbol{x}) + b \tag{3-41}$$

The quadratic programming problem becomes:

$$\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j d_i d_j \Phi(\boldsymbol{x_i})^T \Phi(\boldsymbol{x_j}) \tag{3-42}$$

$$\sum_{i=1}^{N} a_i d_i = 0 \qquad a_i \geq 0, \qquad i = 1,2 \ldots N$$

The kernel function $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \Phi(\boldsymbol{x_i})^T \Phi(\boldsymbol{x_j})$ calculates the dot product of mapped data points in feature space. In this thesis, two different types of kernels were implemented in experiments.

Gaussian kernel:

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = e^{-\frac{1}{2\sigma^2}\|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2} \tag{3-43}$$

Cubic kernel:

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i}^T \boldsymbol{x_j} + \theta)^3 \tag{3-44}$$

3.5 k-nearest neighbors (k-NN)

The k-nearest neighbors classification method was implemented in this research. The logic behind this algorithm is: The testing data point is classified by the majority vote of its neighbors [44]. Let $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)$ be a testing sample and $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)$ be a training sample. The distance between $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be defined as $d(\boldsymbol{X}, \boldsymbol{Y})$.

The commonly used distance functions are:

Euclidean distance $d(\boldsymbol{X}, \boldsymbol{Y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

and

$$\text{Manhattan distance } d(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=1}^{n} |\boldsymbol{x_i} - \boldsymbol{y_i}|$$

After calculating the distance between the testing data point and training points, select the k nearest distances. The selection of parameter k will affect the result. A larger k value could reduce the effects of noise, but make the boundaries of different classes unclear. The k-nearest neighbor classifier can be regarded as the k-nearest neighbors with $\frac{1}{k}$ weight and others with 0 weight. A refined k-NN classification called distance weighted k-NN defines the weight of each neighbor as inverse of the distance. In this research different types of distance metrics and different numbers of k were implemented.

CHAPTER 4. EXPERIMENTS

In the data collection section, 33 subjects remained after preliminary selection. Further subject selection was based on patients' sleep posture. The original purpose was to see the effect of sleep posture on sleep stages. Table 4 shows the selected subjects with all the three different types of sleep postures. Those subjects missing one or more of these postures were eliminated. Whenever the patient changes the sleep posture, there will be a corresponding annotation. The number for each posture represents the number of times this posture appears in the annotation.

Table 4. Nineteen subjects with three sleep postures after selection

subject	right	left	supine
20171128	3	1	3
20171205	1	6	5
20171214	3	2	4
20171227	5	1	6
20180110	4	4	6
20180228	4	7	5
20180319	3	3	3
20180320	6	1	4
20180418	2	5	5
20180423	4	4	7
20180425	2	5	4
20180430	2	4	2
20180522	4	2	3
20180523	1	1	2
20180530	3	1	4
20180605	1	3	1
20180607	1	2	3
20180612	1	4	2
20180618	5	1	4

Furthermore, five subjects were selected randomly for the experiments. Two

scenarios were applied in the experiments: The first was to put all the subjects together and the second one was leave one subject out. Inspired by [35], a hierarchical structure with several layers was implemented in some experiments. The HRV, RV, and LFCC features described in Section 3.3 were extracted for the experiments. Sleep stages NREM1, NREM2, and NREM3 were combined as NREM.

Table 5. Selected subjects with epoch number of each sleep stage

Stage / Subject	WAKE	REM	NREM 1	NREM 2	NREM 3	Total	WAKE(%)	REM(%)	NREM(%)
20180319	203	133	44	295	266	941	21.57%	14.13%	64.29%
20180320	135	150	69	437	112	903	14.95%	16.61%	68.44%
20180418	291	106	75	222	168	862	33.76%	12.30%	53.94%
20180423	366	57	32	160	164	779	46.98%	7.32%	45.70%
20180425	214	140	59	195	109	717	29.85%	19.53%	50.63%

4.1 Put all subjects together

In this scenario, 4202 epochs of the five subjects were put together. The goal was to classify three sleep stages: Wake, REM, and NREM. A 10-fold cross-validation was adopted.

4.1.1 Single classifier

4.1.1.1 Experiment 1: Three sleep stage classification using SVM

In this experiment, all 74 features were fed into a SVM classifier with cubic kernel. After applying 10-fold cross validation, the accuracy is 85.3% and the kappa value is 0.74, which is comparable with Jialei's result 81% in accuracy and 0.70 kappa value. The confusion matrix is shown in Table 6. Some of the wake epochs were misclassified as NREM. Table 7 intuitively indicates 23% of awake

misclassified as NREM. REM and NREM have 80% and 91% of true positive rates respectively.

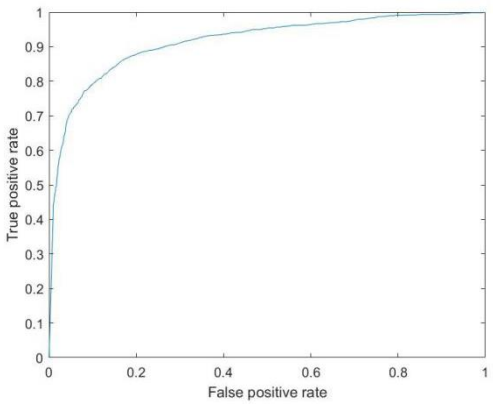
Table 6. Confusion matrix of three sleep stage classifications using cubic SVM classifier

Actual \ Predicted	Wake	REM	NREM
Wake	908	32	269
REM	29	471	86
NREM	161	41	2205

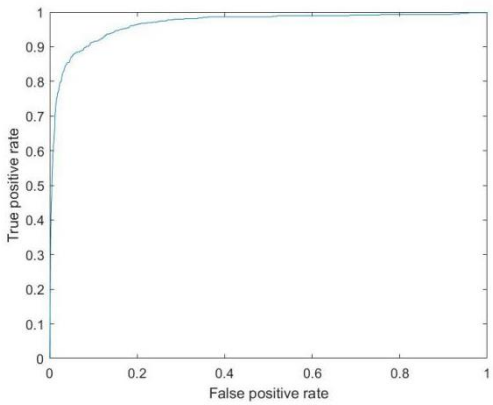
Table 7. True positive rates and false negative rates of three sleep stage classification using cubic SVM classifier

Actual \ Predicted	Wake	REM	NREM
Wake	75%	2%	23%
REM	5%	80%	15%
NREM	7%	2%	91%

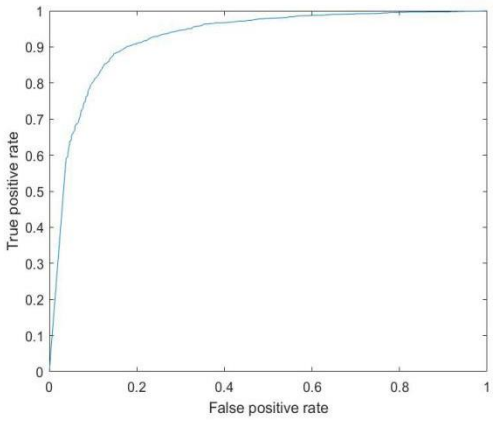
The receiver operating characteristic (ROC) and area under the ROC curve (AUC) were generated shown in Figure 33. For each curve, one of the sleep stage classes was regarded as the positive class; the other two classes were regarded as the negative class. Although the NREM has highest true positive rate, the ROC curve and AUC of REM shows best performance.



(a)ROC curve for wake is positive class, AUC=0. 92



(b) ROC curve for REM is positive class, AUC=0.97



(c)ROC curve for NREM is positive class, AUC=0.92

Figure 33. ROC and AUC of one sleep stage regarded as positive class, other two sleep stages regarded as negative class.

Since wake and NREM stages have different characteristic, the 23% misclassification of wake as NREM was not expected. Besides, 15% of REM misclassified as NREM. So we separated the NREM into light and deep sleep. Table 8 shows the true positive rates and false negative rates. 21% of wake and 11% of REM misclassified as light sleep. It indicates there are some similarities between wake and light sleep. Compared with REM and deep sleep, people are more likely to wake up and have more movements in the light sleep. Table 9 shows that light sleep accounts

for a large percentage of NREM, even entire night. Therefore, misclassification of light sleep has an impact on classification result.

Table 8. True positive rates and false negative rates of four sleep stage classification using cubic SVM classifier

Actual \ Predicted	Wake	REM	Light	Deep
Wake	75%	2%	21%	2%
REM	6%	81%	11%	2%
Light	10%	2%	74%	14%
Deep	2%	1%	36%	60%

Table 9. Selected subjects with percentage of four different sleep stages

Stage / Subject	WAKE(%)	REM(%)	Light(%)	Deep(%)
20180319	21.57%	14.13%	36.03%	28.27%
20180320	14.95%	16.61%	56.04%	12.40%
20180418	33.76%	12.30%	34.45%	19.49%
20180423	46.98%	7.32%	24.65%	21.05%
20180425	29.85%	19.53%	35.43%	15.20%

4.1.1.2 Experiment 2: Three sleep stage classifications using k-NN

In this experiment, all the features were fed into the k-NN classifier. The odd number of neighbors, k, varies from 1 to 19. The Euclidean distance metric and equal distance weight were implemented. Figure 38 and Figure 39 show the accuracy and kappa value changes with varied numbers of neighbor k. The classifier showed the

best performance with five nearest neighbors with the accuracy reaching 83.7% and the kappa value reaching 0.71.

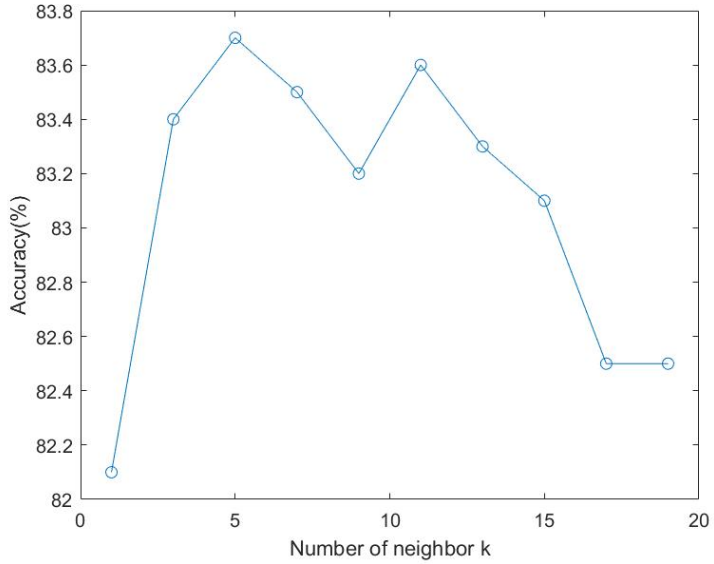


Figure 34. Accuracy with different odd numbers of neighbor k

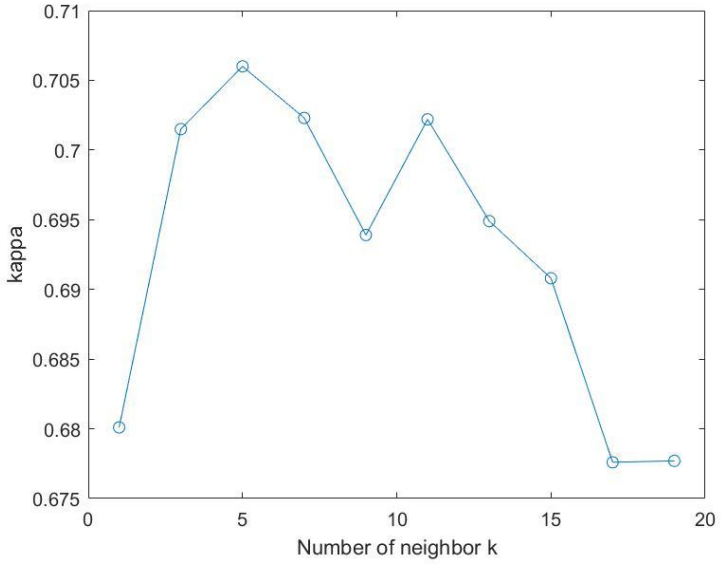


Figure 35. Kappa value with different odd numbers of neighbor k

Table 10. Confusion matrix of three sleep stage classifications using k-NN classifier

Actual \ Predicted	Wake	REM	NREM
Wake	901	29	279
REM	41	432	113
NREM	164	57	2186

Table 11. True positive rates and false negative rates of three sleep stage classification using k-NN classifier

Actual \ Predicted	Wake	REM	NREM
Wake	75%	2%	23%
REM	7%	74%	19%
NREM	7%	2%	91%

The only difference compared to result from experiment 1 is the true positive rate and false negative rates of REM detection. The true positive rate decreased to 74% from 80%, more REM epochs were misclassified as NREM.

4.1.1.3 Experiment 3: Three sleep stages classification with selected features

In this experiment, the importance of features was computed from a decision tree. The importance estimates were calculated as the sum of risk changes caused by a split on each feature divided by the number of branch nodes. Figure 36 shows the importance estimates in the features' number order. Numbers 1–13 are HRV features, numbers 14–24 are RV features, numbers 25–74 are LFCC features. In general, LFCC and RV features have a higher importance than HRV features.

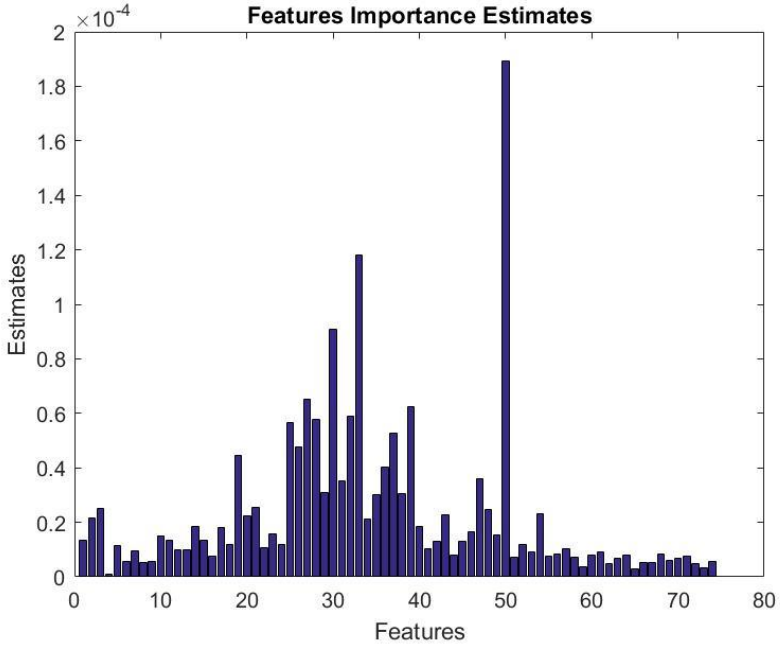
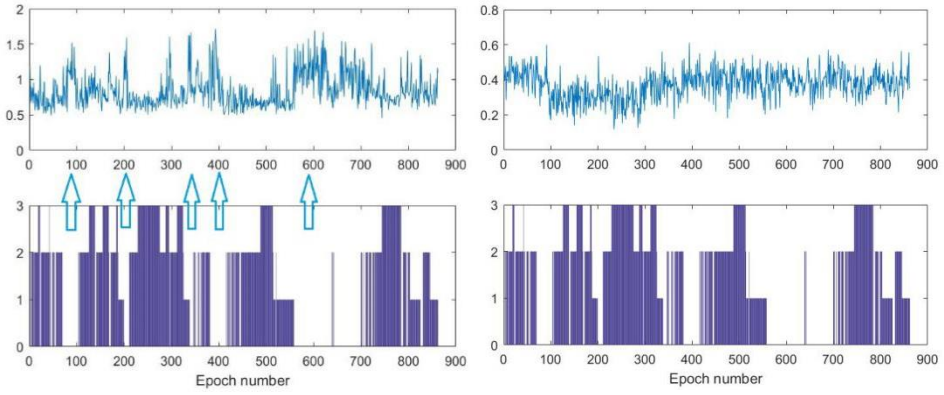


Figure 36. Importance estimates of features



(a) (b)

Figure 37. Features with two different importance estimates for entire night and corresponding sleep stages (0: wake; 1: REM; 2: Light; 3: Deep). (a) Feature number 50 with highest importance estimate and sleep stages. (b) Feature number 4 with least importance estimate and sleep stages.

Figure 37 shows the features with two different level of importance. The feature number 50 on the left has the highest importance estimate. The value changes obviously as the sleep stages change. Especially during wake, there are spikes corresponding to wake. The feature number 4 on the right has least importance.

Compared with feature number 50, the features value didn't show variation as the sleep stages change. So it make sense to have least importance.

Figure 38 is the correlation graph. Compared with the features' importance, the features with lower importance have higher correlation.

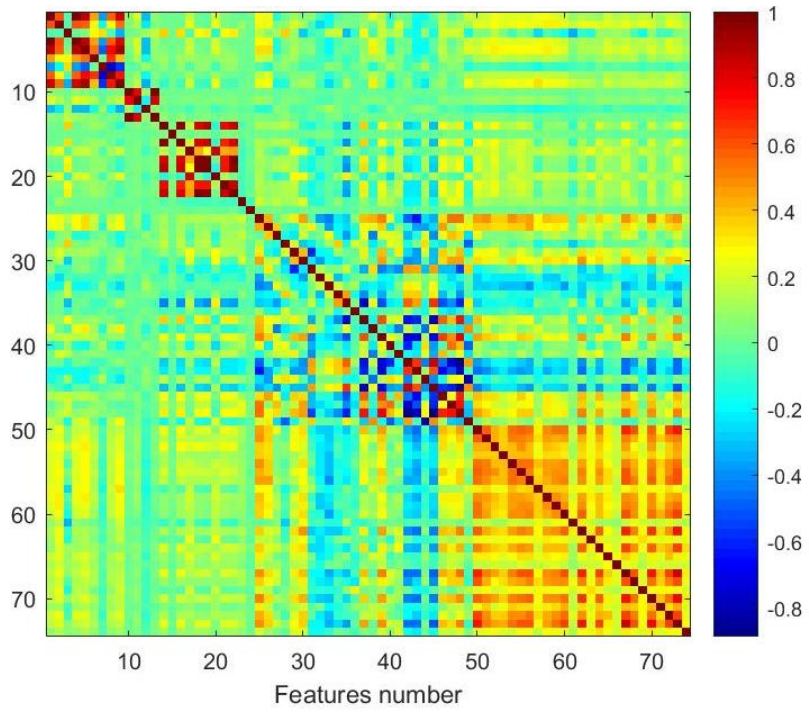


Figure 38. Correlation graph

Fifteen features were selected with the highest importance estimates and fed into the SVM classifier with the cubic kernel. The accuracy and kappa value are 85.4% and 0.74, respectively. Since there was slightly improvement in accuracy, the selected features contained most of the information. From the confusion matrix and true positive rates shown as follow, we can see the true positive rate of REM improved from 80% to 82%. The selected features show better performance in detecting REM.

Table 12. Confusion matrix of three sleep stage classifications using cubic SVM with fifteen selected features

Actual \ Predicted	Wake	REM	NREM
Wake	907	24	278
REM	23	482	81
NREM	139	67	2201

Table 13. True positive rates and false negative rates of three sleep stage classification using cubic SVM with fifteen selected features

Actual \ Predicted	Wake	REM	NREM
Wake	75%	2%	23%
REM	4%	82%	14%
NREM	6%	3%	91%

4.1.1.3 Experiment 4: REM stage detection using SVM classifier

In this experiment Wake and NREM are combined together as one class, the purpose is separate REM from the other classes. All 74 features were fed into the SVM classifier with the cubic kernel and 10-fold cross-validation was applied. The accuracy is 95.3%, kappa value is 0.80. From the confusion matrix we can see that these two classes are imbalanced, only 14% of sleep stages are REM.

Table 14. Confusion matrix of REM detection using cubic SVM

Actual \ Predicted	REM	Wake&NREM
REM	467	119
Wake&NREM	78	3538

The true positive rate of REM is 80%. The true positive rate of detecting combination class is 98%, which made contribution to overall accuracy. The ROC curve approach to left top and AUC is 0.97 indicates the classifier has good performance.

Table 15. True positive rates and false negative rates of REM detection using cubic SVM

Predicted / Actual	REM	Wake&NREM
REM	80%	20%
Wake&NREM	2%	98%

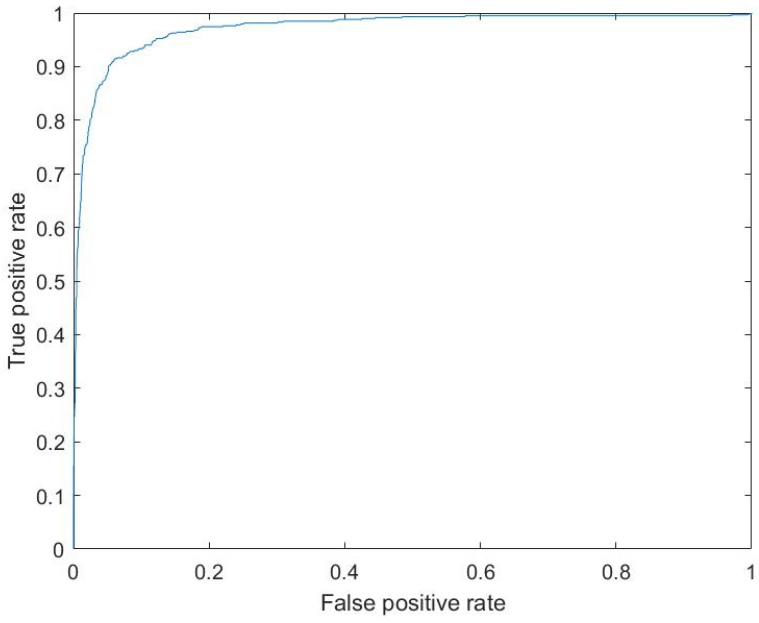


Figure 39. ROC curve of rem detection, AUC = 0.97

4.1.2 Hierarchical classification

A hierarchical structure classification method was implemented in this research. Sleep is a continuous process. The transition from one stage to another happens continuously. So, there are some similarities between different sleep stages.

Multiple layers of the hierarchical method were established. In each layer, only one sleep stage separated from other stages.

4.1.2.1 Experiment 5: Three sleep stage classification using the hierarchical structure

The hierarchical structure is composed of two layers. In the first layer, REM and NREM are combined as one class 'Sleep', and it becomes a binary classification problem. After separating wake from Sleep, all the epochs classified as Sleep were fed into the next layer. In the second layer, further classification was applied on Sleep. The results of each layer are as follows:

First layer: SVM classifier with Gaussian kernel and a total of 4202 epochs.

Accuracy: 88.9% Kappa: 0.69

Table 16. Confusion matrix in first layer wake detection

Predicted / Actual	Wake	Sleep
Wake	839	370
Sleep	139	2854

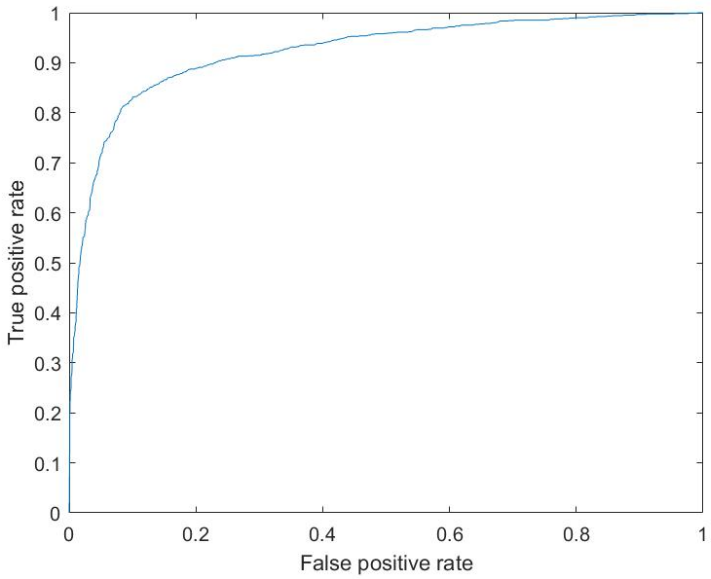


Figure 40. ROC of first layer, AUC = 0.92

Second layer: SVM classifier with Gaussian kernel and a total of 3224 epochs.

Accuracy: 84.6% Kappa: 0.64

Table 17. Confusion matrix in second layer wake detection

Predicted / Actual	Wake	REM	NREM
Wake	127	24	219
REM	19	462	81
NREM	110	44	2138

Table 18. True positive rates and false negative rates of three sleep stage classification using hierarchical method

Predicted / Actual	Wake	REM	NREM
Wake	34%	7%	59%
REM	3%	82%	14%
NREM	5%	2%	93%

All the epochs classified as ‘Sleep’ were fed into second layer from the first layer. However, the first layer confusion matrix indicates that 370 epochs of wake misclassified as ‘Sleep’, and all these epochs were participated in second layer classification. From the confusion above, we can see that in the second layer only 34% of wake epochs were separated from other two stages. 59% of wake still misclassified as NREM. If we accumulate the wake stages which were correctly detected from first and second layer, the total accuracy of wake detection is 80%, higher than accuracy of 75% in previous three experiments. The REM stage and NREM detection accuracy are 82% and 93%, both are slightly higher than previous

experiments.

4.2 Leave-one-subject-out strategy

The experiments using the put-all-subjects-together strategy indicate that the extracted features contain information for classifying the sleep stages. However, it is impossible for continuous monitoring when we do not have ground truth. Therefore, leave one subject out strategy was proposed. For each iteration, one of the subjects is regarded as testing data, while the remaining subjects are used for training.

4.2.1 Three class classifier

4.2.1.1 Experiment 6: Three sleep stage classification using the leave-one-subject-out strategy

Classifier: SVM classifier with Gaussian kernel

Average accuracy: 55.31% Average kappa: 0.21

Table 9 shows the confusion matrix as follow, the rows of confusion represent ground truth sleep stage, from top to bottom are wake, REM, and NREM. The columns of confusion matrix represent predict sleep stages, from left to right are wake, REM, and NREM.

Table 19. Confusion matrix, accuracy and kappa value for each subject using leave-one-subject-out strategy

Leave subject	Confusion matrix			Accuracy(%)	Kappa
subject0319	131	4	68	67.48	0.37
	93	8	32		
	85	24	496		
subject0320	60	11	64	37.98	-0.06
	46	0	104		
	259	76	283		
subject0418	139	8	144	59.74	0.24
	23	3	80		
	72	20	373		
subject0423	205	33	128	60.98	0.31
	1	0	56		
	57	29	270		
subject0425	125	9	80	50.35	0.17
	58	10	72		
	118	19	226		
Average				55.31	0.21

The accuracy and kappa value had significant degradation compared with the put-all-subjects-together strategy. The kappa indicated that the classification results have slightly agreement. The accuracy range from 38% to 67%. From the result of confusion matrix, subject 0320 has 42% of NREM misclassified as wake, which caused the overall accuracy to be 38%. Compare the confusion matrix of each subject, it is almost impossible to detect REM stage for all the subjects. However, there are some differences, subject 0319 misclassified REM as wake, but other subjects more likely misclassified REM as NREM. Then we selected the REM stages from each subject and calculated the average values of features. Figure 41 shows that the average values of

feature number 23 are different, the average values of other features are almost same.

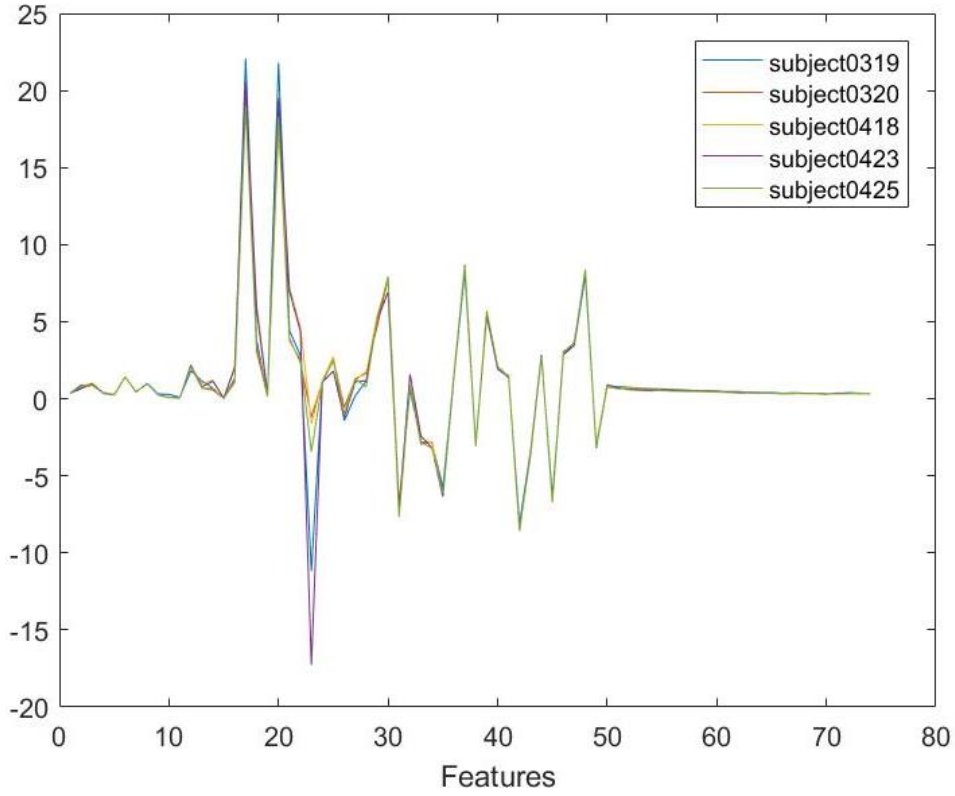


Figure 41. Average value of seventy-four features of REM stage of five subjects.

Figure 42 shows the value of feature number 23 of all REM epochs of subject 0319 and subject 0320. The epoch numbers of REM stage are 131 and 148 for two subjects respectively. The blue line and red line are in different length. The value of feature number 23 shows huge difference for two subjects in REM stage. This is probably one of the reasons that two subjects behave different in REM detection.

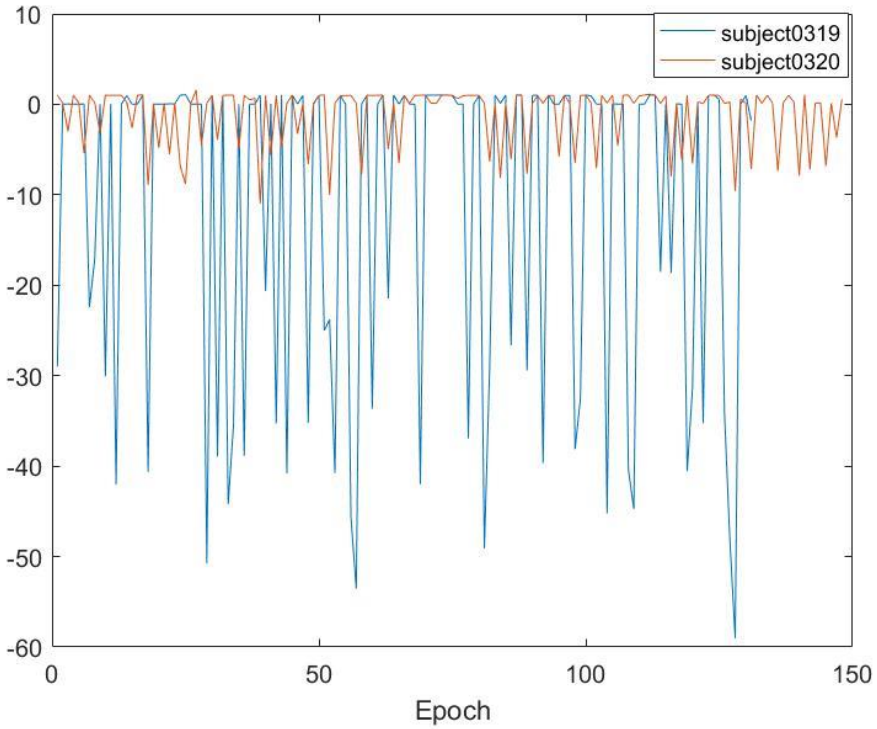


Figure 42. Average value of feature number 23 of REM stage of two subjects.

The feature number 23 is one of the RV features related to the difference between the amplitudes of expiration and inspiration. So we checked respiratory signal of PSG system and bed sensor. Figure 43 shows four epochs of chest band respiratory signal of REM stage. Even the ideal chest band signal is not smooth. The respiratory signal filtered form bed sensor signal shown in Figure 44. There are lots of huge spikes, only few regular respiration cycles can be seen. In this situation, the algorithm didn't work for respiration peaks and troughs detection. In addition, 30 times limb movements were recorded during 20min REM sleep in the annotation. So the features in REM stage are not reliable.

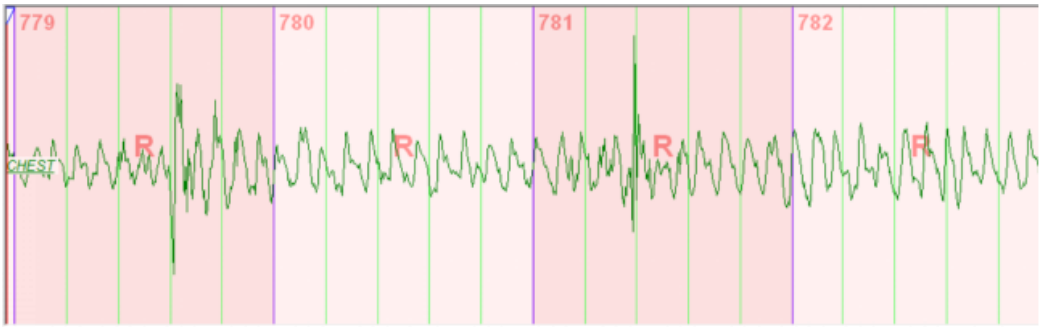


Figure 43. Four epochs of chest band signal during REM sleep of subject 0319

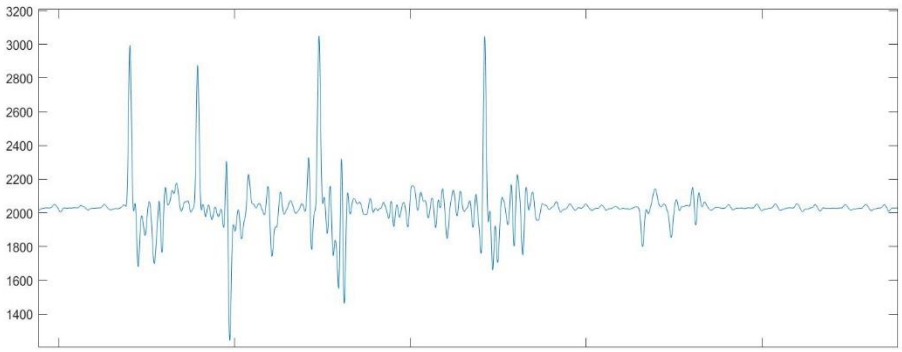


Figure 44. Respiration signal filtered from bed sensor signal in REM stage of subject 0319.

4.2.1.2 Experiment 7: Three sleep stage classification using the leave-one-subject-out strategy after normalization

The scales for different features are varying in large range. Experiment 5 indicates individual differences between subjects. In this experiment, z-score normalization was applied separately for each subject along each feature. All the features are centered to have zero mean and scaled to have standard deviation equal to one for each subject.

Classifier: SVM classifier with Gaussian kernel

Average accuracy: 58.43% Average kappa: 0.21

Table 10 shows the confusion matrix as follow the rows of confusion represent ground truth sleep stage, from top to bottom are wake, REM, and NREM. The columns of confusion matrix represent predict sleep stages, from left to right are wake, REM, and NREM.

Table 20. Confusion matrix, accuracy, and kappa value for each subject using leave-one-subject-out strategy after normalization

Leave subject	Confusion matrix			Accuracy(%)	Kappa
subject0319	101	2	100	68.76	0.33
	71	2	60		
	51	10	544		
subject0320	47	4	84	55.59	0.11
	92	1	57		
	157	7	454		
subject0418	150	6	135	59.86	0.24
	30	0	76		
	94	5	366		
subject0423	151	26	189	58.02	0.24
	20	0	37		
	52	3	301		
subject0425	89	7	118	49.93	0.13
	53	9	78		
	93	10	260		
Average				58.43	0.21

The average accuracy had 3% improvement compared with the result in Experiment 6. The normalization shows different impact on each subject. For subject 0320, the accuracy increased from 38% to 56%, because of NREM detection accuracy increased to 73% from 46%. For subject 0319 and subject 0418, the normalization didn't show much impact. However, for subject 0423 and subject 0425, the wake detection became worse after applying normalization. The kappa value indicated that the classification results have fair agreement.

4.2.2 Hierarchical classification

4.2.2.1 Experiment 8: Three sleep stage classification using leave-one-subject-out hierarchical method

Table 22 shows the confusion matrix and accuracy for each layer after the leave-

one-subject-out strategy was implemented. The rows of confusion matrix are ground truth sleep stages, the columns are predicted stages. For the first layer, two classes from top to bottom are wake and 'Sleep'. For the second layer, three classes from top to bottom are wake, REM, and NREM. The average accuracy for the first layer and second layer was 69.89% and 56.56%, respectively.

In the confusion matrix of first layer over half of the wake stages were misclassified as 'Sleep'. The result of Experiment 1, wake was misclassified as light sleep, may be able to explain this. Subject 0320 has lowest accuracy in first layer. As mentioned in Experiment 6, this subject has low accuracy in detecting NREM. And NREM accounted for a large proportion of sleep stage, which has a decisive influenced on the accuracy. Then all the misclassified wake fed into second layer. In the second layer, only small part of wake and REM were able to be separated from NREM.

Table 21. Confusion matrix and accuracy of two layers for each subject using leave-one-subject-out hierarchical method. Rows are ground truth, columns are predicted stages.

Leave subject	Layer1		Accuracy (%)	Layer2			Accuracy (%)	Kappa
subject0319	94	109	75.56	37	4	68	73.83	0.29
	121	617		37	8	31		
				26	24	491		
subject0320	63	72	62.79	4	11	57	45.31	-0.15
	264	504		17	0	104		
				63	63	257		
subject0418	129	162	72.97	12	6	144	58.16	0.01
	71	500		7	3	80		
				20	20	370		
subject0423	166	200	68.55	41	33	126	54.05	0.12
	45	368		0	0	56		
				17	29	266		
subject0425	99	115	69.60	29	8	78	51.46	0.07
	103	400		22	10	72		
				51	19	226		
Average			69.89				56.56	0.07

4.2.2.2 Experiment 9: Three sleep stage classification using leave-one-subject-out hierarchical method after normalization

Table 23 shows the confusion matrix and accuracy for each layer after the leave-one-subject-out strategy was implemented. In this experiment, z-score normalization was applied for each subject separately along each feature. The average accuracy for the first layer and second layer was 70.52% and 60.63%, respectively. The average accuracy increased by about 4% compared with Experiment 8. Especially the accuracy

for leaving the second subject out, the accuracy increased from 45% to 65%. Comparing the first layer confusion matrix with and without using normalization approach, more 'Sleep' have been detected, however, at the same time, more wake misclassified as 'Sleep'. So the accuracy influenced by the balance of correctly classification of 'Sleep' and wake misclassified as 'Sleep'. Refer to Table 9 the proportion of each sleep stage. The wake stage of subject 0320 accounted for 15% of entire night. Even though some wake misclassified as 'Sleep', the correct classification of 'Sleep' accounted for larger proportion. The final result shows an increase in accuracy from 63% to 73%. Contrary to subject 0320, subject 0423 has 47% of wake for entire night. If wake misclassified as 'Sleep', it has greater impact on final result, which shows a decrease of accuracy from 69% to 63%.

Table 22. Confusion matrix, accuracy, and kappa value for each subject using leave-one-subject-out hierarchical method after normalization. Rows are ground truth, columns are predicted stages.

Leave subject	Layer1		Accuracy (%)	Layer2			Accuracy (%)	Kappa
subject0319	85	118	79.06	28	5	62	72.72	0.16
	79	659		37	3	26		
				29	16	466		
subject0320	44	91	73.20	4	4	83	64.55	0.08
	151	617		58	1	57		
				42	7	452		
subject0418	137	154	72.85	14	6	134	58.45	0.01
	80	491		13	0	76		
				35	4	363		
subject0423	117	249	63.03	35	26	188	53.93	0.10
	39	374		12	0	35		
				23	3	301		
subject0425	54	160	64.44	37	6	117	53.52	0.10
	95	408		17	9	76		
				38	10	258		
Average			70.52				60.63	0.09

Comparing the result of two strategies, the put-all-subject-together shows good performance. All the three experiments of three sleep stage classification have better results than previous Jialei's work. The large percentage of wake misclassified as light sleep, so more work needed to explore the details behind these two stage. The REM detection has comparable result as Jialei's best result. Typically, healthy adults have about 20% REM of entire night sleep. However, from Table 9 we can see that the

average proportion of REM is 12% for the subjects participated in this study. The percentage of REM range from 7% to 20%. In addition, that REM sleep happened in two or more different time periods, which means each continuous REM period is pretty short. And sleep is a continuous process and the sleep stage labelling based on technician's subjective decision. Therefore, the boundary of one sleep stage change to another stage has some uncertainties, the epochs at the beginning and end of one sleep stage may have similar characteristics with neighbor stage. The certain REM stage may even less than it was reported, which makes REM detection more difficult.

The leave-one-subject-out strategy shows 58% accuracy in three stage classification. However, the correctly detection of NREM has great contribution to the final accuracy. It is almost impossible to detect REM in this strategy. As mentioned in Experiment 6, the characteristic of REM is different between subjects may be one of the reasons. In the hierarchical method, the misclassification of wake as 'Sleep' in the first layer fed into the second layer, which has cumulative impact on second layer classification.

CHAPTER 5. CONCLUSIONS

The initial goal of this research was to use noninvasive hydraulic bed sensors to classify the sleep stages. The BCG signals collected using hydraulic bed sensors were used in this study. The ground truth sleep stages provided by the BHC sleep lab is considered to be the most accurate available to date. The subjects who participated in this study were elderly people with different severity levels of apnea. The individual differences are more obvious than the data collected from one subject during several entire nights. The heart beat intervals and breath intervals were calculated from filtered heart rate signal and respiratory signal. Then, HRV features and RV features were extracted based on these intervals respectively. The LFCC features were generated from filtered heart rate signals.

Two strategies were implemented in this thesis. The results of the put-all-subjects-together strategy showed potential in classifying the sleep stages. It indicated that the features extracted from the BCG signals contain characteristics related to sleep stages. Both SVM and k-NN showed acceptable performances with about 85% accuracy and a kappa value of 0.7, which are higher than most of the results reported in the literature. The fifteen selected features showed similar results when compared to all the features fed into the SVM classifier. Reduced dimension features can be considered in future study. Different methods of feature selection need to be implemented and compared. The hierarchical method had an overall accuracy about 85%. The hypothesis that wake is separable from other stages in the first layer had an

88% accuracy rating.

The results of leave-one-subject-out strategy were not meet the expected goal, given the good performance in put-all-subjects-together strategy. Except for the uncertainty of the BCG signal and features caused by sensitivity of bed sensor, the individual differences were an issue. In all the experiments using this strategy, it is difficult to detect REM. Except the reason discussed before that the proportion of REM is too small compared with other sleep stages, find the characteristics of the original signal in each subject may explain these differences. Z-score normalization was applied to reduce the scales for different features. The accuracy for leaving the second subject out increased significantly from 45% to 65%. The accuracy range of testing one subject out reduced to 54%-74% from 45%-74%. The normalization approach reduced individual differences in some extent.

CHAPTER 6. FUTURE WORK

In this study, the PSG system provided the ground truth, which is considered the most accurate and reliable method to date. Hence, we think no improvement is needed in ground truth. However, there is much room for improvement in the bed sensor technology developed in this study. The following reasons are offered to explain the poor results:

1) The bed sensor signal is particularly sensitive to movement. This caused a lot of noise overlapping with the original signal. The BCG signals were not accurate and clean enough to detect heart beats compared with ECG signal. The quality of BCG signal further affected the quality of HRV features extracted from BCG signal. Another problem is that the subjects in this study were elderly; many had a history of sleep disorders and medical complications. A motion detection method or noise removal method needs to be applied before the heart beat detection. For those subjects with apnea, an apnea detection approach should be done before respiratory peak and trough detection.
2) To achieve maximum access to the subjects' heart rate and respiratory information, the four transducers installed under the mattress were placed at equal intervals. However, this led to a transducer selection problem. When processing the signal for each epoch, only one transducer was selected. The selection criteria in this study is based on maximum DC value. We assumed

that large DC value represents more weight on the transducer and a better connection with subjects. But the fact is, the transducer with large DC did not always have the best signal. So, a more accurate transducer selection approach needs to be developed.

3) The features extracted in this study are limited. In the literature, other features were reported to have good performance. Instead of using same features for all detection problems, using different combinations of features in different sleep stages has proven effective in the literature. This is more reasonable because there are differences between the sleep stages.
4) Need to find the criteria for subject selection. Only five out of seventy-seven subjects were selected for the experiments. Different standards will select different combinations of subjects. More experiments are needed to find relatively ideal subjects.

REFERENCES

1. Senaratna, C.V., et al., *Prevalence of obstructive sleep apnea in the general population: a systematic review.* Sleep Medicine Reviews, 2017. 34: p. 70-81.
2. Musiek, E.S., D.D. Xiong, and D.M. Holtzman, *Sleep, circadian rhythms, and the pathogenesis of Alzheimer disease.* Experimental & molecular medicine, 2015. 47(3): p. e148.
3. Peter-Derex, L., et al., *Sleep and Alzheimer's disease.* Sleep Medicine Reviews, 2015. 19: p. 29-38.
4. Videnovic, A. and D. Golombek, *Circadian and sleep disorders in Parkinson's disease.* Experimental neurology, 2013. 243: p. 45-56.
5. Berry, R.B., et al., *AASM scoring manual updates for 2017 (version 2.4).* Journal of Clinical Sleep Medicine, 2017. 13(05): p. 665-666.
6. Kortelainen, J.M., et al., *Sleep staging based on signals acquired through bed sensor.* IEEE Transactions on Information Technology in Biomedicine, 2010. 14(3): p. 776-785.
7. Park, K.S., et al., *Ballistocardiography for nonintrusive sleep structure estimation*, in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE.* p. 5184-5187.
8. Heise, D. and M. Skubic. *Monitoring pulse and respiration with a non-invasive hydraulic bed sensor*. in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE.*
9. Heise, D., et al., *Refinement and evaluation of a hydraulic bed sensor.* Conf Proc IEEE Eng Med Biol Soc, 2011. 2011: p. 4356-60.
10. Lydon, K., et al. *Robust heartbeat detection from in-home ballistocardiogram signals of older adults using a bed sensor*. in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE.*

11. Rosales, L., et al. *Heartbeat detection from a hydraulic bed sensor using a clustering approach*. in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.*

12. Su, B.Y., et al. *Pulse rate estimation using hydraulic bed sensor*. in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.*

13. Devot, S., R. Dratwa, and E. Naujokat. *Sleep/wake detection based on cardiorespiratory signals and actigraphy*. in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE.*

14. Redmond, S.J., et al., *Sleep staging using cardiorespiratory signals.* Somnologie-Schlafforschung und Schlafmedizin, 2007. 11(4): p. 245-256.

15. Willemen, T., et al., *An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification.* IEEE journal of biomedical and health informatics, 2014. 18(2): p. 661-669.

16. Rechtschaffen, A., *A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects.* Brain information service, 1968.

17. Iber, C., et al., *Westchester, IL: American Academy of Sleep Medicine; 2007.* The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, 2007. 4849.

18. American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events*. 2018; Available from: https://aasm.org/clinical-resources/scoring-manual/.

19. Carskadon, M.A. and W.C. Dement, *Normal human sleep: an overview.* Principles and practice of sleep medicine, 2005. 4: p. 13-23.

20. Altevogt, B.M. and H.R. Colten, *Sleep disorders and sleep deprivation: an unmet public health problem*. 2006: National Academies Press.

21. Giovangrandi, L., et al. *Ballistocardiography—a method worth revisiting*. in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE.*

22. Starr, I., et al., *Studies on the estimation of cardiac ouptut in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts; the ballistocardiogram.* American Journal of Physiology-Legacy Content, 1939. 127(1): p. 1-28.

23. Bilchick, K.C. and R.D. Berger, *Heart rate variability.* Journal of cardiovascular electrophysiology, 2006. 17(6): p. 691-694.

24. Baharav, A., et al., *Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability.* Neurology, 1995. 45(6): p. 1183-1187.

25. Scholz, U.J., et al., *Vegetative background of sleep: spectral analysis of the heart rate variability.* Physiology & behavior, 1997. 62(5): p. 1037-1043.

26. Malik, V., D. Smith, and T. Lee-Chiong, *Respiratory Physiology During Sleep.* Sleep Medicine Clinics, 2012. 7(3): p. 497-505.

27. Gutierrez, G., et al., *Respiratory rate variability in sleeping adults without obstructive sleep apnea.* Physiol Rep, 2016. 4(17).

28. Quan, S.F., et al., *Short-term variability of respiration and sleep during unattended nonlaboratory polysomnography—the Sleep Heart Health Study.* Sleep, 2002. 25(8): p. 8-14.

29. Viera, A.J. and J.M. Garrett, *Understanding interobserver agreement: the kappa statistic.* Fam Med, 2005. 37(5): p. 360-363.

30. Cohen, J., *Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.* Psychological bulletin, 1968. 70(4): p. 213.

31. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* biometrics, 1977: p. 159-174.

32. Ohayon, M., et al., *National Sleep Foundation's sleep quality recommendations: first report.* Sleep Health, 2017. 3(1): p. 6-19.

33. Yang, J., *Recognition of sleep stages from sensor data*. 2015: University of Missouri-Columbia.

34. Redmond, S.J. and C. Heneghan, *Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea.* IEEE Transactions on Biomedical Engineering, 2006. 53(3): p. 485-496.

35. Huang, C.-S., et al. *A hierarchical classification system for sleep stage scoring via forehead EEG signals*. *2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*.

36. Heise, D. and M. Skubic. *Monitoring pulse and respiration with a non-invasive hydraulic bed sensor*. *in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE.*

37. (AAST)，, A.A.o.S.T. *Sleep (Polysomnographic) Technologist Job Description*. 2018; Available from: https://www.aastweb.org/job-technologist.

38. Natus Medical Inc. *Natus NeuroWorks®/SleepWorks™* 2018; Available from: https://citrixready.citrix.com/natus-medical-incorporated/neuroworks-sleepworks.html.

39. B, D. *3 Common Artifacts That a Sleep Technologist Must Manage*. March 3, 2016; Available from: https://www.aastweb.org/blog/3-common-artifacts-that-a-sleep-technologist-must-manage.

40. Huang, X., et al., *Spoken language processing: A guide to theory, algorithm, and system development*. Vol. 1. 2001: Prentice hall PTR Upper Saddle River.

41. Davis, S.B. and P. Mermelstein, *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*, in *Readings in speech recognition*. 1990, Elsevier. p. 65-74.

42. Lyons, J. *Mel Frequency Cepstral Coefficient (MFCC) tutorial*. 2013; Available from: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/#related-pages-on-this-site.

43. Shawe-Taylor, J. and N. Cristianini, *Kernel methods for pattern analysis*. 2004: Cambridge university press.

44. Everitt, B.S., et al., *Miscellaneous clustering methods.* Cluster Analysis, 2011: p. 215-255.