

AUTOMATED SLEEP STAGE CLASSIFICATION

A Thesis

Presented to

The Faculty of the Department of Electrical and Computer
Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

In Electrical Engineering

by

Nishant Prakash

May 2014

Acknowledgements

I would like to express my deepest gratitude to my advisor Dr. Ben H. Jansen for all the guidance and encouragement he provided during the course of this research. He has been extremely patient with me and remained always there to clear my doubts. I would also like to thank Dr. Bhavin Sheth and Dr. Ahmet Omurtag for serving in my thesis committee and providing valuable suggestions. I would also like to thank my lab-mates Ruoli Ziyang and Hamza for his help and Ziyang Li for providing me the sleep data for analysis. I thank my parents and my brother for their support and encouragement. Finally, I thank the almighty god for all what I have achieved.

AUTOMATED SLEEP STAGE CLASSIFICATION

An Abstract

of a

Thesis

Presented to

The Faculty of the Department of Electrical and Computer
Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

In Electrical Engineering

by

Nishant Prakash

May 2014

Abstract

Objectives: We have developed a system to perform automated sleep staging. The system is a re implementation of the sleep staging method originally developed by Jansen in late 70's. We extended this system by incorporating transient detection techniques.

Method: The basic method divides the EEG into short intervals (about 1 s) from which autoregressive (AR) model coefficient are extracted. The AR coefficients serve as features in a clustering process to establish a library of elementary patterns. These are used to classify the 1 s EEG segments. Next, histograms showing how frequently each elementary pattern occurs in 30 s interval are obtained, and used for sleep staging. Special transient detection rules are then applied to detect spindles and rapid eye movements.

Results: The method was developed using nine nights of sleep from three subjects and tested on sleep from seven other subjects. An overall classification rate of around 80 % was obtained. We observed that the method for spindle detection did not work as expected, but the REM detection method worked satisfactorily and it increased the classification percentage of stage 1 and REM.

Conclusion: Acceptable sleep staging results were obtained, but improvement may be possible using better spindle and K-complex detection.

Table of Contents

Acknowledgements.....	iv
Abstract.....	vi
Table of Contents.....	vii
List of Figures.....	ix
List of Tables.....	xiii
Chapter 1 Background.....	1
1.1 Introduction.....	1
1.2 Objectives.....	3
Chapter 2 Literature Review.....	5
2.1 Early Work Done in Automatic Sleep Staging.....	5
2.2 Automated Detection of Sleep Transients	7
2.3 Methods Developed for Automated Sleep Staging Incorporating the Non-stationarity of the Signal.....	12
2.4 Method for Automatic Sleep Staging Based on Artificial Neural Networks and Wavelet Transform.....	15
Chapter3 Methods.....	17
3.1 Basic Outline of the Method.....	17
3.2 Data Preprocessing.....	18
3.3 Feature Extraction.....	19

3.4	Clustering.....	22
3.5	Likelihood Test.....	27
3.6	Classification.....	42
3.7	Detection of Sleep Transients.....	46
3.7.1	Spindle Detection.....	46
3.7.2	REM Detection.....	48
3.8	Data and Subjects.....	50
Chapter 4	Results.....	54
4.1	Training Data.....	54
4.2	First Testing Dataset.....	62
4.3	Final Testing Dataset.....	67
Chapter 5	Discussion and Conclusions.....	71
	References.....	74

List of Figures

Figure 1.1	Different types of wave activity present in sleep stages.....	2
Figure 1.2	Sleep spindles and K complexes present in sleep stage 2.....	2
Figure 3.1	Flowchart for the sleep stage Classification.....	18
Figure 3.2	Comparison of power spectrum for awake stage using FFT and AR model.....	20
Figure 3.3	Comparison of power spectrum for sleep stage 1 using FFT and AR model.....	20
Figure 3.4	Comparison of power spectrum for sleep stage 2 using FFT and AR model.....	21
Figure 3.5	Comparison of power spectrum for sleep stage 2 using FFT and AR model.....	21
Figure 3.6	Comparison of power spectrum for REM stage using FFT and AR model.....	22
Figure 3.7	Dendrogram of Awake stage data from Subject 1.....	24
Figure 3.8	Dendrogram of Stage 1 data from Subject 1.....	25
Figure 3.9	Dendrogram of Stage 2 data from Subject 1.....	25
Figure 3.10	Dendrogram of Stage 3 data from Subject 1.....	26
Figure 3.11	Clusters obtained from Subject 1 data.....	26
Figure 3.12	One second time series for the four clusters selected.....	27
Figure 3.13	Histogram of likelihood for Subject 1 data belonging to the patterns selected from Subject 1.....	28
Figure 3.14	Histogram of likelihood for Subject 1 data belonging to the patterns selected from Subject 1 and Subject 2.....	29

Figure 3.15	Histogram of likelihood for Subject 2 data belonging to the patterns selected from Subject 1 and Subject 2.....	30
Figure 3.16	Histogram of likelihood for Subject 1 data belonging to the cluster 1, 2, 3 and 4 from the seven selected patterns....	31
Figure 3.17	Histogram of likelihood for Subject 1 data belonging to the cluster 5, 6 and 7 from the seven selected patterns.....	32
Figure 3.18	Histogram of likelihood for Subject 2 data belonging to the cluster 1, 2, 3 and 4 from the seven selected patterns....	32
Figure 3.19	Histogram of likelihood for Subject 2 data belonging to the cluster 5, 6 and 7 from the seven selected pattern.....	33
Figure 3.20	Histogram of likelihood for Subject 3 data belonging to the cluster 1, 2, 3 and 4 from the seven selected patterns....	33
Figure 3.21	Histogram of likelihood for Subject 3 data belonging to the cluster 5, 6 and 7 from the seven selected patterns.....	34
Figure 3.22	Final clusters selected after clustering.....	35
Figure 3.23	Histogram of likelihood for Subject 4 data belonging to the cluster 1, 2, 3 and 4 from the seven selected patterns....	35
Figure 3.24	Histogram of likelihood for Subject 3 data belonging to the cluster 5 and 6 from the seven selected patterns.....	36
Figure 3.25	One second segment using the centroid of cluster 1.....	37
Figure 3.26	One second segment using the centroid of cluster 2.....	37
Figure 3.27	One second segment using the centroid of cluster 3.....	38
Figure 3.28	One second segment using the centroid of cluster 4.....	38
Figure 3.29	One second segment using the centroid of cluster 5.....	39
Figure 3.30	One second segment using the centroid of cluster 6.....	39

Figure 3.31	Power spectrum of the one second segment closest to the centroid of alpha clusters.....	40
Figure 3.32	Power spectrum of the one second segment closest to the centroid of theta clusters.....	40
Figure 3.33	Power spectrum of the one second segment closest to the centroid of spindle clusters.....	41
Figure 3.34	Power spectrum of the one second segment closest to the centroid of delta clusters.....	41
Figure 3.35	Power spectrum of the one second segment closest to the centroid of rem clusters.....	42
Figure 3.36	Reference profile for Awake stage.....	43
Figure 3.37	Reference profile for Stage 1.....	44
Figure 3.38	Reference profile for Stage 2.....	44
Figure 3.39	Reference profile for Stage 3.....	45
Figure 3.40	Reference profile for REM.....	45
Figure 3.41	Power spectrum of the spindle cluster.....	46
Figure 3.42	Histogram of the spindle distribution in each sleep stage.....	47
Figure 3.43	Histogram of the spindle distribution in REM sleep stage.....	48
Figure 3.44	Histogram for the number of times the standard deviation exceeds 30, 40, 50 and 60 in REM stage.....	49
Figure 3.45	Histogram for the number of times the standard deviation exceeds 30, 40, 50 and 60 in stage 1.....	50
Figure 3.46	International 10-20 principle of electrode placement.....	51

Figure 4.1	Graph of the location of the one second segments from awake stage classified as alpha (stars) and spindle activity (circles) in the Euclidean space spanned by the first two AR coefficients and the variance58
Figure 4.2	Sleep hypnogram of Night 2 for subject 3.....59
Figure 4.3	Sleep hypnogram of Night 1 for subject 2.....59
Figure 4.4	Histogram showing the epoch difference where stage transition occurs in automated scoring as compared to visual scoring (1 epoch = 30 second).....60

List of Tables

Table 3.1	Sleep wave activity and their frequency range.....	23
Table 3.2	Overall correct classification and spindle threshold.....	48
Table 3.3	Sleep data available for each subject.....	52
Table 4.1	Confusion matrix using only profile classification.....	54
Table 4.2	Confusion matrix using profile classification and REM detection rule.....	55
Table 4.3	Confusion matrix using profile classification and spindle detection rule.....	55
Table 4.4	Confusion matrix using profile classification, spindle detection rule and REM detection rule.....	56
Table 4.5	Percentage correct classification for training subjects per night per sleep stage.....	56
Table 4.6	Sleep onset time comparison for the training subjects.....	61
Table 4.7	Sleep Quality comparison for training subjects.....	61
Table 4.8	Confusion matrix for initial batch of testing data using profile classification, spindle detection rule and REM detection rule.....	63
Table 4.9	Percentage correct classification for the first testing subjects per night per sleep stage.....	63
Table 4.10	Sleep onset time comparison for the first testing subjects.....	65
Table 4.11	Sleep Quality comparison for the first testing subjects.....	66
Table 4.12	Confusion matrix for final batch of testing data using profile classification, spindle detection rule and REM detection rule.....	67

Table 4.13	Percentage correct classification for the final batch of testing subjects per night per sleep stage.....	68
Table 4.14	Sleep onset time comparison for the final batch of testing subjects.....	69
Table 4.15	Sleep Quality comparison for the final batch of testing subjects.....	69

Chapter 1 Background

1.1 Introduction

The human brain produces electrical activity that can be recorded from electrodes attached to the scalp. This recording is referred to as the electroencephalogram (EEG). The amplitude and frequency characteristics of the EEG change as a function of the waking/sleeping cycle. Retschaffen and Kales [1] classified the human sleep in five recurring stages: Rapid Eye Movement (REM) and Non rapid Eye Movement (NREM). REM is characterized by low voltage, mixed frequency EEG and rapid eye-movements. NREM is further divided into stages S1, S2, S3, and S4 (stages S3 and S4 were grouped into a single stage in 2007) [2].

During wakefulness (eyes closed) the EEG is characterized by alpha activity (sinusoidal rhythm in the 8 to 12 Hz range) that appears in more than 50% of the epoch. Stage 1 is characterized by a low voltage mixed frequency EEG without rapid eye movements. The waves present in this stage are theta waves (oscillatory patterns in 4-7 Hz range). In Stage 2 there is presence of sleep spindles (oscillations of 12–15 Hz and duration of 0–1.5 s) and K-complexes (High-amplitude biphasic waves of at least 0.5 second duration, with an initial sharp positive Voltage followed by a negative deflection slow wave). Besides the sleep spindles and K complexes, stage 2 also has beta activity (in 15-30 Hz range). Stages 3 and 4 are defined by delta activity with a frequency of less than 4 Hz and peak-to-peak amplitude greater than 75 μ V. This is the deepest stage of sleep and is also called as slow wave sleep (SWS). Figure 1.1 and 1.2 below show different types of wave activity present in sleep stages.

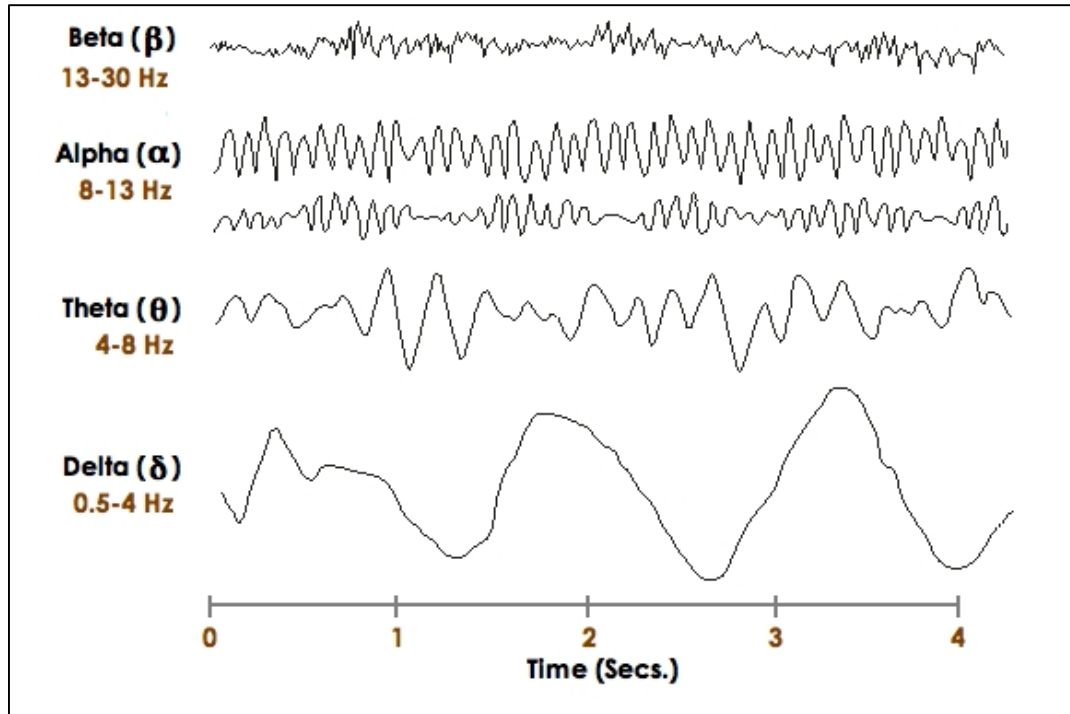


Figure 1.1 Different types of wave activity present in sleep stages [3].

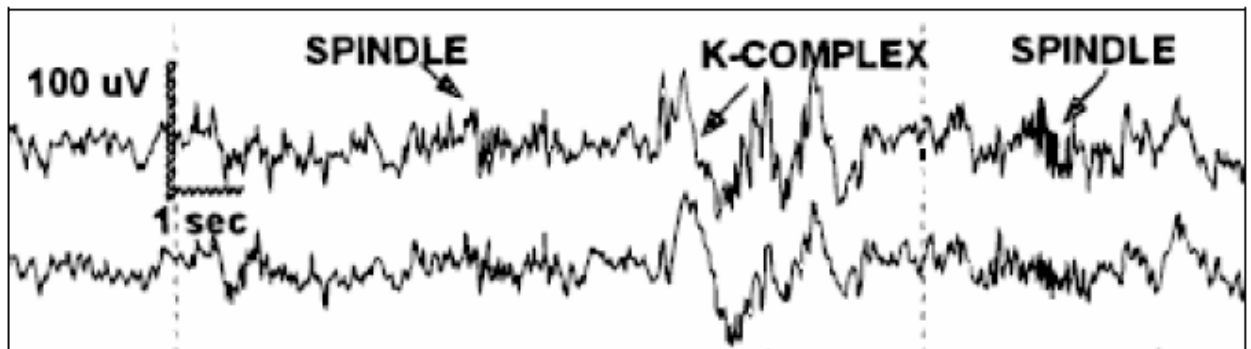


Figure 1.2 Sleep spindles and K complexes present in sleep stage 2 [4].

Classifying the sleep data into one of the above mentioned stages and then plotting the whole night's trend as a sleep hypnogram is called sleep scoring. Sleep is scored mostly on the basis of 20 or 30 seconds of epochs, using three signals: the electroencephalogram (EEG), electrooculograms (EOG) and electromyogram (EMG). EOG is the recording of eye movements and EMG is the recording of chin movements. EMG channel is used for detection of the REM as EMG signal is flat during the REM

sleep. The analysis of EEG signals is very challenging because of its non-stationary (signal whose frequency content and statistical parameters changes with time) nature and introduction of artifacts during recording such as sweating, physical movements by the subject, settling time needed for electrodes etc. The task of manually scoring the sleep stages is very onerous and cumbersome and even if it is performed by trained personnel, the reliability and coherence of manual sleep scoring are unsatisfactory.

Thus, there have been efforts earlier in the direction of automatically scoring the sleep stages and one such method was developed by Jansen [4] in the late 1970's/early 1980's, which takes into account the non-stationary nature of the EEG. This method underwent limited testing, thus it is anticipated that the availability of contemporary computer technology and advanced signal processing methods could lead to an improvement of performance.

1.2 Objectives

This thesis aims to achieve the following objectives:

1. Re-implement the Jansen method for sleep staging using contemporary computer technology. We plan to use Matlab and EEGLAB for EEG processing, and PR-Tools for clustering and classification.
2. Evaluate the method on sleep data. Sleep data will be obtained. These data will be scored visually by an expert Somnologist and the computer-generated sleep stages will be compared with the visual scoring.
3. Propose and implement improvements of the Jansen method if needed. We will use AR modeling to characterize the EEG segments and also add the detection of sleep spindles.

The thesis is organized into five chapters. First chapter is the introduction. The second chapter describes the relevant research work already done in this field which started in late 1970's. Chapter three focusses on the detailed methodology which we followed to design the system. Results obtained are discussed in chapter four and the fifth chapter consists of conclusion, discussion and scope of further improvement.

Chapter 2 Literature Review

Automated sleep staging has received a lot of attention over the years and a substantial amount of research work has been done in this field. This chapter has been divided in four sections. The first section describes some of the very early work done to perform automated sleep staging. The second section specifically focuses on methods for the detection of sleep transients. The third section describes sleep staging methods addressing the issue of the non-stationarity of the EEG signals. The fourth section reviews other techniques such as artificial neural networks and wavelet transforms used for automated sleep staging.

2.1 Early Work Done in Automatic Sleep Staging

Various methods were developed by researchers [6, 7, and 8] in late 1960's. One of the first studies reporting on automated sleep staging and comparing its results to human scoring was done by Smith et al. [9]. A special-purpose Sleep Analyzing Hybrid Computer (SAHC) was constructed to process four EEG channels simultaneously (frontal, parietal, central and occipital) as it was found that having more than one EEG channel improved the system's accuracy. The system used four waveforms for the classification of sleep stages: alpha activity, delta activity, spindles and K complexes.

The frontal channel was used for detection of delta activity, the parietal and central channels for the detection of spindles and K complexes, and the occipital channel for alpha activity detection. Each sleep stage was determined based on an epoch duration of 60 seconds. They used the power content in the frequency bands for determination of different types of waveform activity including alpha, spindle and delta activity. A rule-

based scoring was used for classification of sleep stages. An epoch was classified as stage 0 (awake) if alpha activity occurred for more than 30 second in an epoch. If alpha activity occurred for less than 30 seconds or delta activity for less than 12 seconds and a maximum of one K-complex or spindle per epoch then it was classified as stage 1. Stage 2 was classified similar to stage 1 except K complexes or spindle activity occurred more than once per epoch. They classified stage 3 when delta activity was present for more than 12 seconds but less than 30 seconds and an epoch having delta activity for more than 30 seconds was classified as stage 4. Here in the rule, they did mention about the K-complexes but the K-complex detector was not used during analysis. In this work only four recordings of 2 hour each were used for comparison against manual scoring which may not be sufficient to fully assess the method. The accuracy obtained for each of the records were 89%, 82%, 72% and 87%, respectively, giving an average accuracy of 82.5%.

Later, using the pattern-recognition techniques, Martin et al. [10] again showed around 83% agreement with visual scoring on 30-second epochs of five healthy subjects. The EEG data was sampled at frequency of 68.3 Hz and it was filtered to remove all the frequencies above 28 Hz. The EOG data signal was also filtered to remove the frequencies above 14 Hz. For each night recording, Fourier analysis was performed on successive 30 seconds EEG to compute the spectral components for each 1 Hz increase till 34 Hz. These components were combined to form one indicator and the 34 spectral intensity values were used by the pattern recognition system for classification of sleep stages. There was a separate delta measurement program where a 30 second epoch was examined to locate peaks more than 0.5 sec apart corresponding to activity of 0-2 Hz and

delta activity was confirmed if the amplitude exceeded $0.75 \mu\text{V}$. Stage 3 and 4 were assigned based on percentage of delta activity in the 30 second epoch. If there was no classification of stage 3 or 4 then the recognition system used the 34 spectral indicators to assign the epoch as stage 2, awake and “stage 1 or REM”. If the epoch was assigned as “stage 1 or rem” then the corresponding EOG was looked at and if eye movements were found then the epoch was assigned as REM stage.

2.2 Automated Detection of Sleep Transients

In continuation of the research done by Smith et al. [9], early work on K complex detection was done by Bremer et al. [11] with an aim to incorporate it into the SAHC. They developed a mathematical description of a K complex which was to be used as a basis for detection. Hybrid pattern recognition was used which mimicked the way human scoring is performed and an electronic circuit was designed that closely matched the requirements of a mathematical model of a K complex. The parameters of the system were adjusted empirically based on the K complexes detected by human scorers. The detection of K-complex by the system had an agreement rate of 68% with the human scorers. Another issue they faced while comparing their sleep scoring with the manual scoring was the inter-rater disagreement. Out of total 102 K-complexes scored by scorer 1 and 80 K-complexes scored by scorer 2, there was an agreement on only 51 of them and the detection system agreed with 32 out of these 51.

Further work in the detection of K complexes using multi-layer back propagation artificial neural network was done by Jansen [12]. One hour data from all-night recordings of 6 subjects was used for analysis. All data was passed through low-pass

filter having a threshold frequency of 30 Hz. Seven EEG channels were stored (Fpl, F3, F1, C3, C4, T3, and T4), but data from Fpl only was used in the study. Ten-second intervals, either containing K-complexes, delta activity, or isolated waves bearing some resemblance to the K complexes, were selected and classified by an expert Somnologist. These segments were sampled at 100 Hz and used for further processing. Two different methods were used to prepare the inputs for the ANN. In the first case, the data was passed through 1st order Butterworth band pass filter (0.5 - 2.3Hz range) and the output was scaled to obtain a series of values in the range of 0 to 1. In the second case the data was down sampled by a factor of 2 and again the output was scaled differently to obtain a series of values between 0 and 1. These were directly used as an input to ANN with 500 input nodes. Four data-sets (one for training and three for testing) were generated. About half of the data of four subjects were used for training, with the remainder to be used as testing set. Two testing sets were composed from data subject 1 and subject 2 respectively. During training, small amounts of noise were added to the original training input patterns to create a large number of slightly different patterns. The results obtained were not very convincing with the detection rate ranging from 42% to 67%, suggesting that the neural net approach does not seem to be adequate for the detection of K-complexes.

Later, an automatic method for K-complexes detection based on feature extraction and the use of likelihood thresholds was developed by Devuyst et al. [13]. The features extracted included minimum peak to peak amplitude, average amplitude of the background EEG for 15 seconds surrounding the K-complex and the frequency content of 5 second segment surrounding the K-complex. The training data set consists of 2 whole

night recordings and 1 excerpt of 2 hours coming from 3 healthy subjects. The testing data set consists of 5 excerpts of 30 minutes extracted from whole-night recordings of 5 other healthy subjects sampled at 200Hz. These records were given independently to two experts for K-complexes scoring. The strategy was to limit the number of false detections using a multi-level approach. At each stage the characteristic of a testing K- complex was compared to a threshold to confirm its legitimacy. Each testing K- complex was assigned a likelihood corresponding to a real K-complex. The threshold was computed using a feature value and thresholding curve which was established from the training data set. The threshold value of 0.69 was chosen, which correspond to the optimal point of the ROC curve obtained by varying the minimal likelihood from 0.5 to 1 on the testing data set. After computing the final likelihood for each testing K-complex and comparison with the threshold value the detection was confirmed. The correct detection rate of K complex obtained was 61.72% and 60.94%, respectively with scorer 1 and scorer 2. However, for a total of 209 K-complexes scored by scorer 1 and 64 K-complexes scored by scorer 2, there was a mutual agreement on only 43. This corresponded to sensitivities of only 20.57% and 67.18% respectively. The detection system agreed with 33 of these K-complexes, which corresponded to an agreement rate of 76.74% (when a K-complex was considered as real when both scorers marked it as such).

Later, Duman et al. [14] presented a sleep spindle detection method using the combination of three approaches: STFT (Short Time Fourier Transform), MUSIC (Multiple signal classification algorithm) for frequency analysis and TEO (Teager operator) algorithm for amplitude analysis. The EEG data was sampled at 128 Hz and band pass filter was applied to select frequency between 0.5-35 Hz. The processed signal

was decomposed into its sub bands by using wavelet transform and a hamming window with 2 sec interval. A spindle was finally confirmed in 30 second epoch when all the three algorithms detected the spindle in a 30 second epoch.

TEO is an energy operator that depends on the first two derivative of an input signal and makes the sudden amplitude changes and discontinuities become sharper. In a 30 second epoch, a spindle was detected by the TEO only if its amplitude was more than 60% of average amplitude and it lasted for at least 0.5 second. MUSIC is a robust technique to extract the signal by projecting its frequency components to subspace. In this method, a correlation matrix is formed from the signal and its eigenvalues are determined. The eigenvalues are divided into two groups; eigenvalues greater than some threshold, and Eigen values less than that threshold. The assumption was that Eigen values corresponding to the signal sub-space have values greater than the noise threshold level, with the remaining eigenvalues corresponding to the noise sub-space. The spectrum was obtained by projecting the signal eigenvalues to noise sub-space and it gave sharp peaks at the signal frequencies which are related to the spindles.

In STFT, a sliding hamming window of 64 taps was selected and frequency band between 0-32 Hz was obtained. A sleep spindle was detected when the normalized value of STFT was one and frequency around 12 Hz. This algorithm was tested using 16 subjects and the overall sensitivity and specificity was found to be 96.17% and 95.54%. One major advantage of this approach was the rejection of false waveforms which reduced the false alarm.

Early work on automated REM detection was also done by Smith et al. [15]. A Multi Hybrid System for REM detection was developed and was it used an electronic

system consisting of both analog and digital circuitry. The output of the system used two channels of serial digital pulses, each pulse corresponding to single REM detection. Both the EOG channels were used to eliminate any confusion with other signals with similar characteristics. The minimum threshold for the detection of eye movement by the analog circuit was set such that the amplitude of a REM be must be greater than or equal to $50\mu\text{V}$ in one EEG channel and at least greater than $30\mu\text{V}$ in the coincident channel. After being detected by the analog circuit, the signal was passed to the digital circuit where the final decision about REM stage was made based on detected and processed information. Three sleep EEG recordings were used for REM detection. Each of these records were scored by a human scorer and from these records, REM periods of 30 second intervals were selected for analysis. Real time analysis of the recorded EOG data was done to test for system's repeatability as determined by human expert on EEG's. The system's repeatability was assessed by comparing the result of real time analysis of REM period from three EEG recordings. The system was capable of accurate REM detection and had high repeatability from the same record or between different records.

Another automated method for REM detection using EOG (Electro-oculograms) was developed by Gopal and Haddad [16]. Their method was formulated as a sequential decision process based on slope and amplitude threshold criteria. The EOG data was sampled at 25Hz and filtered using a 3rd order low pass Butterworth filter with cut-off frequency of 6.5Hz as they believed that 95% of the signal power was concentrated below 6.5 Hz. From the EOG data obtained, 5 minutes of REM data was divided into 10-s epochs. Two sleep experts independently counted the eye movements present in these 10 seconds epoch. The score of the two observers were average and used as the visual

score in further analysis. The sampled points in the EOG signal that were thought to contain information on the presence or absence of eye movements were identified and termed as the decision points. At each of these points, a decision was made based on the amplitude and slope thresholds.

It was observed that the slope associated with the initial rise or the initial fall of an eye movement wave form was, on average, steeper than the slope associated with any other segment of the EOG signal and the peaks associated with an eye movement waveforms were, on average, higher- than peaks not associated with other segments of eye movement waveforms. The epoch was called as REM only if the value of slope and the amplitude were greater than the threshold.

The optimal threshold values for the slope and amplitude were decided as the ones which provided the highest value of correlation coefficient between visual and system count of REM, the first order regression coefficient obtained should be between 0.9 and 1.10 and zero order regression coefficient should not statistically be different from zero. The system was tested using EOG data of three infants and the automated count of eye movements was compared against the visual counts from human observers. The correlation coefficients obtained between visual and system count of REM were almost always greater than 0.8.

2.3 Methods Developed for Automated Sleep Staging Incorporating the Non-stationarity of the Signal

A number of computer methods for automated sleep staging incorporate the challenge posed by the non-stationary nature of EEGs. One work of this kind was done by Gaillard and Tissot [17] in 1973, which was in continuation with their own work done

earlier where they had described a hybrid computer capable of performing automatic sleep staging of entire night tape recordings of human sleep EEG, EMG, ECG and EOG. They used the power content in various frequency bands as features for detection of five elementary activities mentioned below and their method consisted of two stage analysis: 1) Identification of the elementary patterns in the recorded data for each 4 second segment; 2) Classification of each 60 second epoch based on the frequency of occurrence of the identified elementary patterns. They detected five elementary activities: alpha rhythm (7-12 Hz), spindles (burst activity of 11-17 Hz), K-potentials, saw tooth waves and delta activity (1-3 Hz). In this work, 15 sleep night recordings of 10 subjects were used and five activities were detected: alpha activity, spindles, K-complexes, saw tooth waves and delta waves.

They achieved 83% agreement with visual scoring and the classification accuracy of stage 1 and stage 2 was more than the classification accuracy of stage 3 and stage 4, but there was very good agreement in the classification between the alternate or higher stages like stage 1 and 3, stage 4 and 2 or awake and 4 but larger disagreement was observed between the consecutive stages. In spite of these limitations, this method worked well for awake, stage 3 and stage 4 but had lower accuracy for stage 1 and 2.

Further work in this field was based on the assumption of piecewise stationarity, i.e., the EEG is considered to consist of variable length segments and each one of them is stationary. This kind of approach to deal with non stationarity was first suggested by Bodenstein and Praetorius [18]. Their work was based on three propositions: 1. The EEG consists of quasi-stationary segments on which the transients can be super-posed. 2. Each segment is specified by its length and power spectral estimate and each transient by a

grapho-element and time of occurrence. 3. The EEG consists of finite number of recurrent states. The third proposition needed a clustering method to cluster the similar elements together and then use pattern recognition method to do automatic sleep staging although they did not perform the sleep staging.

Bodenstein and Praetorius used linear prediction or auto regressive modeling. Such a model predicts the next signal value on a weighted summation of past samples. The weights, or model coefficients, are estimated from a data segment, and completely characterize the spectral content of the signal. The inverse model, or inverse filter, was constructed next and the signal was passed through it, producing white noise as long as the signal 'fits the model.' A deviation from whiteness indicates then the end of a stationary segment and the need for a new model. The EEG signal was band filtered between 1 Hz- 25 Hz and sampled at 50Hz. The predictor order was chosen as 6 based on the fact that rarely more than two fast activities occurred in a segment. Using this method, a single channel of EEG could be compressed by a ratio of 1:50.

Work along similar lines was done by Jansen et al. [5]. They too considered the EEG as composed of a finite set of graphic objects (elementary patterns). These elementary patterns were obtained by dividing the EEG in short intervals (1.28 s), which were presumed to be stationary, and applying a clustering procedure. Various features were extracted from these segments, but the most useful were found to be the fifth order autoregressive model coefficients estimated using a Kalman filter. The feature vectors were input to a clustering process to identify a library of elementary patterns. Once the library was established, EEG segments could be classified by identifying the elementary pattern most resembling the EEG segment under study. Next, 'profiles' were computed

depicting how often each elementary pattern occurred in a longer (say 30 s) segment. Average profiles were computed for each sleep stage and used to classify new sleep data. This method achieved almost 80% correct classification using 3 hour recording per night for three subjects.

2.4 Method for Automatic Sleep Staging Based on Artificial Neural Networks and Wavelet Transforms

More recently, neural networks and wavelet transforms have been used for automatic sleep stage scoring. In 2008, Ebrahimi et al. [19] used neural networks and wavelet packets coefficients for automatic sleep staging. Here wavelet transformation [20] was applied to 30 second epochs of Fpz-Cz/Pz-Oz EEG recordings. The EEG signal was classified into 6 frequency bands: delta activity (0-3.5 Hz), alpha activity (8-10 Hz), spindle (10-16 Hz), theta activity (3-8 Hz) and beta1 (15-21 Hz) and beta 2 (21-30 Hz) activity. Feature vectors were extracted from wavelet packet coefficients. They extracted five features: 1) Energy of the wavelet packet for each of the 6 bands; 2) Total energy which was the sum of energy for each of the 6 bands; 3) Ratio of the different energy values; 4) Mean of the absolute coefficient values; 5) Standard deviation of the coefficients in each sub bands. Then these feature vectors were processed using an artificial neural network with one hidden layer and four neurons in the output layer for discrimination between sleep stages Awake, Stage 1 + REM, Stage 2 and Stage 3.

The error backpropagation algorithm was used to train the network. Best performance was produced using 8 neurons in the hidden layer, with no significant enhancements observed beyond this number. They used the data available from five subjects each 3 hour recording per night. Out of the five subjects the first two subjects

were used for training purpose and remaining three was used for testing the system. The results indicated that this method could discriminate between awake, Stage 1 + REM Sleep, Stage 2 and stage 3 with a specificity of $94.4 \pm 4.5\%$ and sensitivity of $93.0 \pm 4.0\%$.

Chapter 3 Methods

The method used here involves piecewise analysis of the EEG along the lines described by Jansen et al. [3]. The original method was modified by adding detection of sleep transients including sleep spindles and REM. Also, a larger data set than the three subjects studied by Jansen was analyzed. We have used the Parietal channel data for the EEG analysis

3.1 Basic Outline of the Method

1. First, the EEG data is preprocessed.
2. Features are extracted from each 1 second segment of the EEG signal.
3. During the training phase, these features are fed into a clustering process to determine a library of elementary patterns. This is followed by a classification stage in which ‘reference profiles’ for each sleep stage are determined. These profiles show the average frequency of occurrence of each elementary pattern in a 30 second interval.
4. In the testing phase, profiles for each 30 sec epoch are obtained, which are compared to reference profiles to determine the sleep stage. We then apply the transient rules as mentioned below for detection of REM and spindles.
5. Since, REM stage cannot be separated using EEG only, we will use the EOG (eye movement) channel to identify REM. In case of stage 2, we will use a secondary analysis involving the number of spindles detected in a 30-s interval. Details of all these steps are provided in subsequent sections. Figure 3.1 graphically represents these procedures.

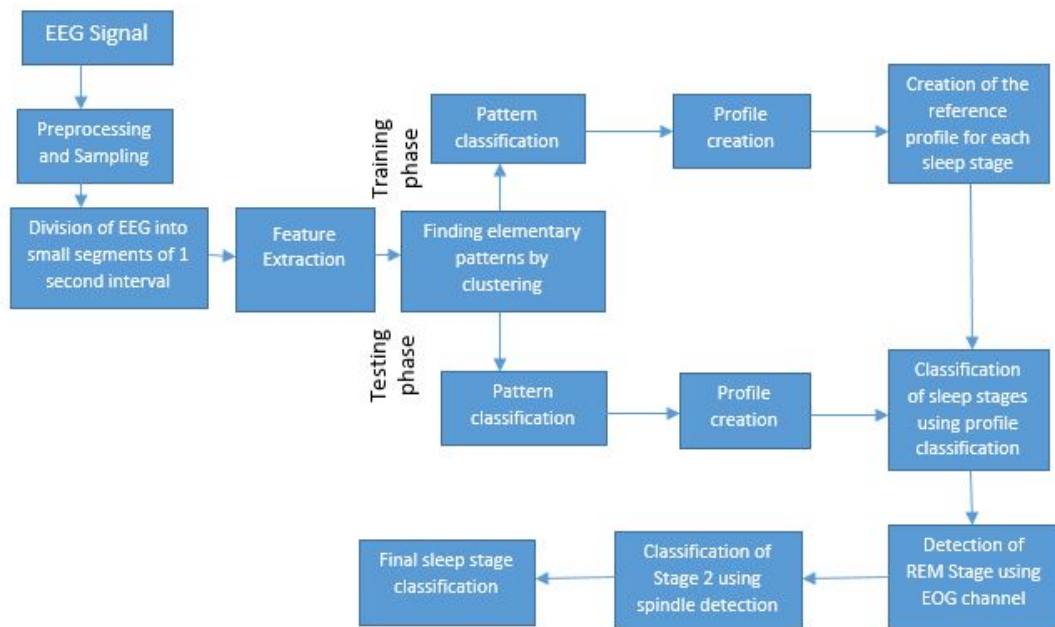


Figure 3.1 Flowchart for the sleep stage Classification.

3.2 Data Preprocessing

The steps performed for data preprocessing are as follows:

1. The data was filtered using a low pass Butterworth filter of 10th order and cut-off frequency of 30 Hz because the frequencies of interest for waveforms present in EEG are in range of 0-30Hz. The low pass filter acts as an anti-aliasing filter which restricts the bandwidth of signal to satisfy the sampling theorem.
2. Down sampled the signal to 64 Hz. This frequency was selected after some experimentation.
3. The EEG was segmented into 1 second, non-overlapping epochs.
4. Mean of the signal was removed on a second-by-second basis. This was done by subtracting the mean of each 1 second segment from its samples.

3. 3 Feature Extraction

Autoregressive modeling was used for feature extraction. The mathematical form of the AR model is

$$x(N) = \sum_{i=1}^p a_i x(N - i) + \epsilon(N). \quad \text{Equation 3.1}$$

Here p is the model order, the weights a_i are the autoregressive coefficients and $\epsilon(N)$ is the error term which is assumed to be white Gaussian noise. The Burg algorithm [21] was used to estimate the model coefficients.

The number of frequency components present in the signal that can be modeled by the AR model is determined by the order p . The type of EEG activity seen during sleep rarely has more than two frequency components as can be seen from Figures 3.2-3.6. These figures present the average spectra, obtained from 1-s intervals, for each of the sleep stages using fast Fourier transform and 5th order AR model, since no more than one clear peak (and a low frequency component) are observed, a fifth order model is sufficient. The same model order was also used by Jansen et al. [6]. We also computed the variance for each 1 second segment as a sixth feature. We used all the six features for the classification of the sleep stages as all the features were found to be helpful.

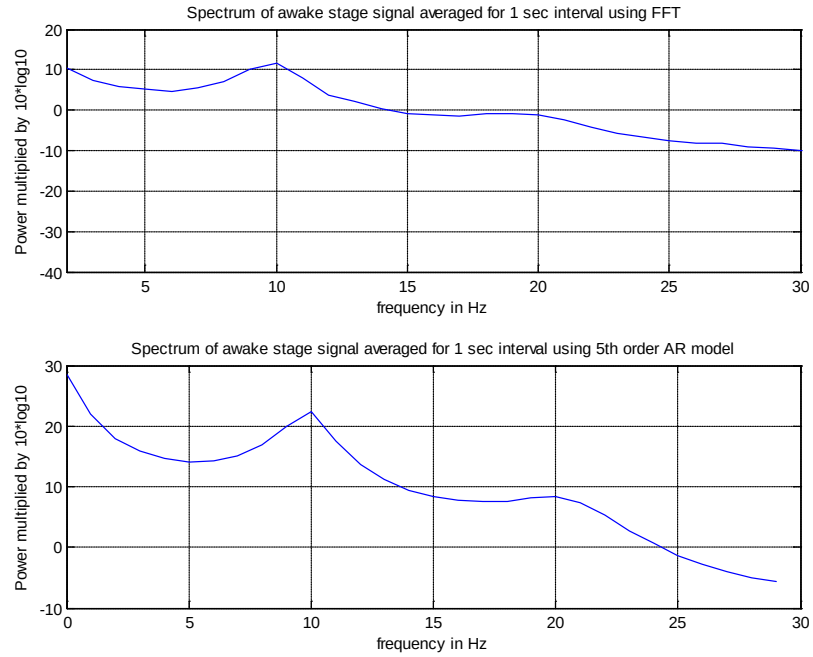


Figure 3.2 Comparison of power spectrum for awake stage using FFT and AR model.

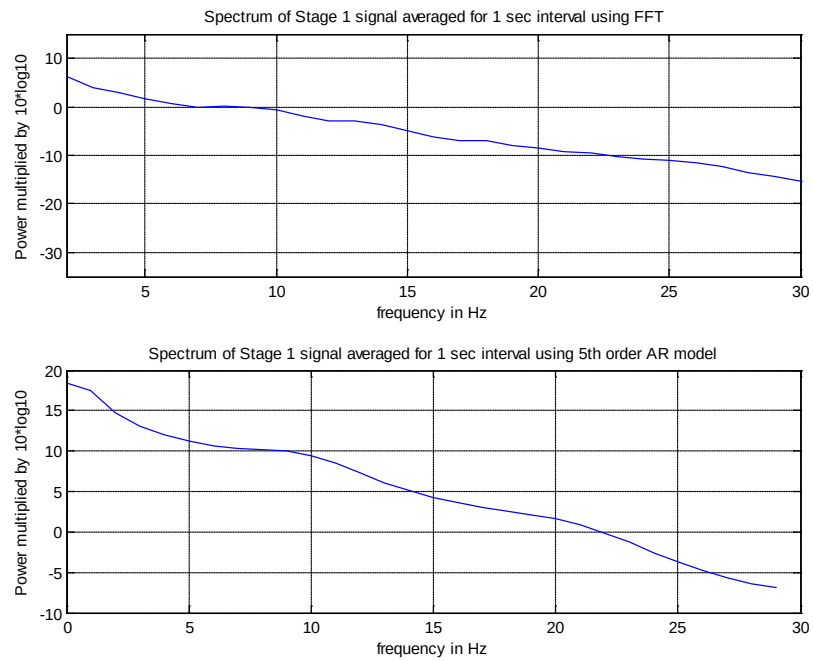


Figure 3.3 Comparison of power spectrum for sleep stage 1 using FFT and AR model.

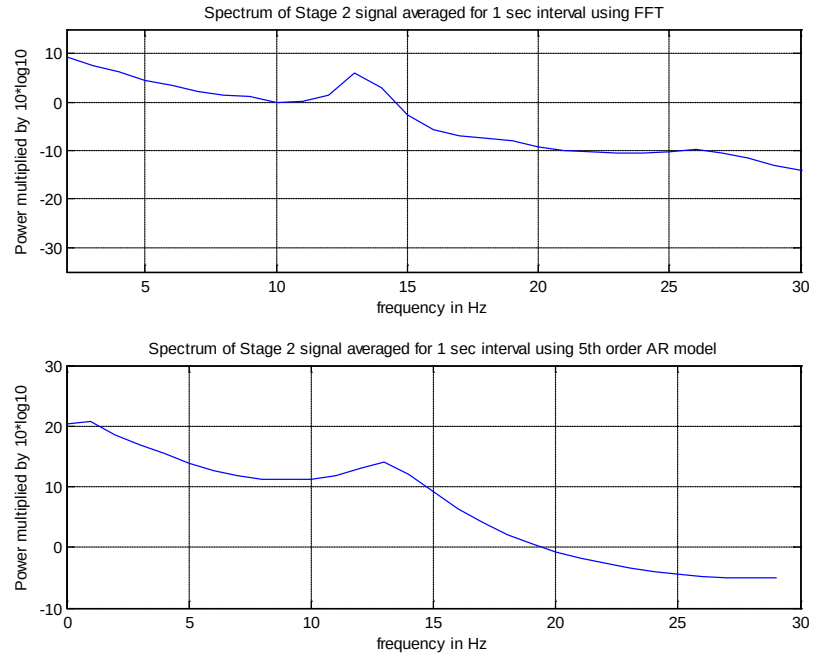


Figure 3.4 Comparison of power spectrum for sleep stage 2 using FFT and AR model.

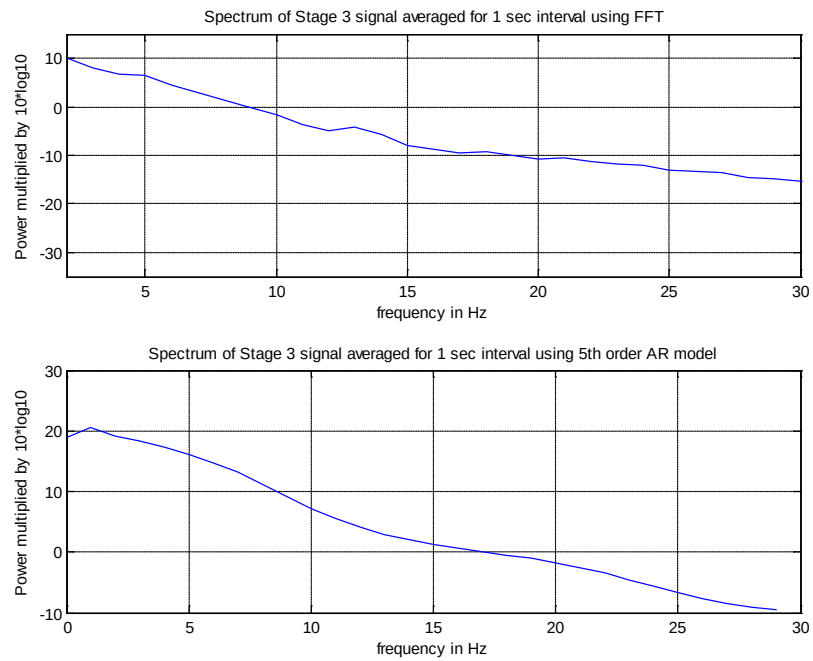


Figure 3.5 Comparison of power spectrum for sleep stage 3 using FFT and AR model.

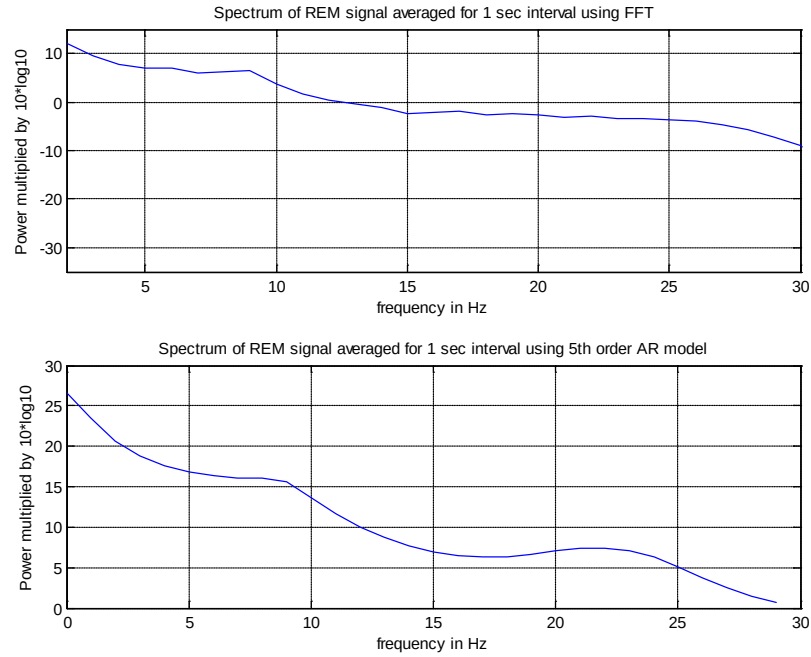


Figure 3.6 Comparison of power spectrum for REM stage using FFT and AR model.

3.4 Clustering

The extracted features were fed as an input in the clustering process. Since the number of clusters present in our data was not known, we used hierarchical clustering to determine the number of clusters. In the hierarchical clustering algorithm, data is grouped over a variety of scales by creating a cluster tree or dendrogram. In our case, the scale used is the Euclidean distance between the feature vectors. The cluster tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next level. Specifically, we separated the features based on the sleep stage they belonged to and the clusters were obtained individually from each sleep stage. Hierarchical clustering was performed using the data of subject 1 and we obtained 4 clusters. Each cluster represents a specific sleep wave activity as given below. By cutting

the hierarchical tree at some empirically-determined level, clusters are obtained. The approach used to obtain clusters from EEG data is as follows:

1. Hierarchical clustering was performed for each individual sleep stage separately, the cluster tree obtained from the hierarchical clustering was cut at a level of 5 and the clusters obtained after chopping off the cluster tree were analyzed for their power spectrum and wave activity.
2. The clusters showing similar power spectrum and sleep wave activity were combined together to get a single cluster for each kind of sleep activity such as alpha, theta, spindle, delta and REM.
3. Thus, separate clusters for each sleep activity was selected after analyzing the power spectrum of the sleep data present in that cluster according to table 3.1.

The table below lists the frequency range used for the selection of clusters for each kind of sleep activity.

TABLE 3.1 Sleep wave activity and their frequency range.

Sleep Activity	Frequency Range
Alpha activity	8-12 Hz
Theta activity	4-8 Hz
Spindle activity	12-15 Hz
Delta activity	0-4 Hz

Our approach was to first get the clusters from a single subject. These clusters were then used to classify the data of a second subject, and if this was done successfully, as

indicated by the Likelihood test (discussed in the next section), the process was halted. If classification was not successful, clusters from the second subject were extracted and added to the cluster library. This process was repeated until no new clusters were needed. The dendrograms of each sleep stage and the level at which the cluster tree was cut are shown in Figures 3-7 through 3-10. We cut the cluster tree at a level of 5 in all the cases.

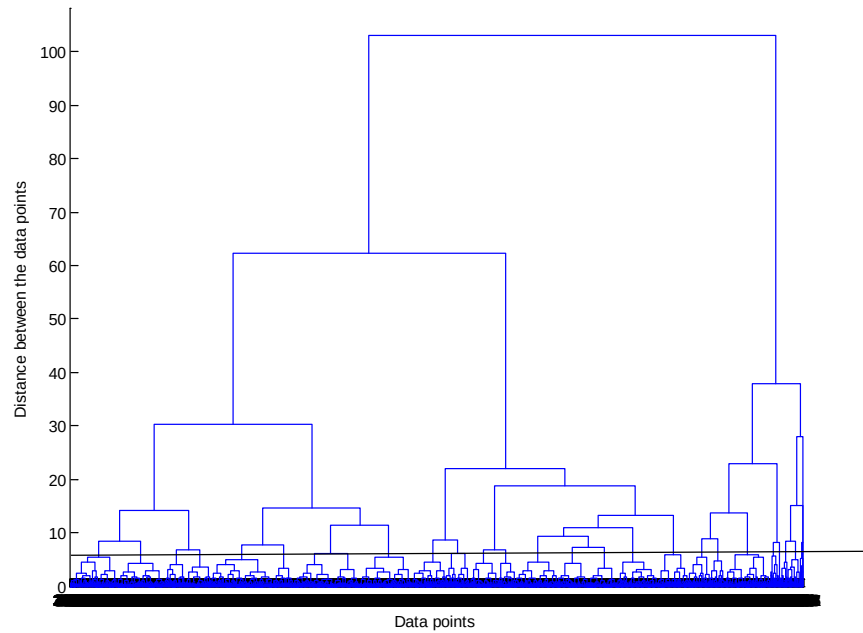


Figure 3.7 Dendrogram of Awake stage data from Subject 1.

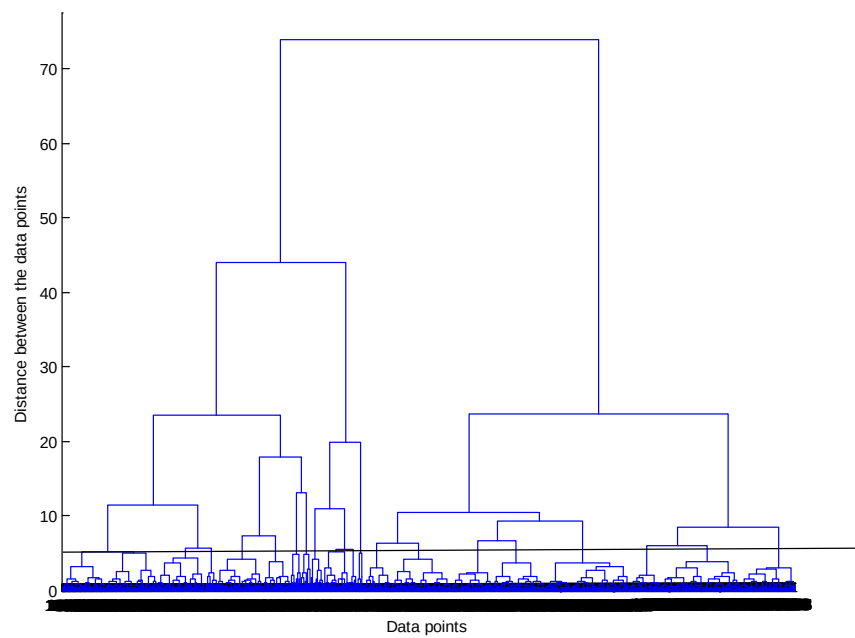


Figure 3.8 Dendrogram of Stage 1 data from Subject 1.

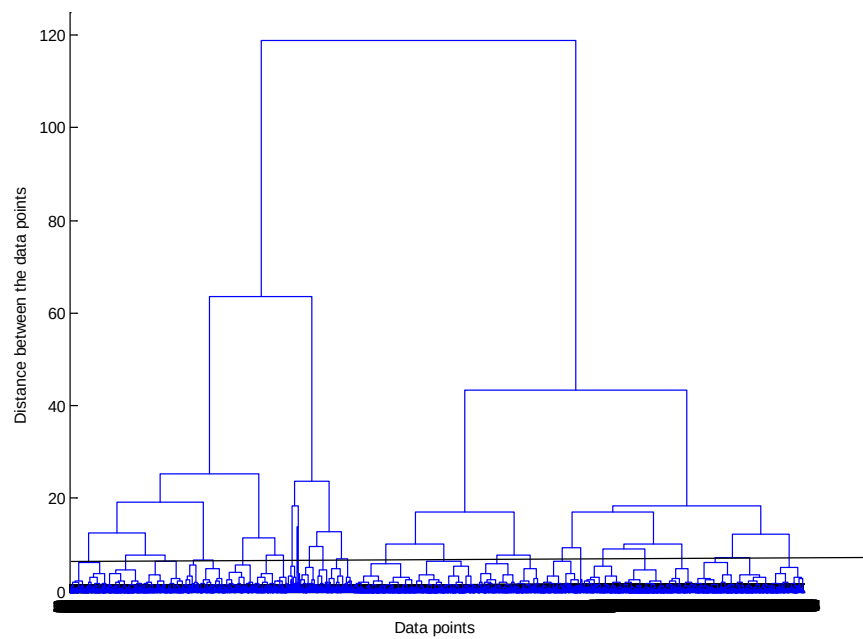


Figure 3.9 Dendrogram of Stage 2 data from Subject 1.

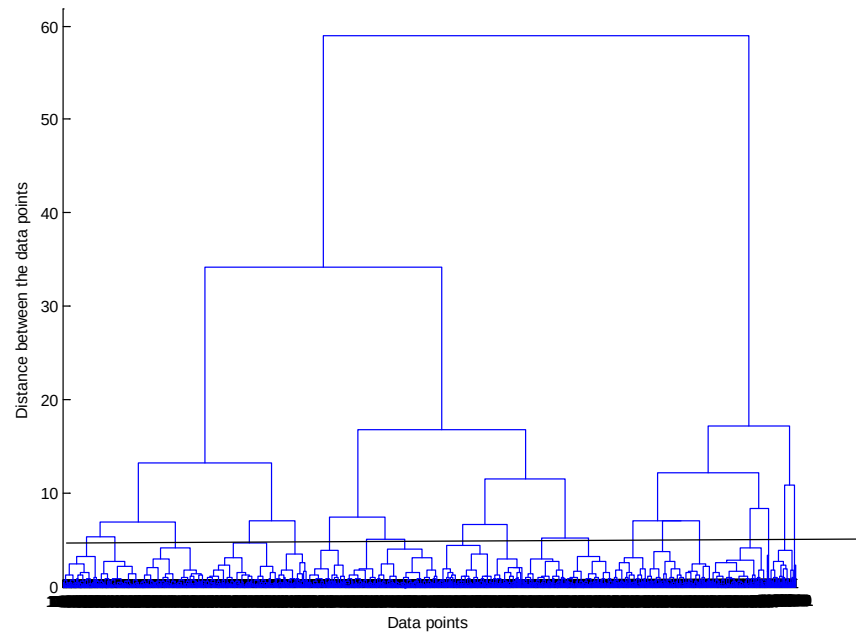


Figure 3.10 Dendrogram of Stage 3 data from Subject 1

The clusters obtained for each kind of sleep activity and the 1 second segment closest to the centroid of each cluster are shown in figure 3.11 and 3.12 respectively.

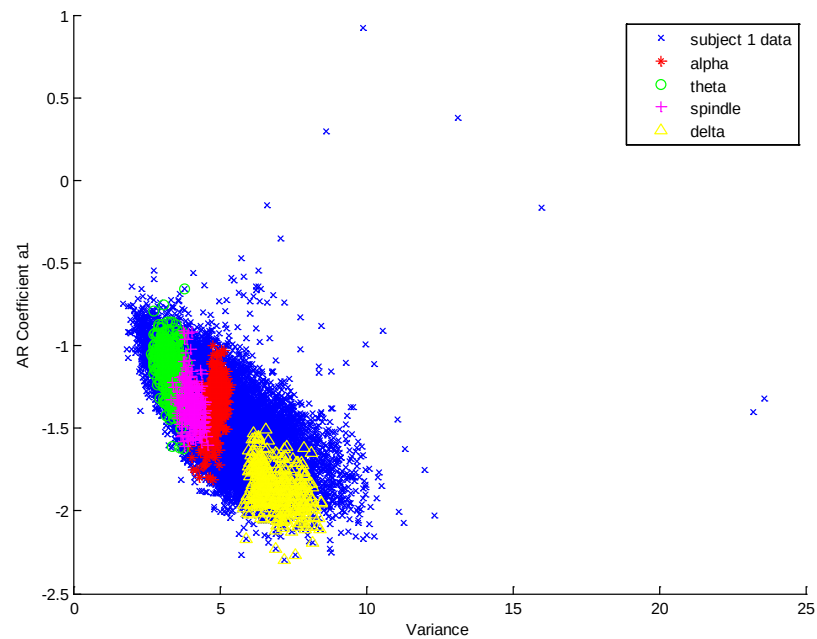


Figure 3.11 Clusters obtained from Subject 1 data.

The time series for the 1-second segment closest to each cluster centroid obtained from subject 1 are shown in the Figure 3.12.

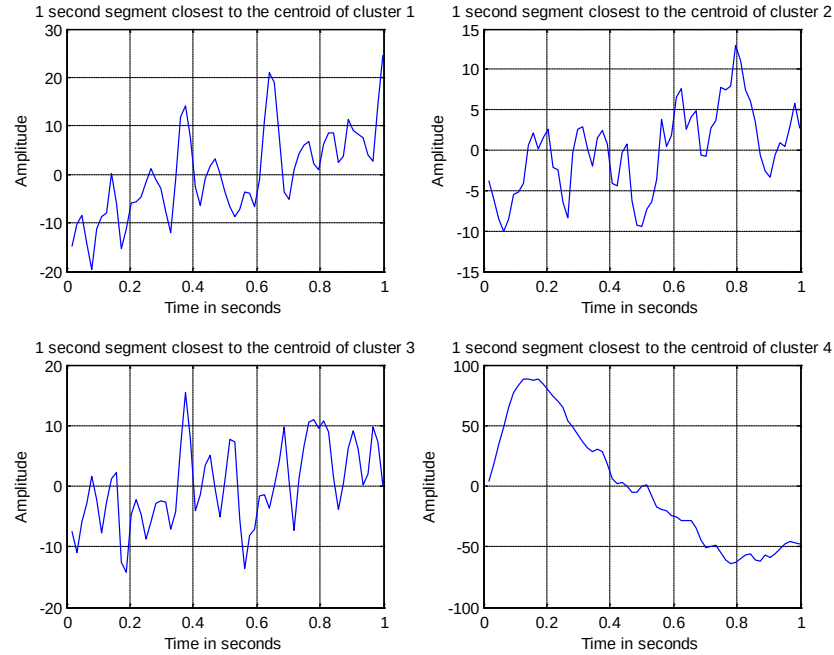


Figure 3.12 One second time series for the four clusters selected.

In order to check whether the clusters obtained were good enough to cover the entire data of subject 1 and the other training subjects, we conducted a likelihood test. This test is explained below.

3.5 Likelihood test

In the likelihood test, we classify the sleep data of training subjects using the cluster library and compute the likelihood of each feature vector belonging to its assigned pattern. The histogram of these likelihoods is then computed and inspected visually. Ideally, the likelihood histogram will be a monotonically increasing function as the

likelihood increases, with preferably no small values (say less than 0.5). This will indicate that the data is covered well by the clusters.

We begin with the sleep data of subject 1. This data was classified using 4 clusters and the likelihood of the data belonging to each cluster was calculated. The histogram of the likelihood for data covered by each cluster is shown in Figure 3.13.

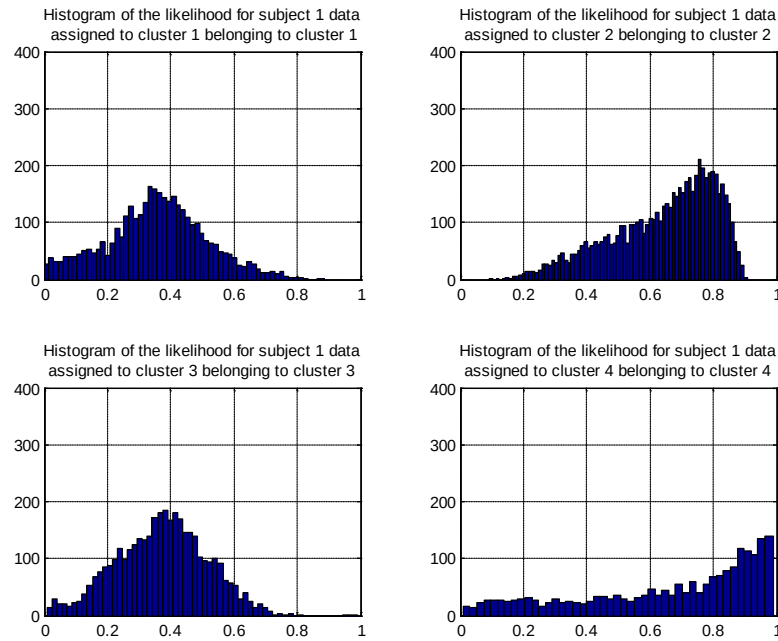


Figure 3.13 Histogram of likelihood for Subject 1 data belonging to the patterns selected from Subject 1.

It can be seen from Figure 3.13 that the likelihood of data being covered by pattern 2 is good but the likelihood histograms for the other patterns deviates a lot from our expectations. The likelihood histogram for pattern 4 can be ignored because the 4th cluster (pattern 4) is far from other three patterns and in spite of the likelihood of data belonging to pattern 4 being low, we expect to get good classification using that pattern. Since the problem seems to be with pattern 1 (alpha activity) and pattern 3 (spindle

activity), we now obtain the clusters for alpha, spindle and delta activity from subject 2 and we replace pattern 3 obtained from subject 1 by pattern 3 obtained from subject 2 and re-compute the likelihood histogram for both the subjects separately. The obtained histograms are shown in Figures 3.14 and 3.15, respectively.

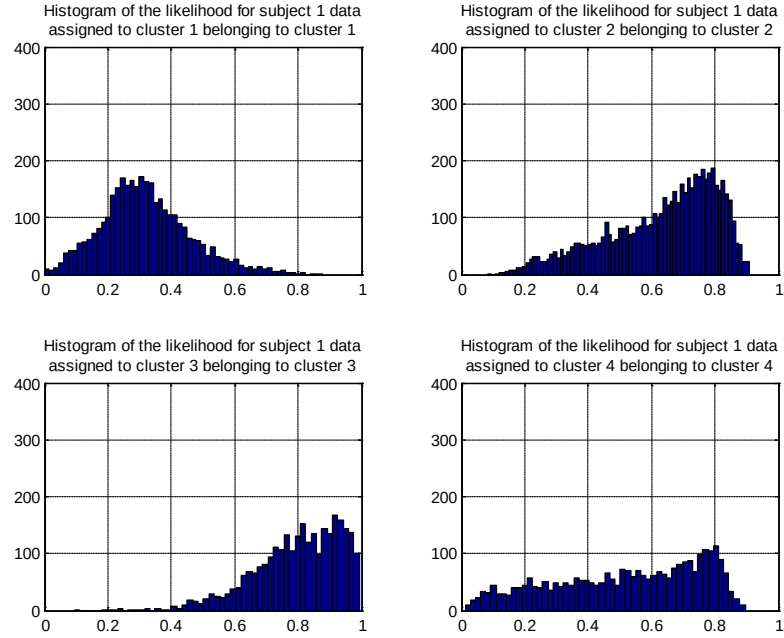


Figure 3.14 Histogram of likelihood for Subject 1 data belonging to the patterns selected from Subject 1 and Subject 2.

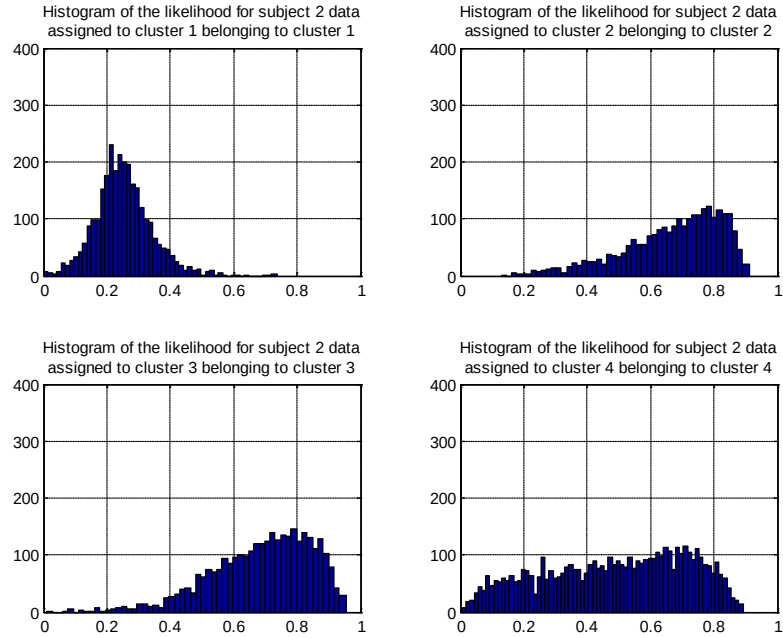


Figure 3.15 Histogram of likelihood for Subject 2 data belonging to the patterns selected from Subject 1 and Subject 2.

Here we observe that the likelihood histogram for the first subject improves for pattern 3 but for subject 2 the histogram of pattern 1 (alpha activity) is not as expected. Therefore, we will represent alpha activity by two clusters; one represented by the alpha activity pattern of subject 2, and the other by the alpha activity pattern of subject 1. We combine the clusters of delta activity from subject 1 and subject 2 into a single cluster representing pattern 4 and the theta activity is represented by two clusters: one by the theta activity pattern of subject 3 and other with the theta activity pattern of subject 1.

Thus, now we have total of seven clusters: two clusters for alpha activity representing pattern 1, two clusters for theta activity representing pattern 2, one cluster for spindle activity representing pattern 3, one cluster for delta activity representing pattern 4 and one cluster for REM representing pattern 5. Next, the likelihood histograms

of for each of the 3 subjects using the seven clusters were obtained and the histograms are shown in Figures 3.16 through 3.21.

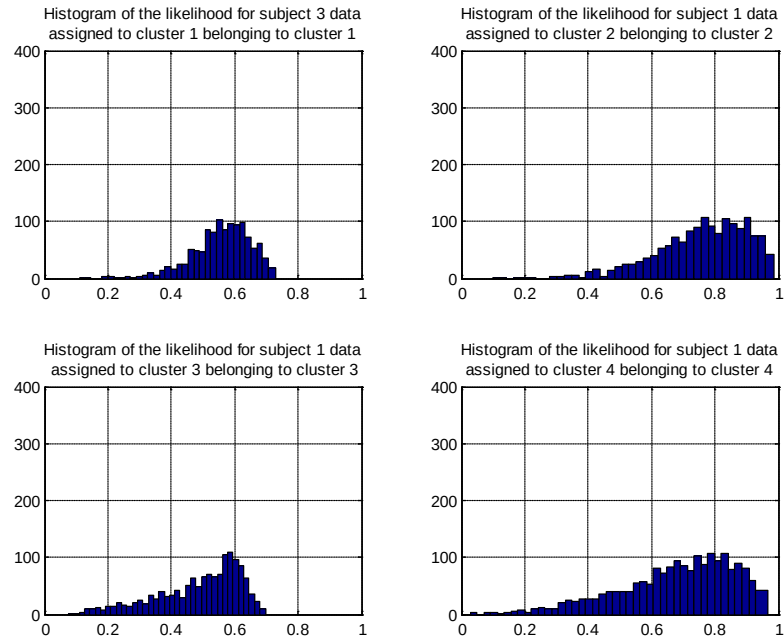


Figure 3.16 Histogram of likelihood for Subject 1 data belonging to the cluster 1, 2, 3 and 4 from the seven selected patterns.

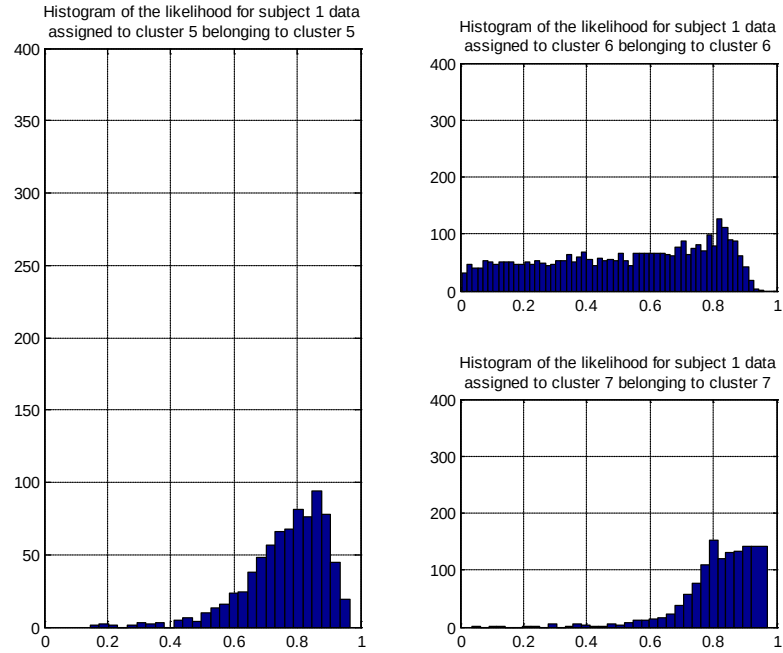


Figure 3.17 Histogram of likelihood for Subject 1 data belonging to the cluster 5, 6 and 7 from the seven selected patterns.

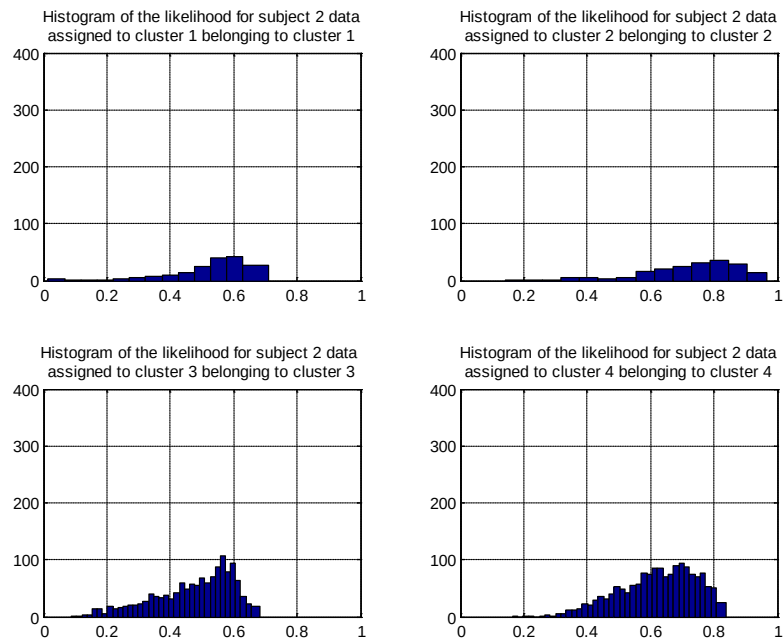


Figure 3.18 Histogram of likelihood for Subject 2 data belonging to the cluster 1, 2, 3 and 4 from the seven selected patterns.

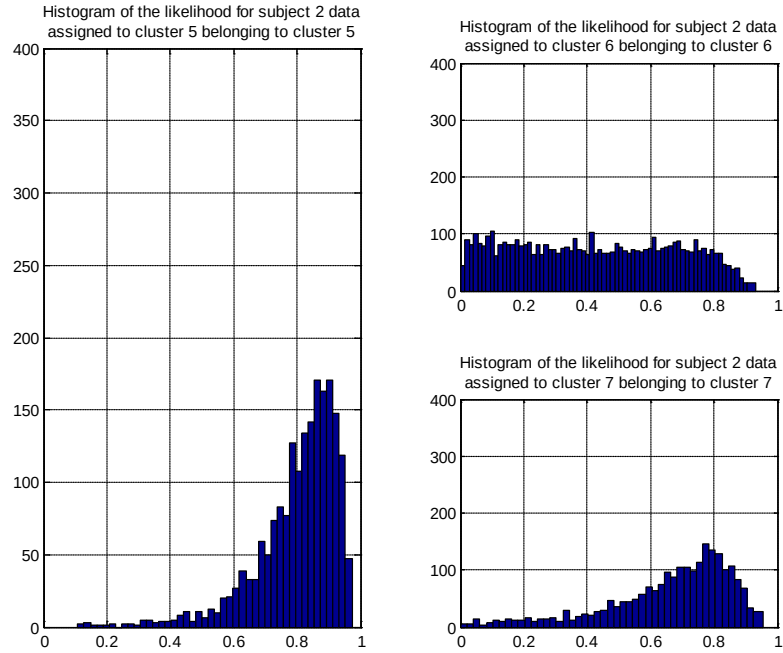


Figure 3.19 Histogram of likelihood for Subject 2 data belonging to the cluster 5, 6 and 7 from the seven selected patterns.

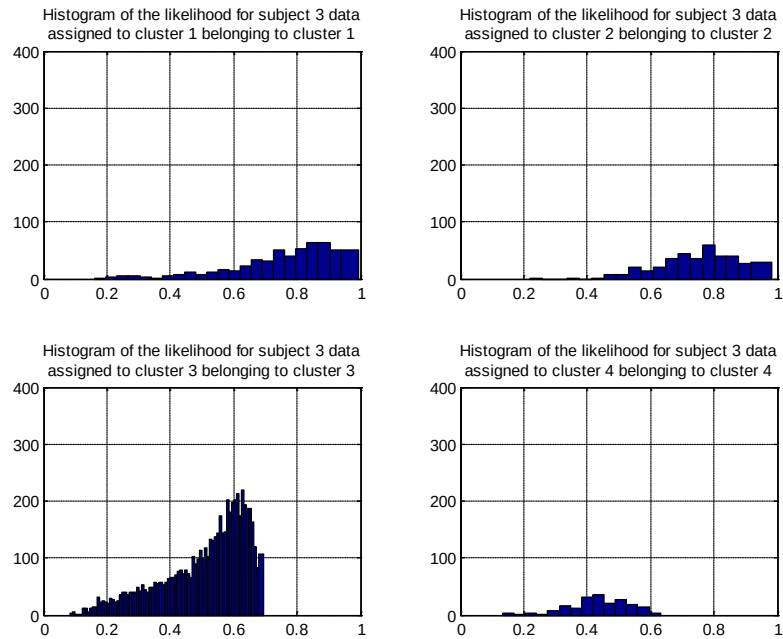


Figure 3.20 Histogram of likelihood for Subject 3 data belonging to the cluster 1, 2, 3 and 4 from the seven selected patterns.

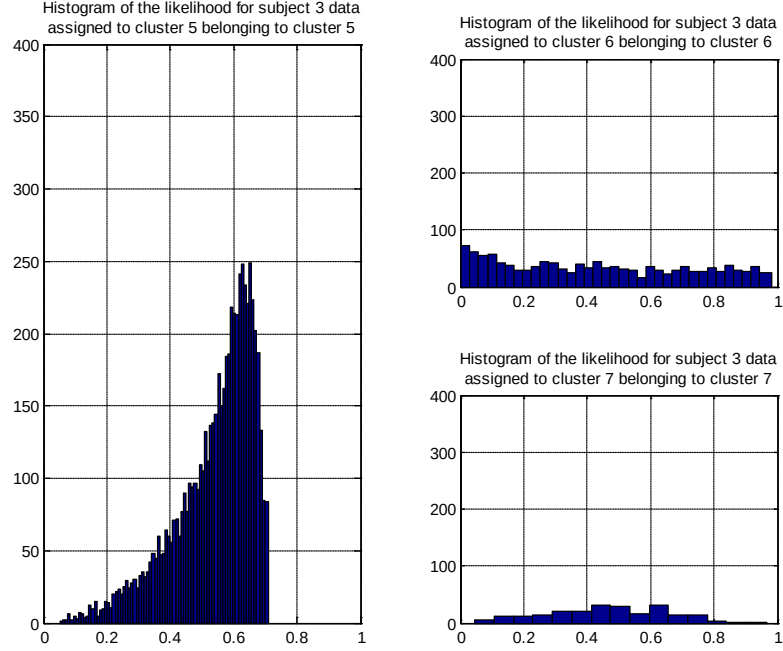


Figure 3.21 Histogram of likelihood for Subject 3 data belonging to the cluster 5, 6 and 7 from the seven selected patterns.

To further refine the classification of single-second epochs, we decided to label epoch as ‘unclassifiable’ if their likelihood was less than a certain threshold. Inspecting Figures 3.16 through 3.21, we choose the threshold value of the likelihood as 0.15 so that there is no significant loss of data and for further classification the 1-second segment having a likelihood lower than the threshold value will be rejected.

Finally the seven clusters selected for classification are shown in Figure 3.22. In order to verify that these clusters are satisfactory, their coverage of subject 4 data was tested by classifying this subject’s data using the seven clusters and inspecting the resulting likelihood histograms, which are shown in figures 3.23 and 3.24. We conclude from the likelihood histograms that data of the training subjects as well data of subject 4 is covered well by our clusters.

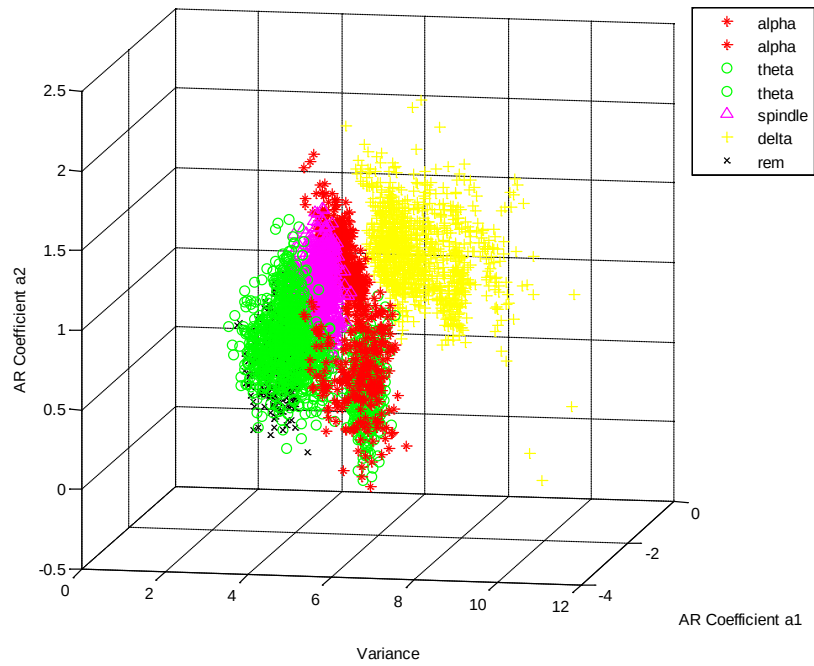


Figure 3.22 Final clusters selected after clustering.

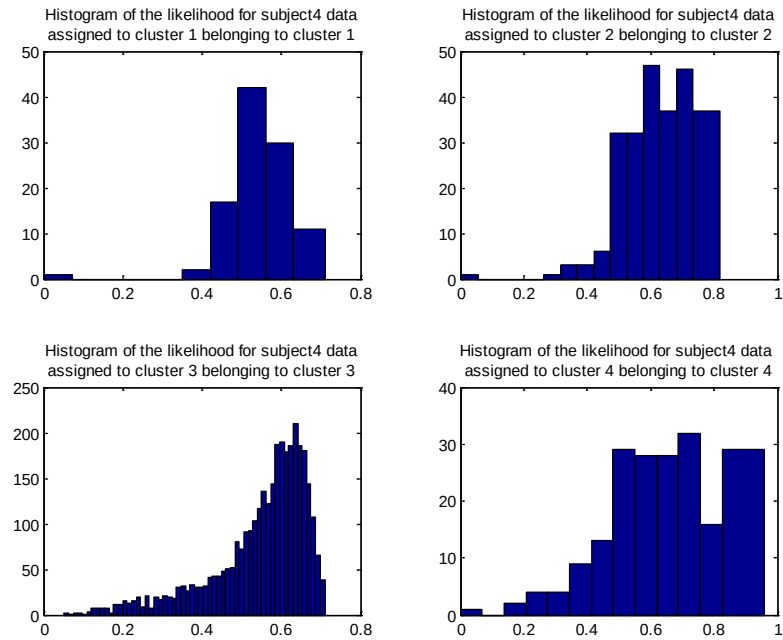


Figure 3.23 Histogram of likelihood for Subject 4 data belonging to the cluster 1, 2, 3 and 4 from the seven selected patterns.

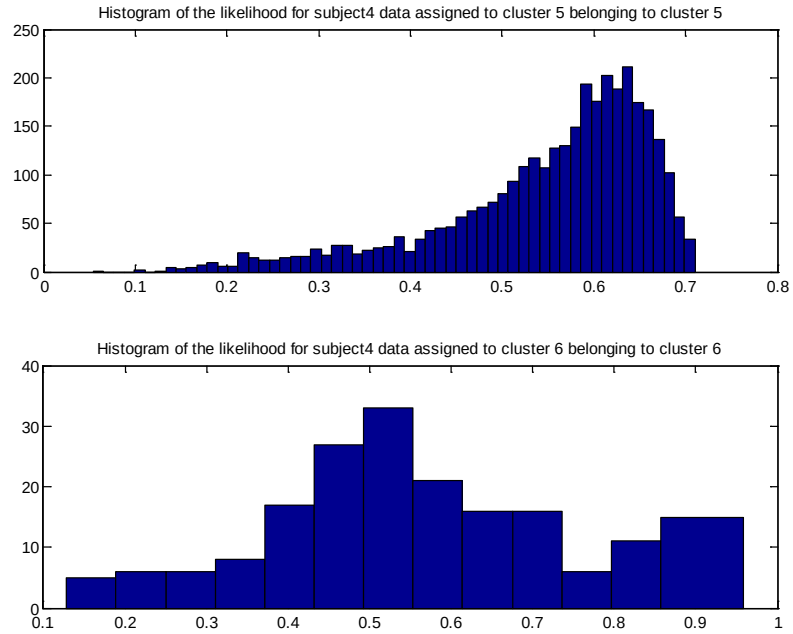


Figure 3.24 Histogram of likelihood for Subject 3 data belonging to the cluster 5 and 6 from the seven selected patterns.

The time series and power spectrum for the 1-second segment closest to each class centroid are shown in Figure 3.25 through 3.35 based on which we can say that the selected clusters are different from each other and each cluster represents a different sleep wave activity.

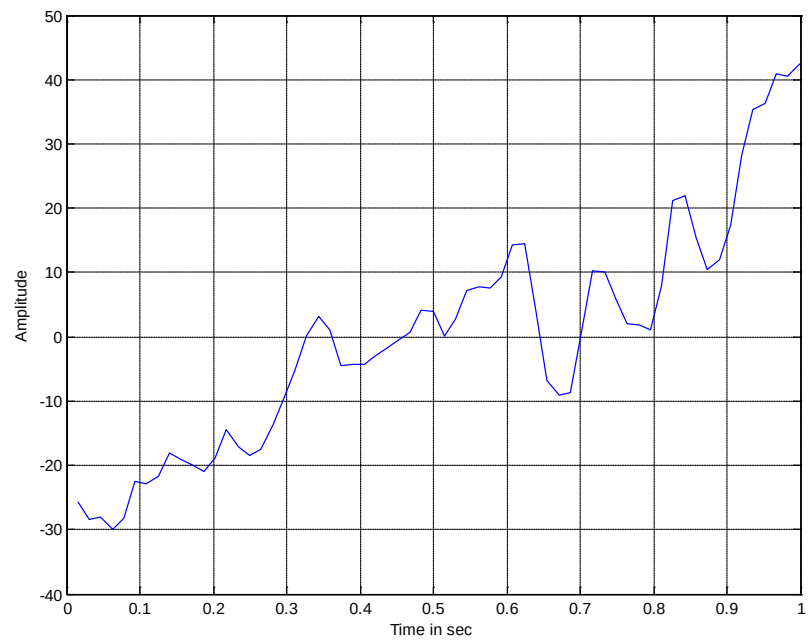


Figure 3.25 One second segment using the centroid of cluster 1 (alpha).

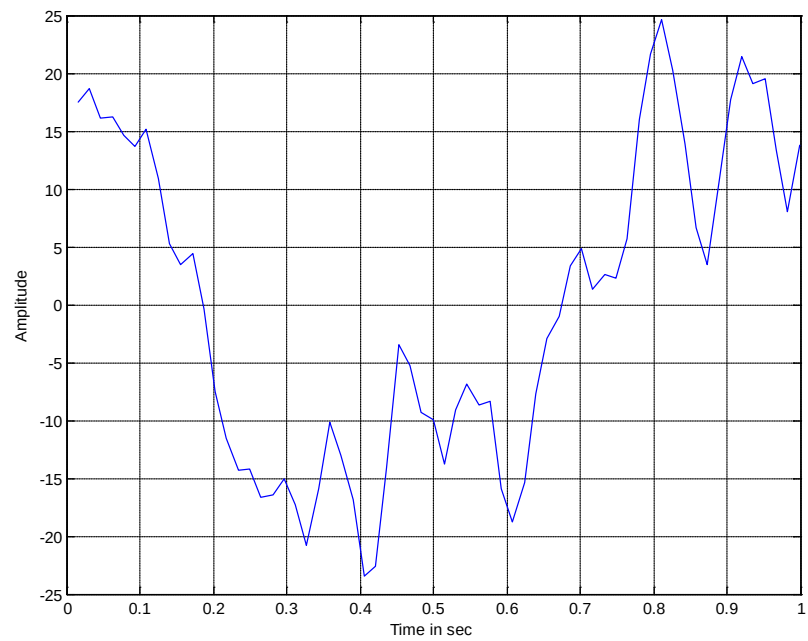


Figure 3.26 One second segment using the centroid of cluster 2 (alpha).

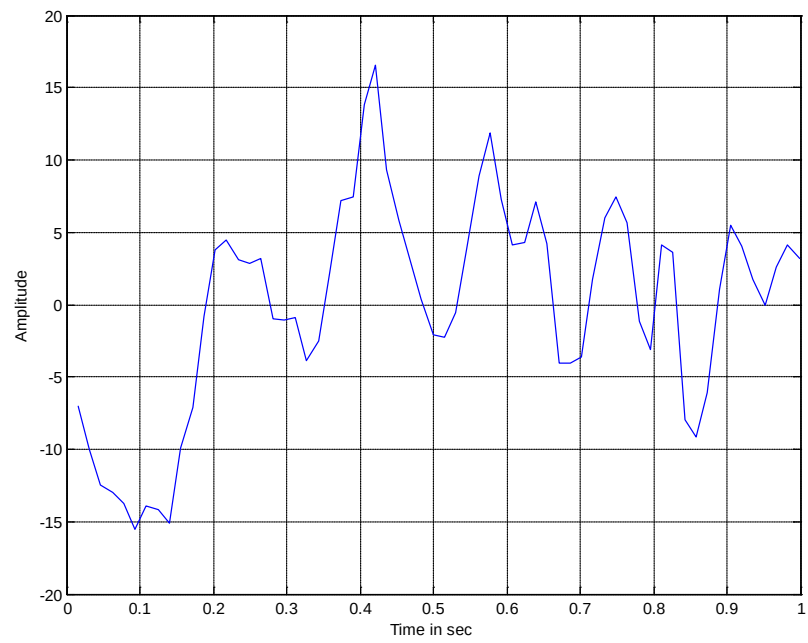


Figure 3.27 One second segment using the centroid of cluster 3 (theta).

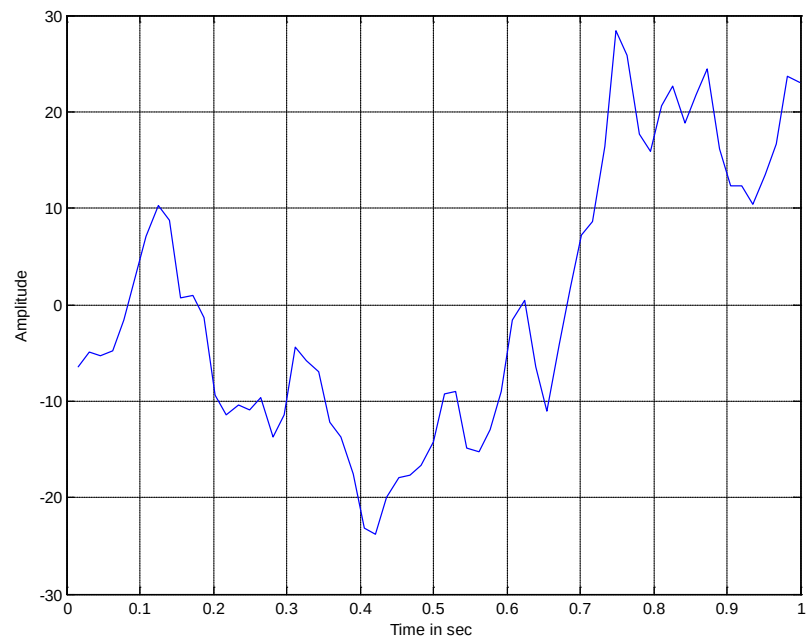


Figure 3.28 One second segment using the centroid of cluster 4 (theta).

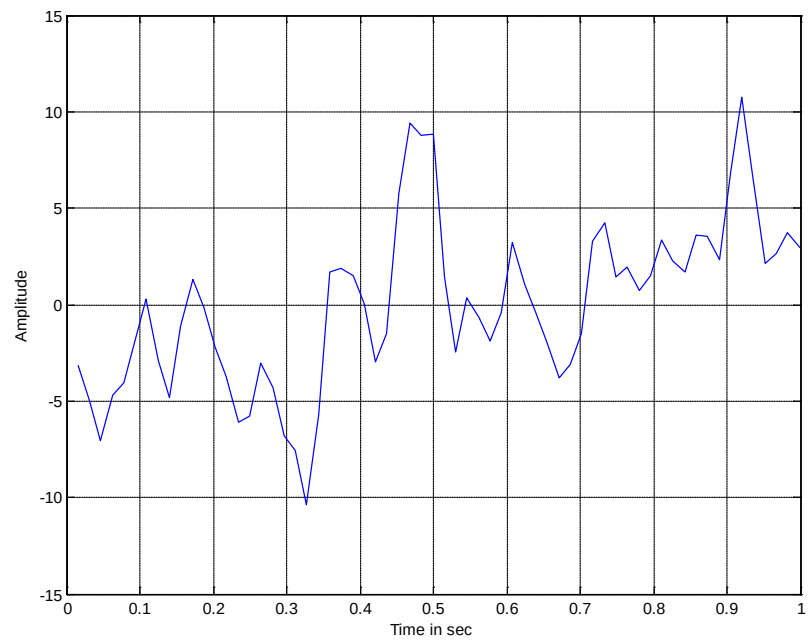


Figure 3.29 One second segment using the centroid of cluster 5 (spindle).

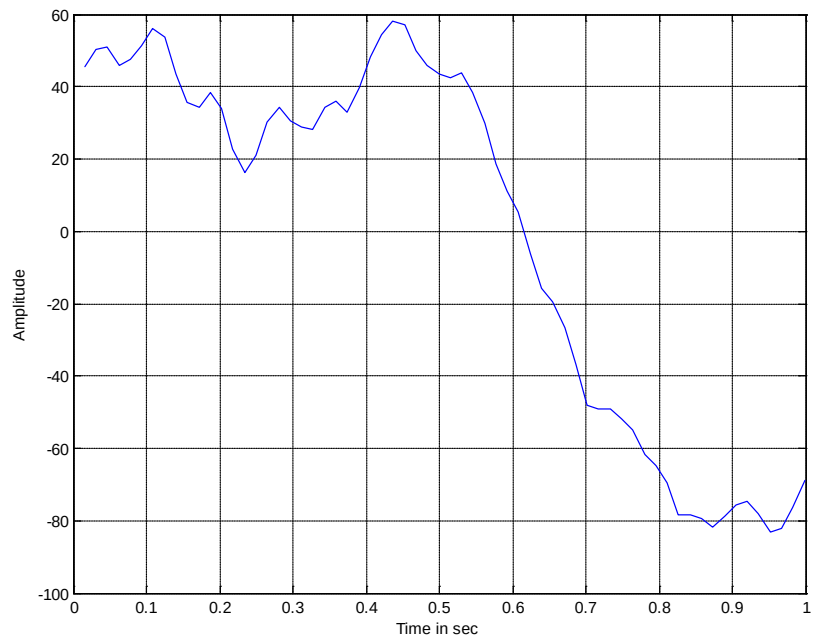


Figure 3.30 One second segment using the centroid of cluster 6 (delta).

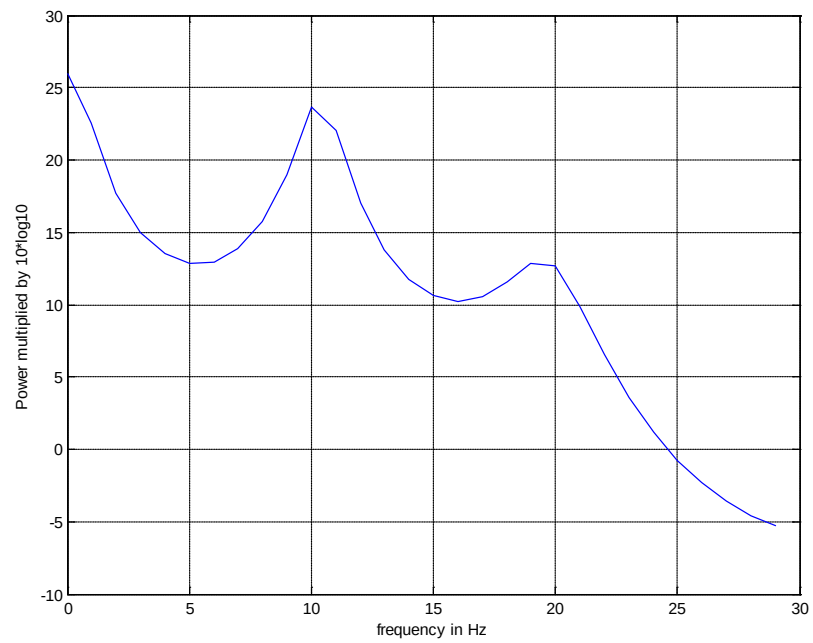


Figure 3.31 Power spectrum of the one second segment closest to the centroid of alpha clusters

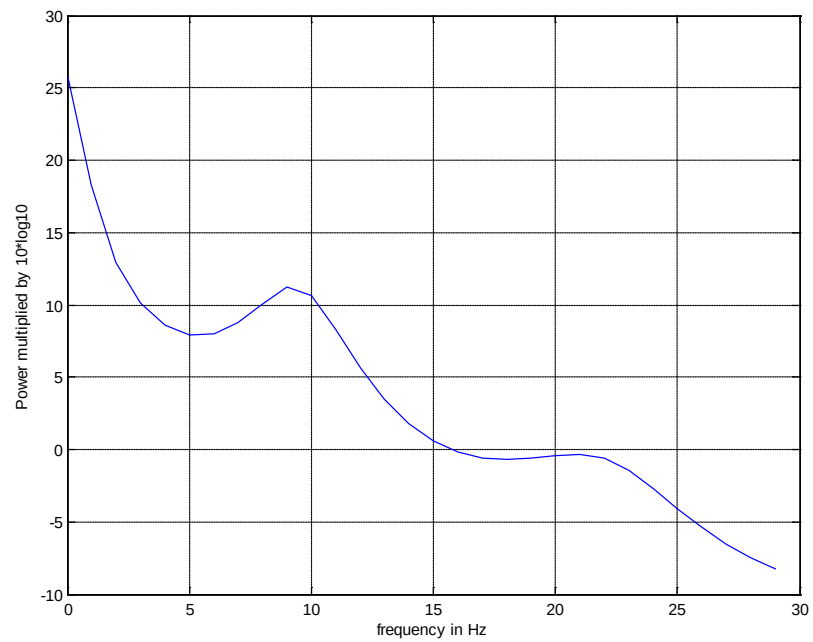


Figure 3.32 Power spectrum of the one second segment closest to the centroid of theta cluster

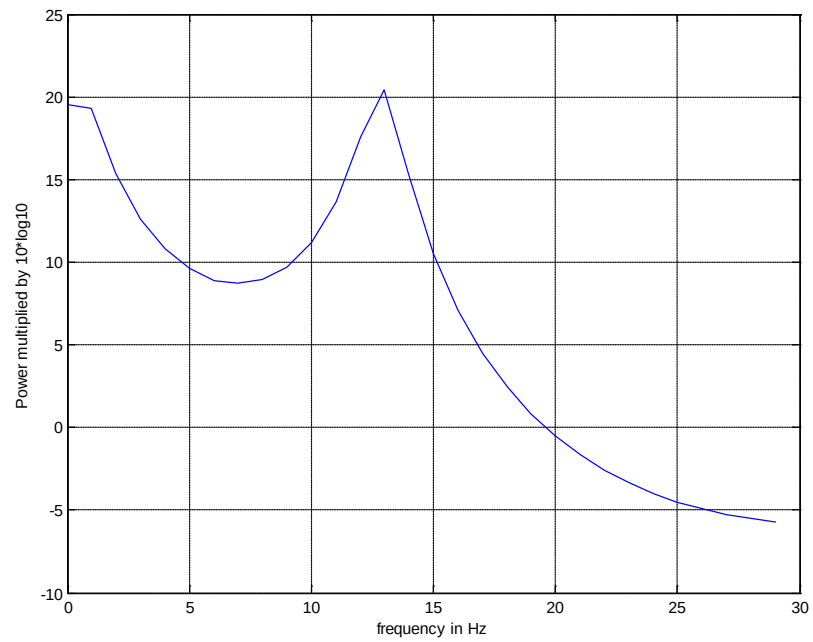


Figure 3.33 Power spectrum of the one second segment closest to the centroid of spindle clusters

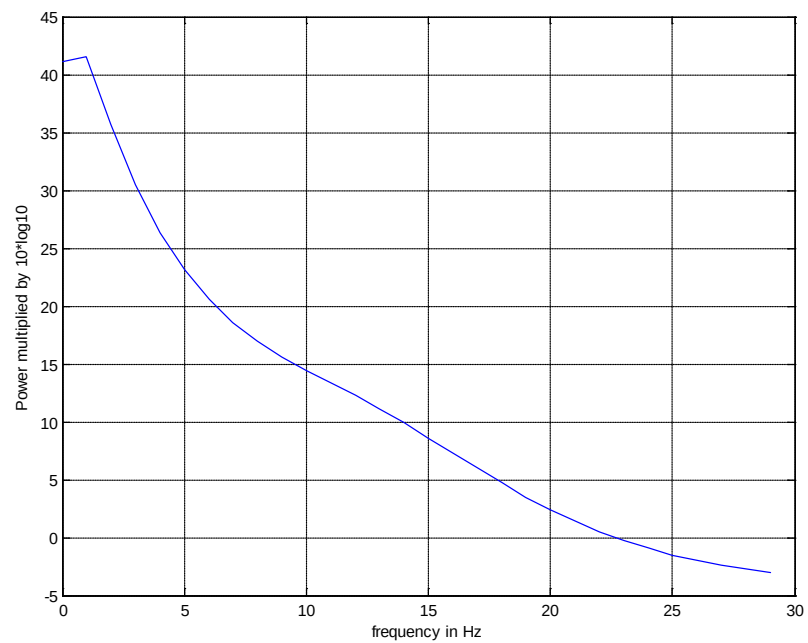


Figure 3.34 Power spectrum of the one second segment closest to the centroid of delta cluster

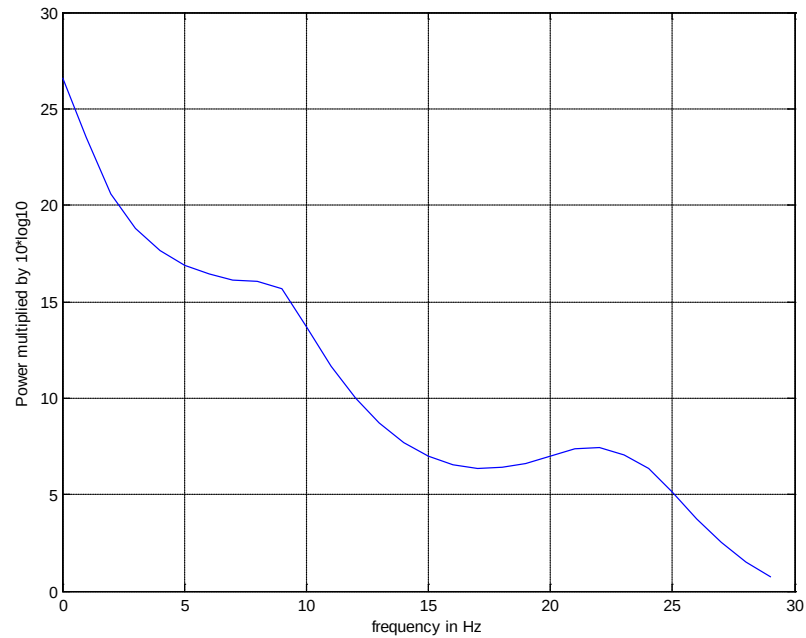


Figure 3.35 Power spectrum of the one second segment closest to the centroid of rem cluster

3.6 Classification

Once the clusters are determined, we can proceed with the next step, i.e., creating classification profiles. A classification profile is a histogram which indicates the frequency of occurrence of each elementary pattern in a 30 second epoch of the EEG signal. A number of classification profiles for 30 second epoch of each sleep stage are averaged to create a reference profile for individual sleep stages as we discussed earlier. The profile obtained from 30 second epoch of testing segment is compared with the reference profile to determine its sleep stage using chi square statistics. The testing epoch is assigned to the sleep stage for which chi square value is minimum. The chi square value is generally computed using

$$x^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}. \quad \text{Equation 3.2}$$

Here,

O_i is the observed profile value of the testing epoch, E_i is the expected profile value of the individual sleep stage, and n is the number of patterns used which is 5 here.

Since the expected value E_i in our case can be 0 and then the expression becomes indeterminate, we have modified equation 3.2 to

$$x^2 = \sum_{i=1}^n (O_i - E_i)^2. \quad \text{Equation 3.3}$$

The averaged Reference profiles created for each sleep stage are shown in Figures 3.36 through 3.40.

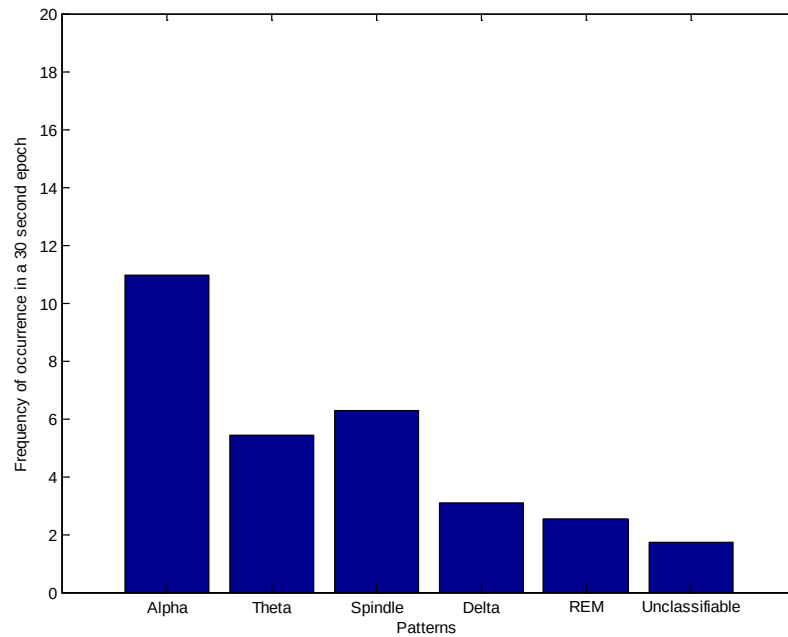


Figure 3.36 Reference profile for Awake stage.

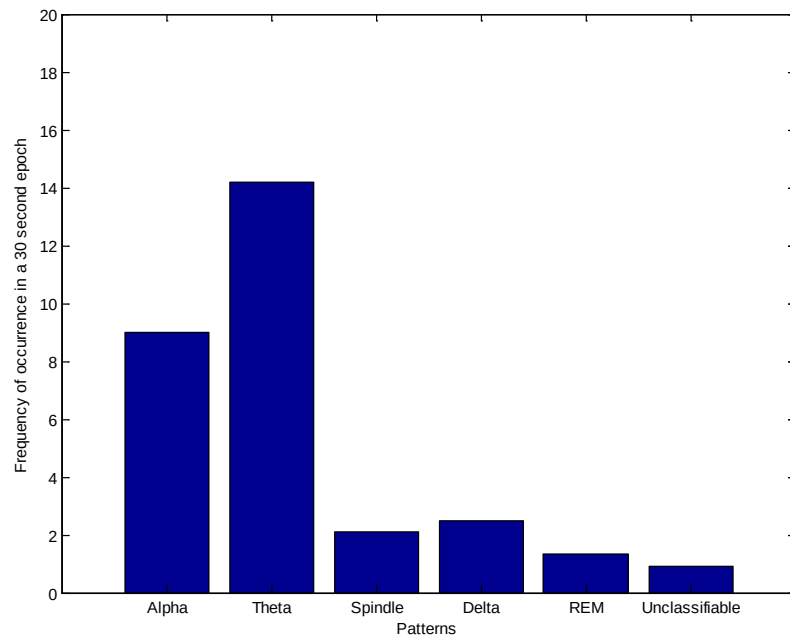


Figure 3.37 Reference profile for Stage 1.

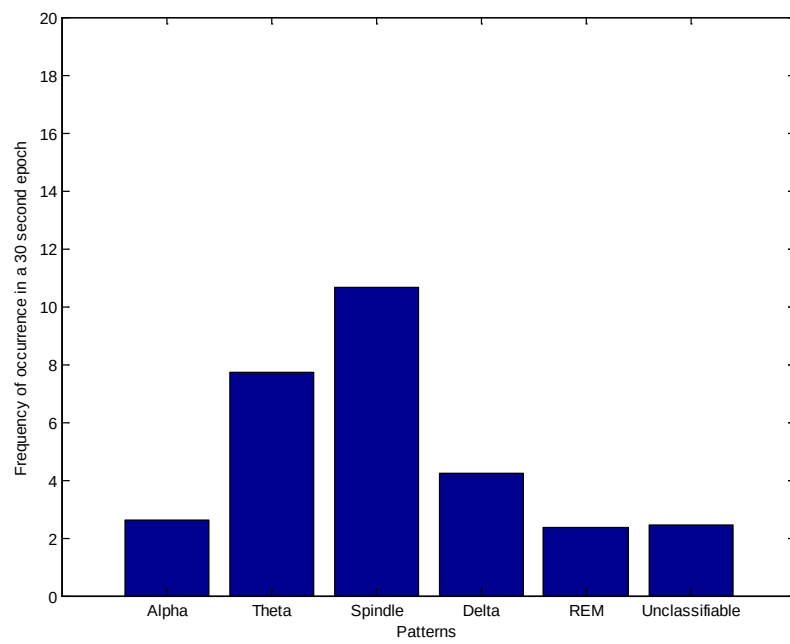


Figure 3.38 Reference profile for Stage 2.

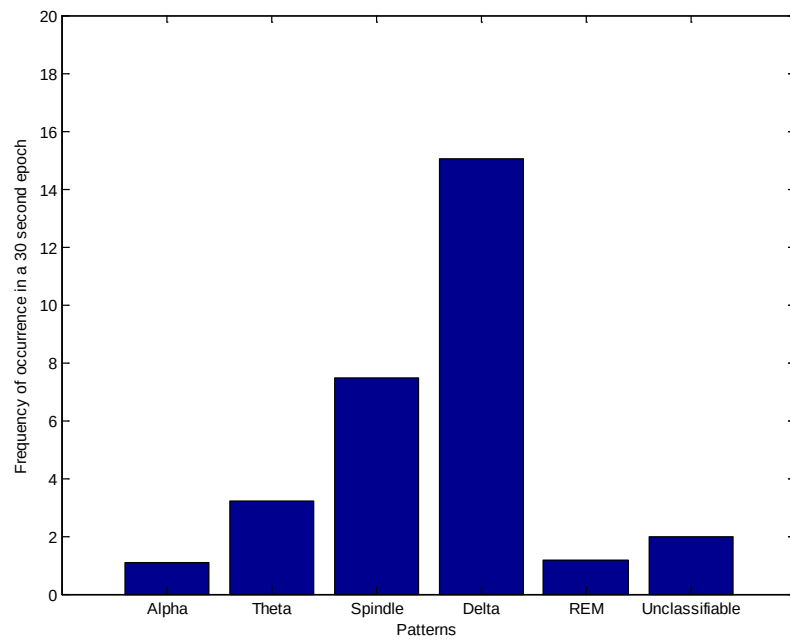


Figure 3.39 Reference profile for Stage 3.

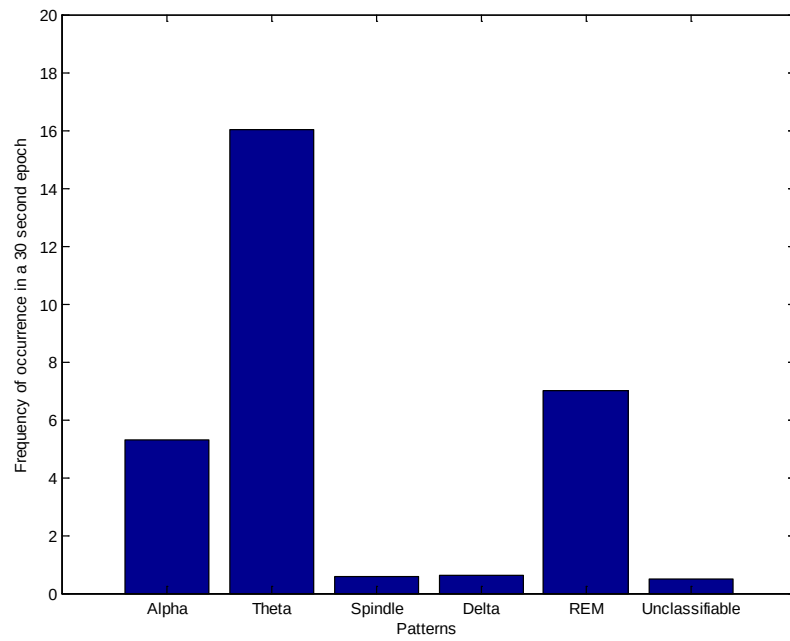


Figure 3.40 Reference profile for REM.

3.7 Detection of Sleep Transients

3.7.1 Spindle Detection

Spindles were detected on the basis of their frequency range. We inspected each cluster to see if the power spectrum had a peak between 12-15 Hz. If that was the case, the corresponding data were used to form the spindle cluster in the sleep stage classification. The power spectrum of the spindle cluster is shown in figures 3.41.

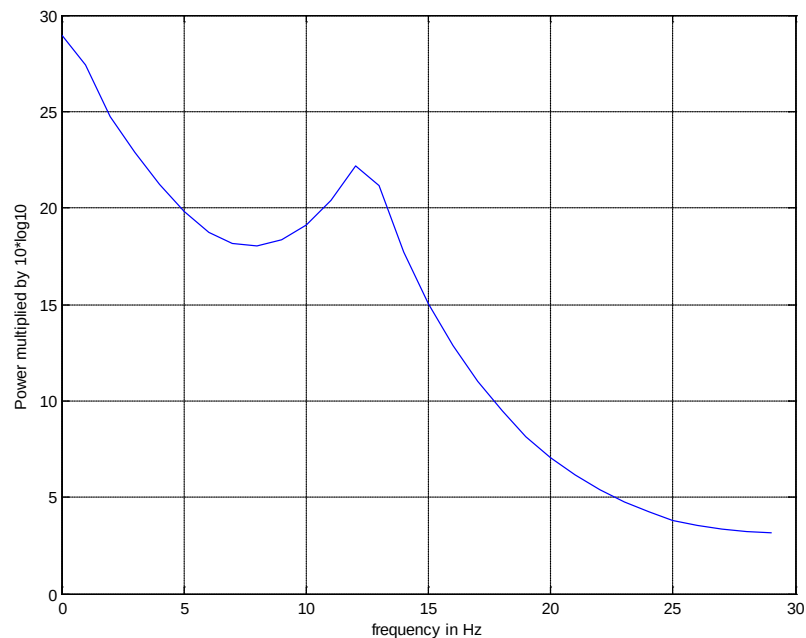


Figure 3.41 Power spectrum of the spindle cluster.

Our system detected the presence of spindles in all sleep stages but the maximum number of spindles was detected in stage 2. To minimize the risk of misclassifying a 30 s interval, we created a rule requiring that a 30 s epoch can only be classified as stage 2 if it has more than some threshold and this threshold value is used along with the profile classification. The histograms of spindle in each sleep stage are shown in Figures 3.42 and 3.43. The histograms suggest a threshold value of around 15. We then determined the

percentage correct classification for threshold values of 11, 12, 15 and 16 and the results are shown in Table 3.2. It was found that the threshold of 15 gave the largest correct classification percentage.

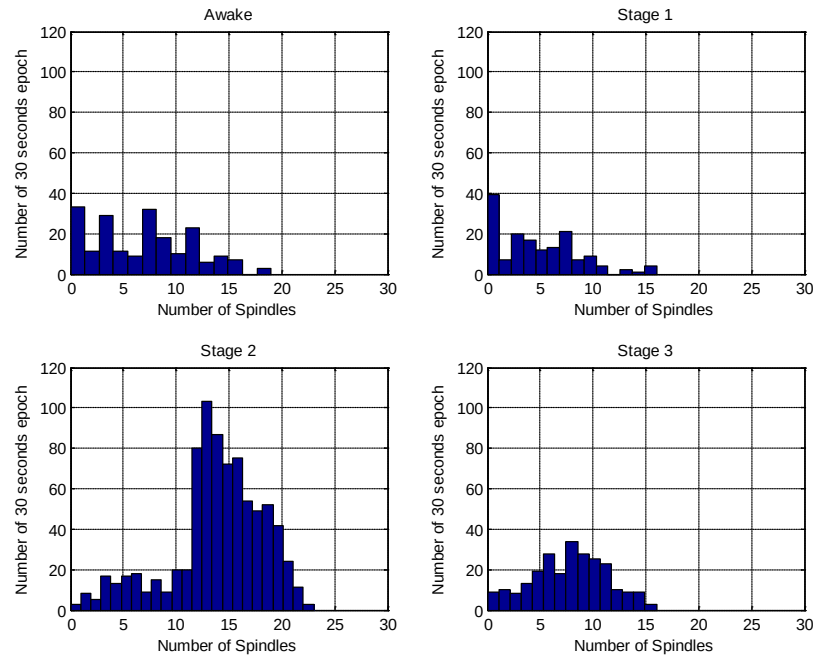


Figure 3.42 Histogram of the spindle distribution in each sleep stage.

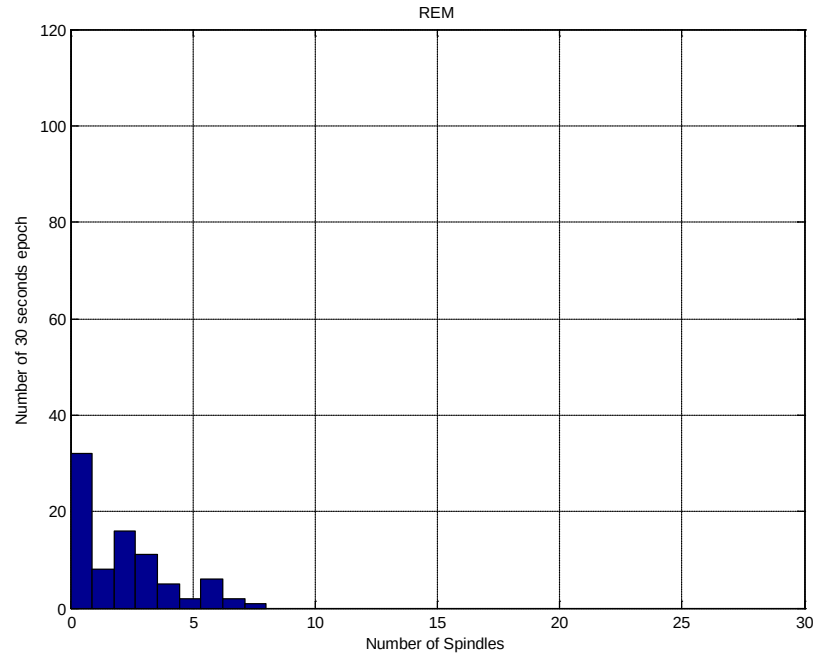


Figure 3.43 Histogram of the spindle distribution in REM sleep stage.

Table 3.2 Overall correct classification and spindle threshold

Threshold value for number of spindles	Overall correct classification percentage
11	76.63
12	77.02
15	78.58
16	77.26

3.7.2 REM Detection

The EOG (Electro-oculograms) channel was used for separating REM from stage 1. Thus, 30 second segments scored as REM or stage 1 were further analyzed using their corresponding EOG channel data. EOG data was preprocessed using the same steps as described in Section 3.2 and the standard deviation of each 1 second segment was

computed. It was observed that in stage 1 epochs the standard deviation rarely exceeded 50 for any 1 second segment but in case of the REM epoch, the standard deviation was more than 50 for a number of 1 second segments. This was tested on the 30 second epochs of REM and stage 1 from training subjects. In order to choose a threshold for the number of times a standard deviation of 50 or more occurred in both sleep stages we plotted the histogram of the standard deviation for each 1 second segment of REM and stage 1 separately. These are shown in Figures 3.44 and 3.45. The histograms were plotted for standard deviation greater than 30, 40, 50 and 60.

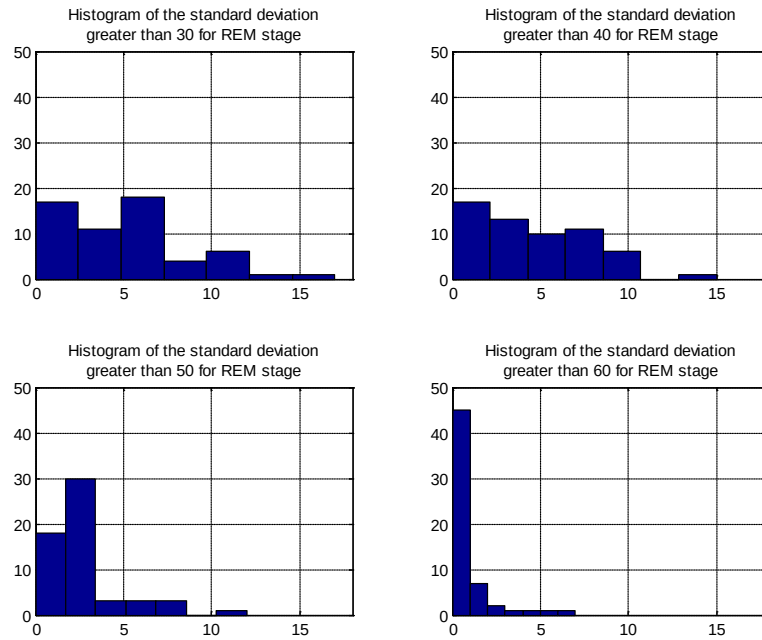


Figure 3.44 Histogram for the number of times the standard deviation exceeds 30, 40, 50 and 60 in REM stage.

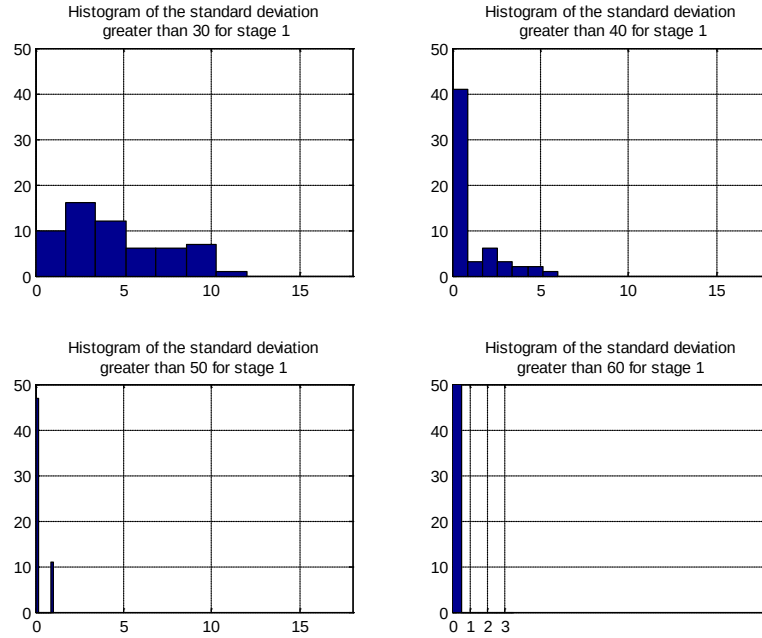


Figure 3.45 Histogram for the number of times the standard deviation exceeds 30, 40, 50 and 60 in stage 1.

From the distribution, we made a rule according to which a 30 second epoch having 2 or more instances of standard deviation greater than 50 was classified as REM.

3. 8 Data and Subjects

We first tested the method on simulated EEG data to ascertain that our method worked well on the EEG signal of known model and order. This helped us to develop our system efficiently and test its accuracy since we already had the knowledge about the EEG signal with which we had to experiment. After getting convincing results with simulated data, the method was applied on real EEG data with some modifications as required.

EEG data

Ten volunteers slept three nights in the laboratory and their EEGs were recorded for 1.5 hours for each night using scalp electrodes placed at 6 locations, 4 electrodes for eye and chin movement, and two reference electrodes. The recordings were done by Ziyang Li in Dr. Sheth's lab and sleep scoring was performed by a sleep scoring expert from Wedmed organization. The 10-20 principle for electrode placement (see Figure 3.46) was used for EEG recording with electrodes placed at P3, F3, O1, P4, O2, and F4.

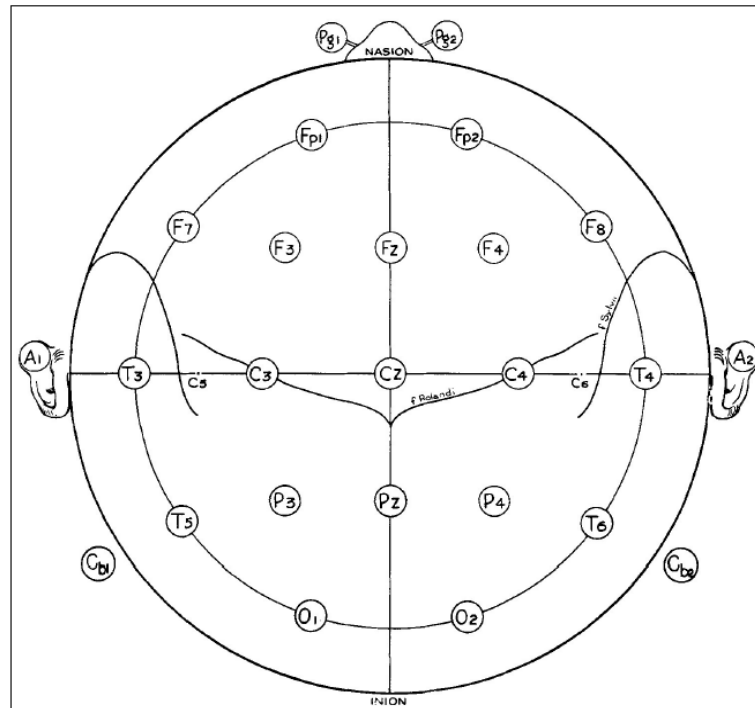


Figure 3.46 International 10-20 principle of electrode placement.

The 10–20 system is used to ensure standardized reproducibility so that a subject's studies could be compared over time and subjects could be compared to each other. The "10" and "20" refer to the fact that the actual distances between adjacent electrodes are either 10% or 20% of the total front–back or right–left distance of the skull. Each site has

a letter to identify the lobe and a number to identify the hemisphere location. Even numbers (2, 4, 6, and 8) refer to electrode positions on the right hemisphere, whereas odd numbers (1, 3, 5, and 7) refer to those on the left hemisphere. The recordings were sampled at 512 Hz. Visual sleep staging was done by a trained electroencephalographer applying the Rechtschaffen & Kales criteria, classifying each 30 sec epoch as Wake (0), Stage 1 (1), Stage 2 (2), Stage 3 (3) and rapid eye movement (5). The data available for each sleep stage for each subject is shown in Table 3.3.

Table 3.3 Sleep data available for each subject.

Subjects	Awake (Number of 30 sec Epochs)	Stage1 (Number of 30 sec Epochs)	Stage 2 (Number of 30 sec Epochs)	Stage 3 (Number of 30 sec Epochs)	REM (Number of 30 sec Epochs)
Subject 1	152	75	230	65	0
Subject 2	19	23	181	175	27
Subject 3	31	60	403	87	56
Subject 4	10	60	229	110	33
Subject 5	4	28	117	183	0
Subject 6	16	7	162	112	23
Subject 7	17	40	194	154	11
Subject 8	95	42	149	127	0
Subject 9	89	22	158	151	0
Subject 10	43	49	132	195	5

From the entire data available, we decided to use the data from subject 1, subject 2 and subject 3 as the training data and the data from the remaining 7 subjects as the testing data set. Out of the 7 testing data set, we used subject 4, subject 5, subject 8, subject 9 and subject 10 for initial testing and the other 2, i.e., subject 6 and subject 7 were kept for final testing of the classifier.

Chapter 4 Results

4.1 Training data

The results obtained on the training data have been represented in the form of confusion matrices. Each column of confusion matrix shows the instances in a predicted class and each row shows the instances in an actual class. Our results were compared with the visual scoring of an expert. In order to assess the effect of incorporating the transient detection method for REM and Spindle detection, we obtained separate confusion matrices using profile classification alone, combining the profile classification and the REM detection rule, combining the profile classification and the spindle detection rule and lastly by combining the profile classification with both the REM and spindle detection rules. The resulting matrices are presented in Tables 4.1-4.4, respectively. We also present the correct classification rate of each subject per night and overall correct classification rate for all the 3 training subjects combined in Table 4.5.

Table 4.1 Confusion matrix using only profile classification.

	Awake	Stage 1	Stage 2	Stage3/4	REM	Percent Correct
Awake	155	18	21	4	4	76.73
Stage 1	5	105	19	1	28	66.45
Stage 2	6	37	695	53	23	85.38
Stage 3/4	0	0	10	317	0	96.94
REM	0	24	3	2	54	65.06
Overall accuracy:						78.11

Table 4.2 Confusion matrix using profile classification and REM detection rule.

	Awake	Stage 1	Stage 2	Stage3/4	REM	Percent Correct
Awake	155	14	21	4	8	76.73
Stage 1	5	114	19	1	19	72.15
Stage 2	6	41	695	53	19	85.38
Stage 3/4	0	0	10	317	0	96.94
REM	0	14	3	2	64	77.10

Overall accuracy: 81.66

Table 4.3 Confusion matrix using profile classification and spindle detection rule.

	Awake	Stage 1	Stage 2	Stage3/4	REM	Percent Correct
Awake	152	16	26	4	4	75.24
Stage 1	5	104	20	1	28	65.82
Stage 2	2	21	734	36	21	90.17
Stage 3/4	0	0	11	316	0	96.63
REM	0	24	3	2	54	65.06

Overall accuracy: 78.58

Table 4.4 Confusion matrix using profile classification, spindle detection rule and REM detection rule.

	Awake	Stage 1	Stage 2	Stage3/4	REM	Percent Correct
Awake	152	14	26	4	6	75.24
Stage 1	5	113	20	1	19	71.51
Stage 2	2	19	734	36	23	90.17
Stage 3/4	0	0	11	316	0	96.63
REM	0	14	3	2	64	77.10

Overall accuracy: 82.13

Table 4.5 Percentage correct classification for training subjects per night per sleep stage.

Subject 1						
Nights	Awake (Percentage correct classification)	Stage 1 (Percentage correct classification)	Stage 2 (Percentage correct classification)	Stage 3/4 (Percentage correct classification)	REM (Percentage correct classification)	Overall Percentage correct per subject per night
Night 1	77.17	71.67	90.90	97.05	N/A	84.19
Night 2	72.94	78.57	92.45	96.83	N/A	85.19
Night 3	78.96	76.14	88.00	87.77	N/A	82.71
Overall percentage correct classification for subject 1						84.03
Subject 2						
Night 1	50.00	50.00	96.35	97.86	N/A	73.55
Night 2	70	73.22	89.61	95.00	74.26	80.41
Night 3	72.27	65.20	86.72	100	73.75	79.58
Overall percentage correct classification for subject 2						77.84
Subject 3						
Night 1	70	80.82	90.71	96	N/A	84.38
Night 2	80	83.35	85.00	98.50	80.10	85.39

Table 4.5 (continued)

Night 3	78.56	67.66	86.46	100	72.33	81.60
Overall percentage correct classification for subject 3						83.79

From the confusion matrices we observe that after the REM detection rule has been applied, the percentage of correctly classified stage 1 and REM epochs increases by almost 5% and 11%, respectively, without negatively affecting the classification accuracy of the other sleep stages and thus increasing the overall correct classification rate. Thus we concluded that the rule for REM detection works well.

We also observe that after the spindle rule is applied, the percentage correct classification of stage 2 increases but simultaneously, the classification accuracy of the awake stage decreases, with the overall classification rate remaining virtually unchanged. The cause for this may be that a large number of spindles in the awake stage are (incorrectly) detected. To investigate this further, the location of the one-second segments of the awake stage which were classified as spindles and as alpha activity, respectively, were graphed in the Euclidean space spanned by the first two AR coefficients and the variance (see Figure 4.1). Inspection of this graph reveals that the spindles are located almost on top of the alpha segments thus leading to false detection of spindles in the awake stage by the system. Since the alpha and the spindle segments are overlapping, we are unable to make any further separation between these categories using the features used here.

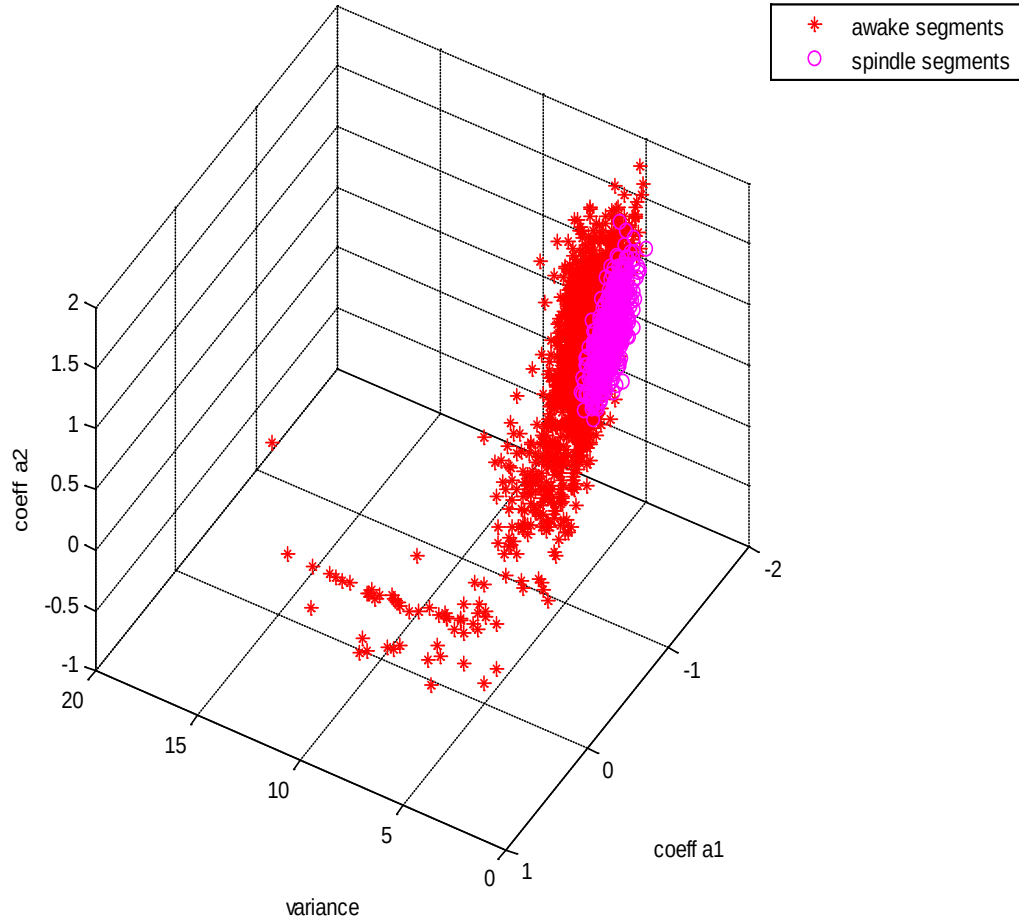


Figure 4.1 Graph of the location of the one second segments from awake stage classified as alpha (stars) and spindle activity (circles) in the Euclidean space spanned by the first two AR coefficients and the variance

Two hypnograms, one of night 2 for Subject 3 (most percentage of correct classification) and one of night 1 for Subject 2 (least percentage of correct classification), are shown in Figures 4.2 and 4.3, respectively. These figures suggest that many of the misclassifications occur around sleep stage transitions, often lasting only a few 30 s epochs.

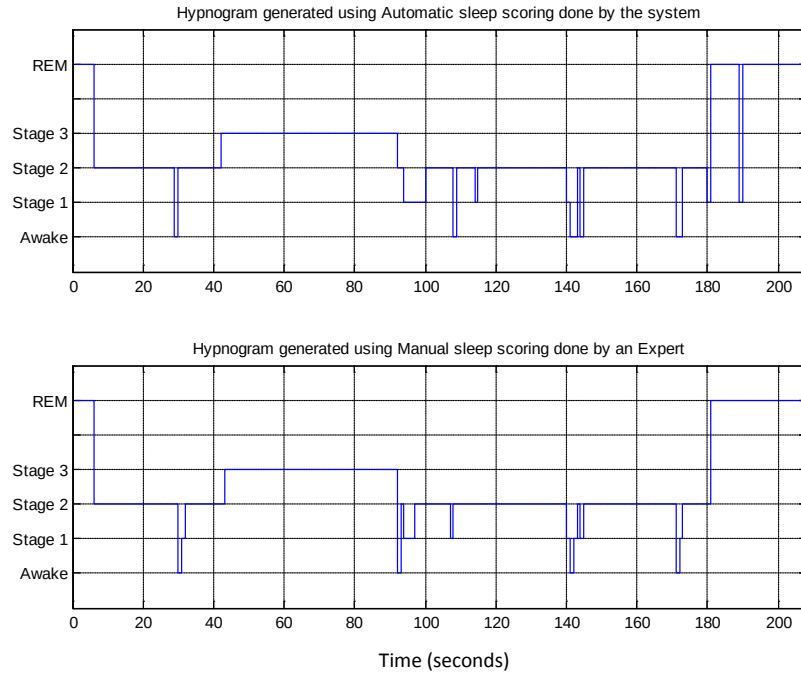


Figure 4.2 Sleep hypnogram of Night 2 for subject 3.

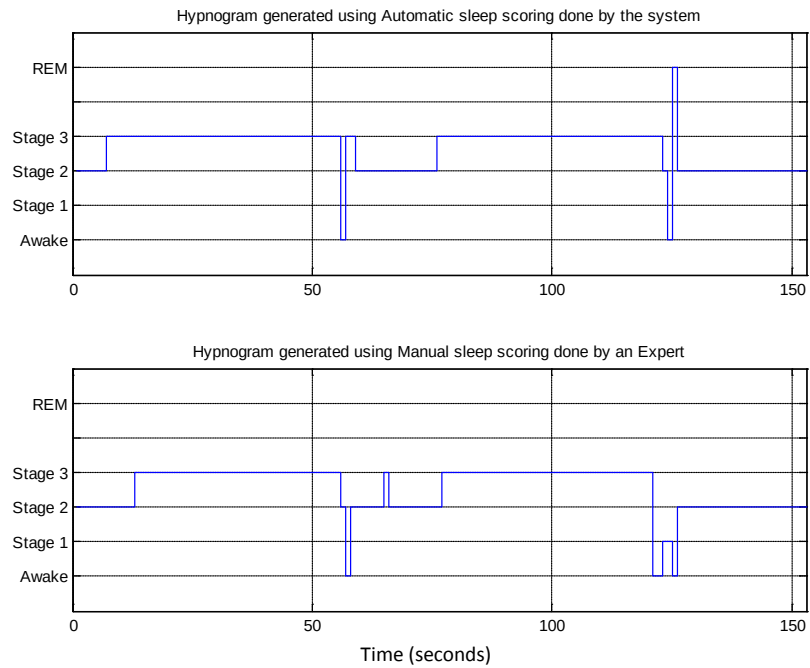


Figure 4.3 Sleep hypnogram of Night 1 for subject 2.

In order to perform a more comprehensive quantitative analysis, we obtained the histogram showing where the misclassifications occurred relative to the nearest sleep stage transition. This histogram is shown in Figure 4.2 and it is clear that most of the misclassifications occur within one to three epochs from the sleep stage.

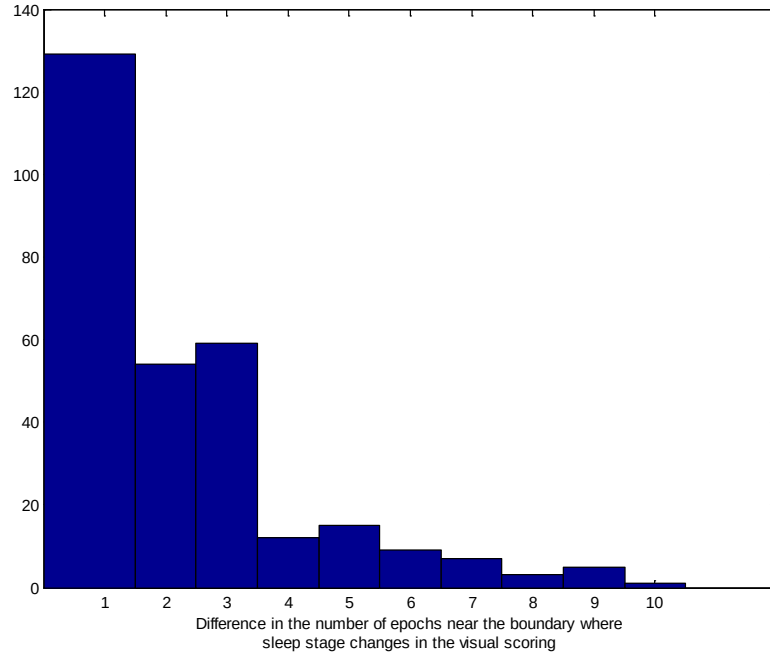


Figure 4.4 Histogram showing the epoch difference where stage transition occurs in automated scoring as compared to visual scoring (1 epoch = 30 second).

The results obtained by the automated system were further compared with that of visual scoring using sleep onset time and sleep quality measures.

Sleep onset time (or sleep latency) is the time between full wakefulness (i.e., going to bed) and falling asleep. The sleep onset times for visual scoring and automated scoring have been compared in Table 4.6. Sleep quality measures the percentage of deep sleep (stage 3 and REM) and the total number of awakenings during a night of sleep.

Sleep quality measures for the visual scoring and automated scoring are compared in Table 4.7.

Table 4.6 Sleep onset time comparison for the training subjects.

Subject 1			
Nights	Sleep onset time computed from Visual Scoring (seconds)	Sleep onset time computed from System Scoring (seconds)	Difference between the sleep onset time for visual and System Scoring (seconds)
Night 1	1410	1440	30
Night 2	0	0	0
Night 3	30	30	0
Subject 2			
Night 1	0	0	0
Night 2	660	510	150
Night 3	1740	1770	30
Subject 3			
Night 1	0	0	0
Night 2	900	990	90
Night 3	0	0	0

From Table 4.6 we observe that the average difference between the sleep onset time computed from visual scoring and automated scoring is 33 seconds and it varies in the range of 0-150 seconds.

Table 4.7 Sleep Quality Comparison for training subjects.

Subject 1							
Nights	Visual Scoring				System Scoring		
	Sleep duration (seconds)	Stage 3 (%)	REM (%)	No. of awakenings	Stage 3 (%)	REM (%)	No. of awakenings
Night 1	3900	26.15	0	12	29.40	0.58	15
Night 2	2790	25.80	0	13	27.20	2.77	10
Night 3	4230	4.96	0	11	8.35	0	11
Subject 2							
Night 1	4470	59.06	0	3	61.62	0	5
Night 2	4020	25.37	11.94	5	27.48	13.27	8
Night 3	4050	39.25	17.03	3	36.95	15.90	2

Table 4.7 (continued)

Subject 3							
Night 1	5250	10.28	0	5	9.89	0	5
Night 2	6120	24.01	16.17	1	22.93	18.29	3
Night 3	6480	9.25	4.62	14	10.09	6.69	16

From Table 4.7, we observe the following:

1. The average difference in the percentage of stage 3 computed from visual and automated scoring is 1.92%.
2. The average difference in the presence of REM computed from visual and automated scoring is 1.08%.
3. The average difference in the number of awakenings computed from visual and automated scoring is approximately equal to 1.77%.

4.2 First testing data set

The results obtained using the first testing data set are presented in the form of confusion matrix shown in table 4.8.

Table 4. 8 Confusion matrix for initial batch of testing data using profile classification, spindle detection rule and REM detection rule.

	Awake	Stage 1	Stage 2	Stage3/4	REM	Percent Correct
Awake	168	20	35	9	9	69.70
Stage 1	10	147	20	2	32	69.66
Stage 2	11	21	682	43	28	86.87
Stage 3/4	7	2	37	719	1	94.35
REM	0	5	2	2	29	76.31
Overall accuracy:						79.37

The overall percentage correct classification rate is 79.36, which is somewhat lower than the 83.79% obtained on the training data, but a binomial test [22] did not find a difference to be significant. We also assessed the correct classification rate of each subject per night and overall correct classification rate for all the initial batch of testing subjects and these results are shown in Table 4.9.

Table 4.9 Percentage correct classification for the first testing subjects per night per sleep stage.

Subject 4						
Nights	Awake (Percentage correct classification)	Stage 1 (Percentage correct classification)	Stage 2 (Percentage correct classification)	Stage 3/4 (Percentage correct classification)	REM (Percentage correct classification)	Overall Percentage correct per subject per night
Night 1	N/A	83.33	89.74	90.27	N/A	87.78
Night 2	N/A	77.77	85.47	89.23	N/A	84.15
Night 3	33.33	84.61	85.10	92.73	N/A	73.94
Overall percentage correct classification for subject 4						81.96
Subject 5						
Night 1	75.00	72.22	89.18	94.87	75.00	81.25
Night 2	75.00	70.58	84.28	100	81.81	82.33
Night 3	50.00	50.00	96.74	93.10	N/A	72.46
Overall percentage correct classification for subject 5						78.68
Subject 7						
Night 1	80.20	60.00	76.19	87.60	N/A	75.99
Night 2	66.23	80.00	79.42	88.71	N/A	78.59
Night 3	81.33	66.67	86.00	90.58	N/A	81.11
Overall percentage correct classification for subject 7						78.56
Subject 8						
Night 1	80	71.42	78.19	100	70	79.92
Night 2	80.00	66.66	81.20	95.00	N/A	78.21

Table 4.9 (continued)

Night 3	50.00	57.14	80.83	90.82	N/A	69.69
Overall percentage correct classification for subject 7						75.94
Subject 9						
Night 1	70	75.00	71.42	94.96	N/A	77.84
Night 2	100	66.67	94.00	97.28	N/A	89.48
Night 3	50	70.00	86.16	100	N/A	76.54
Overall percentage correct classification for subject 9						81.28

Table 4.9 shows that there are a few nights in which the correct percentage classification rate goes below 50% such as Night 3 of subject 4, Night 3 of subject 5 and Night 3 of subject 8. These are the nights in which there were only a small number of awake or stage 1 epochs, so we cannot really make a good comparison. For example in the case of Night 3 for subject 4, there was only three awake epochs and out of that our system detected only one awake epoch correctly, thus giving a classification rate of 33.33%. Similarly, in the case of night 3 for subject 5 there were only four stage 1 epochs and our system detected just two of them thus giving a classification rate of 50%. We have shown the comparison of sleep onset time and sleep quality using manual scoring and automated scoring as we did for the training data in tables 4.10 and 4.11, respectively.

Table 4.10 Sleep onset time comparison for the first testing subjects.

Subject 4			
Nights	Sleep onset time computed from Visual Scoring (seconds)	Sleep onset time computed from System Scoring (seconds)	Difference between the sleep onset time for visual and System Scoring (seconds)
Night 1	0	0	0
Night 2	0	0	0
Night 3	0	0	0
Subject 5			
Night 1	0	0	0
Night 2	0	0	0
Night 3	0	30	30
Subject 7			
Night 1	0	30	30
Night 2	0	0	0
Night 3	0	0	0
Subject 8			
Night 1	60	150	90
Night 2	0	30	30
Night 3	0	0	0
Subject 9			
Night 1	1110	1260	150
Night 2	0	30	30
Night 3	90	150	60

From Table 4.10, we observe that the average difference between the sleep onset time computed from visual and automated scoring is 46.66 seconds and it varies in the range of 0-150 seconds.

Table 4.11 Sleep quality comparison for the first testing subjects.

Subject 4							
Visual Scoring					System Scoring		
Nights	Sleep duration (seconds)	Stage 3 (%)	REM (%)	No. of awakenings	Stage 3 (%)	REM (%)	No. of awakenings
Night 1	3510	61.53	0	0	64.83	4.27	2
Night 2	3630	60.33	0	0	61.48	1.65	3
Night 3	2940	39.79	0	3	43.40	0	2
Subject 5							
Night 1	4050	23.70	8.14	2	25.95	6.74	4
Night 2	4260	26.05	15.49	3	27.10	13.33	7
Night 3	3960	21.96	0	2	24.30	2.27	3
Subject 7							
Night 1	2880	22.91	0	11	24.42	1.04	15
Night 2	3840	65.62	0	4	68.23	1.56	6
Night 3	3840	53.12	0	4	54.19	3.12	5
Subject 8							
Night 1	4410	44.21	2.72	4	44.93	4.08	2
Night 2	3840	56.25	0	0	57.15	1.56	2
Night 3	2940	59.18	0	0	60.18	2.34	3
Subject 9							
Night 1	2070	43.47	0	2	45.07	2.89	1
Night 2	3210	72.89	0	1	76.10	0	0
Night 3	4650	21.07	0	23	24.29	3.22	30

From Table 4.11 we observe the following:

1. The average difference in the percentage of stage 3 computed from visual and automated scoring is 1.97%.
2. The average difference in the presence of REM computed from visual and automated scoring is 1.92%.
3. The average difference in the number of awakenings computed from visual and automated scoring is approximately equal to 2.4%.

4.3 Final testing data

Since we get satisfactory results using the first batch of testing data, we test our algorithm on the remaining subjects: subject 6 and subject 10. The combined confusion matrix for these two subjects is shown in table 4.12.

Table 4.12 Confusion matrix for final batch of testing data using profile classification, spindle detection rule and REM detection rule.

	Awake	Stage 1	Stage 2	Stage3/4	REM	Percent Correct
Awake	22	2	8	0	0	68.75
Stage 1	3	30	3	0	8	68.18
Stage 2	26	5	208	3	8	85.95
Stage 3/4	1	7	13	176	1	88.88
REM	2	6	2	0	25	71.42

Overall accuracy: 76.63

We observe that the overall correct percentage classification for the final testing data is again slightly less than the training data but the binomial test did not find this to be significant. The overall percentage correct classification per subject per sleep stage per night is shown in table 4.13 so that we can assess the results obtained in more detail.

Table 4.13 Percentage correct classification for the final batch of testing subjects per night per sleep stage.

Subject 6						
Nights	Awake (Percentage correct classification)	Stage 1 (Percentage correct classification)	Stage 2 (Percentage correct classification)	Stage ¾ (Percentage correct classification)	REM (Percentage correct classification)	Overall Percentage correct per subject per night
Night 1	60	64.28	78.94	97.56	N/A	75.12
Night 2	66.66	66.66	88.70	81.25	N/A	75.81
Night 3	66.66	84.61	83.11	82.85	N/A	79.30
Overall percentage correct classification for subject 6						76.74
Subject 10						
Night 2	71.42	25	86.08	80.88	N/A	65.84
Night 3	55.55	66.66	77.08	86.36	63.63	71.41
Overall percentage correct classification for subject 10						68.62

In the table 4.13 we see that percentage correct classification of stage 1 in night 2 for subject 10 is only 25%. Here, there were only four stage epochs and our system detected just one of them thus we get an accuracy of 25%. Also, we had the sleep data for Night 1 of subject 10 but there was no manual scoring available for that night so we did not use that data. In the case of Night 3 for subject 10, the sleep data had only 6 channels, missing the eye movement channels. To compensate, we used the frontal channel as a substitute for the eye-movement channel, which may explain why the correct classification percentage of REM sleep is relatively smaller for this subject.

We observe the difference between visual and automated scoring across sleep onset time and sleep quality. The comparison is shown in tables 4.14 and 4.15 respectively.

Table 4.14 Sleep onset time comparison for the final batch of testing subjects.

Subject 6		
Nights	Sleep onset time computed from Visual Scoring (seconds)	Sleep onset time computed from System Scoring (seconds)
Night 1	0	0
Night 2	0	0
Night 3	0	0
Subject 10		
Night 2	60	30
Night 3	0	30

From Table 4.14 we observe that the average difference between the sleep onset time computed from visual and automated scoring is 12 seconds and it ranges between 0-30 seconds.

Table 4.15 Sleep quality comparison for the final batch of testing subjects.

Subject 6							
Visual scoring				System scoring			
Nights	Sleep duration (seconds)	Stage 3 (%)	REM (%)	No. of awakenings	Stage 3 (%)	REM (%)	No. of awakenings
Night 1	3240	51.85	0	3	52.05	2.77	6
Night 2	3870	31.00	0	3	35.00	0	1
Night 3	4860	35.80	0	3	39.20	4.25	4
Subject 10							
Night 2	4920	41.46	0	5	45.29	1.82	8
Night 3	4200	16.42	16.42	3	13.90	9.14	6

From Table 4.15 we observe the following:

1. The average difference in the percentage of stage 3 computed from visual and automated scoring is 2.79%.
2. The average difference in the presence of REM computed from visual and automated scoring is 3.22%.

3. The average difference in the number of awakenings computed from visual and automated scoring is approximately equal to 2.4%.

Chapter 5 Discussions and Conclusion

Overall we obtain around 80% correct sleep stage classification. The computer-generated hypnograms compared favorably with the visual hypnograms, with most of the misclassifications occurring during sleep stage transitions. The system continued to perform well even on data from subjects not used during the training phase.

We also observe that our attempt to make improvement in Jansen's method did not work completely as expected because of the poor performance of the spindle detection method. On the other hand, the REM detection rule did help to increase the correct detection of stage 1 and REM intervals.

We have used 5th order autoregressive coefficients and variance computed for each 1 second segment as the features. These features work quite well in capturing the general EEG characteristics, such as the slowing of the frequency seen during stage 2, and the alpha activity during the awake state. However, it does not capture transient activity such as sleep spindles, which may occupy only a fraction of a 1 s segment. We may have to use other methods, such as the ones proposed by Duman et al. [12] where the sleep spindle was detected by a combination of three algorithms which were based on time localization, amplitude threshold and frequency localization (mentioned in chapter 2). The system can be further improved if a method for the detection of K- complex is incorporated in the system, since a spindle is usually accompanied by a K-complex. One of the method to consider includes the detection of K-complex based on feature extraction and likelihood threshold presented by Devuyst et al. [11]. This method has been reviewed in chapter 2.

Furthermore, in the clustering process the clusters were selected on the basis of AR coefficients (and variance) which determine the power spectrum. However, very similar power spectra can be obtained from substantially different signals.

For separating REM from sleep stage 1, we used the EOG (eye movement channel). We have used the standard deviation for each 1-second segment as a feature. It did not give 100% correct classification. There may be a few stage 1 epochs having standard deviation of more than the threshold. Another problem in the REM detection result is the non-availability of sufficient REM data. Out of 29 nights recording from all the subjects, REM data was present only in eight nights and only for relatively short duration, thus we cannot completely rely on the REM detection results we have obtained. We did not use the EMG channel for detection of REM but using the EMG channel along with the EOG may give much better classification.

For the data analyzed in this study, there were a few subjects for which there was no awake and/or stage 1 data for some nights. In these cases we had only stage 2 and stage 3 data available which did make the results of stage 2 and stage 3 reliable but simultaneously the results of awake and stage 1 not very reliable.

The results we obtained using automated scoring are quite consistent except for a few sleep stages in some nights where the amount of data available was small especially for awake and sleep stage 1. For rest of the nights the percentage correct classification in comparison with manual scoring is around 80 -85 percent which is agreeable as even the inter rater disagreement between sleep scorers is also 10-15 percent [22]. Thus, we can infer that even though the spindle detection method did not work as expected the overall system works quite well. Therefore, we believe that if we design a method for more

accurate detection of spindles including the time domain analysis then accuracy of the system can definitely be increased to 85-90%.

References:

1. Rechtschaffen A, Kales A. "A Manual of Standardized Technology, Techniques and Scoring System for Sleep Stages of Human Subjects," Eds., Brain Inform. Service/Brain Res. Inst., University of California, Los Angeles, 1968.
2. Iber C, Ancoli I S, Chesson A, Quan, S F, "The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications," American Academy of Sleep Medicine, Westchester, IL, 2007.
3. <http://www.meditation-mp3.org/brainwave-entrainment>.
4. H. Park, J. Oh, D. Jeong, and K. Park, "Automated sleep stage scoring using hybrid rule- and case-based reasoning," Computers and Biomedical Research, vol. 33, pp 330-349, 2000.
5. Jansen B H, Hasman A, Lenten R. "Piecewise Analysis of EEGs Using AR-Modeling and Clustering," Computers and Biomedical Research 14, pp. 168- 178, 1981.
6. Itil T M, D M, Shapiro M F, and Kassebaum D. "Digital computer classifications of EEG sleep stages," Electroencephalography and Clinical Neurophysiology, vol. 27, pp. 76-83, 1969.
7. Johnson L, Lubin A, Naitoh P, Nute C, Austin M. "Spectral analysis of the EEG of dominant and non-dominant alpha subjects during waking and sleeping," Electroencephalography and Clinical Neurophysiology, vol. 26, pp. 361-370, 1969.
8. Larsen L E, Walter D O. "On automatic methods of sleep staging by EEG spectra," Electroencephalography and Clinical Neurophysiology, vol. 28, pp. 459-467, 1970.

9. Smith J, Negin M, Nevis A H. "Automatic Analysis of Sleep Electroencephalograms by Hybrid Computation," IEEE Trans. on Systems Science and Cybernetics, vol. SSC-5, pp. 278-284, 1969.
10. Martin W B, Johnson L C, Viglione S S, Naito P, Joseph R D, Moses J D. "Pattern Recognition of EEG-EOG as a technique for all-night sleep stage scoring," Electroencephalography and Clinical Neurophysiology, vol. 32, pp. 417-427, 1972.
11. Bremer G, Smith R. J, Karacan I. "Automatic Detection of the K-Complex in Sleep Electroencephalograms," Biomedical Engineering, IEEE Transactions on, Biomedical Engineering vol.BME-17, no.4, pp.314,323, Oct. 1970.
12. Jansen B H, "Artificial neural nets for K-complex detection," Engineering in Medicine and Biology Magazine, IEEE, vol.9, no.3, pp.50-52, Sept. 1990.
13. Devuyst S, Dutoit T, Stenuit P, Kerkhofs M. "Automatic K-complexes detection in sleep EEG recordings using likelihood thresholds," Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE , pp. 4658-4661, Aug. 2010.
14. Smith J R, Michael J C, Karacan I. "A multichannel hybrid system for rapid eye movement detection (REM detection)," Computers and Biomedical Research 4, no. 3, pp.275-290, 1971.
15. Gopal I S, Haddad G. G. "Automatic detection of eye movements in REM sleep using the electrooculograms," American Journal of Physiology - Regulatory, Integrative and Comparative Physiology Sept. 1981.

16. Duman F, Erdamar A, Osman E, Telatar Z, Yetkin S. "Efficient sleep spindle detection algorithm with decision tree, Expert Systems with Applications," vol. 36, Issue 6, pp. 9980-9985, Aug. 2009.
17. Gaillard J M, Tissot R. "Principles of Automatic Analysis of sleep records with a hybrid system," Computers and Biomedical Research, vol. 6, pp. 1-13, 1973.
18. Bodenstein G, Praetorius H M, "Feature Extraction from the Electroencephalogram by Adaptive Segmentation," Proceedings of IEEE 5, pp. 542-652, 1977.
19. Ebrahimi F, Mikaeili M, Estrada E, Nazeran H. "Automatic Sleep Stage Classification Based on EEG Signals by Using Neural Networks and Wavelet Packet Coefficients," 30th Annual International IEEE EMBS Conference, pp. 20-24, Aug. 2008.
20. Burg J P, "Maximum entropy spectral analysis," in Proceedings 37th Meeting Society Exploration Geophysics, Oklahoma City, OK, 1967.
21. Introduction to Wavelets and Wavelet Transform by C. Sidney Burrus, Ramesh A. Gopinath. First edition, Prentice Hall, 1998.
22. Bendat S. J, Piersol G. A, "Random Data: Analysis and Measurement Procedures, Wiley Publishers, 4th Edition 2010.

