AUTOMATED SLEEP SCORING SYSTEM USING LABVIEW

A Thesis

by

PARIKSHIT BAPUSAHEB DESHPANDE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2005

Major Subject: Electrical Engineering

AUTOMATED SLEEP SCORING SYSTEM USING LABVIEW

A Thesis

by

PARIKSHIT BAPUSAHEB DESHPANDE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Co-Chairs of Committee,	Charles Lessard
	Shankar Bhattacharyya
Committee Members,	Lihong Wang
	Aniruddha Datta
Head of Department,	Chanan Singh

December 2005

Major Subject: Electrical Engineering

ABSTRACT

Automated Sleep Scoring System Using Labview. (December 2005)

Parikshit Bapusaheb Deshpande, B.E., Govt. College of Engineering, Pune, India

Co-Chairs of Advisory Committee: Dr. Charles Lessard
Dr. Shankar Bhattacharyya

Sleep scoring involves classification of polysomnographic data into the various sleep stages as defined by Retschaffen and Kales. This process is time-consuming and laborious as it involves experts visually scoring the data. During recent years, there has been an increasing focus on automated sleep scoring systems and professional software programs are finding increased use. However, these systems are not relied on for scoring and are often used as a tool that facilitates easy visual scoring.

This thesis proposes a neural network based approach to automatic sleep scoring using LabVIEW. Effort has been made to give the sleep expert more control over key parameters such as the frequency bands, and thus come up with scores that are more in agreement with the individual scorer than being a rigid interpretation of the R&K rules. Though this thesis is limited to the development of an offline software program, given the data acquisition facilites in LabVIEW, a complete system from data acquisition to sleep hypnograms is a fair possibility.

To my parents…

ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Lessard for his constant support. In all the time I have spent with him, he encouraged hands-on projects, which involved real world data and solutions. It has cleared away my feeling that all real world solutions are complex and replaced it with the confidence that they are doable.

I would especially like to thank Dr. Bhattacharyya, who provided me numerous opportunities to be involved in interesting work. I also learned that all fields eventually come together – like when he plays his Sarod, he relies on feedback from his ear to play the correct note. It inspired me to marvel at the complexity behind apparently simple things in the world.

I would also like to thank Dr. Datta and Dr. Wang for taking an interest in my thesis and Juma, at the College Station Sleep Lab, for helping me with knowledge of real world sleep scoring systems.

Finally I would like to thank my parents for their constant encouragement.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

A. Introduction

Studies indicate that almost 80 million Americans suffer from some kind of sleep disorder [1],[2]. These disorders have psychological and/or physiological origins and may have an impact on the psychological and physiological well-being of a person. While some qualitative aspects (e.g. nightmares, insomnia due to worries, etc) of sleep can be narrated by the patient himself or noted by an observer, physiological aspects that provide a quantitative insight into the body's state during sleep are difficult to observe, and as such scientific methods need to be employed for their measurement.

Such methods involve recording of the Electro-encephalograms (EEG), Electro-cardiograms (ECG), Electro-myograms (EMG), and Electro-oculograms (EOG), together known as Polysomnograms. Once acquired, these parameters can be used to determine the state of sleep.

B. Sleep Scoring

Dement and Kleitman observed cyclic variations in EEG data that were closely related to the functional levels of the brain [3]. To analyze the large sets of data, they classified sleep into four stages, from Stage 1 to Stage 4. Later, Retschaffen and Kales

---

This thesis follows the style and format of *IEEE Transactions on Biomedical Engineering*.

standardized these into five stages – four Non-Rapid eye movement (NREM) stages from Stage 1 to Stage 4, and one Rapid eye movement (REM) stage, to increase compatibility between results in different laboratories [4]. These subsequently became the gold standard for quantitative analysis of sleep.

Sleep scoring involves classifying 30 second epochs of the polysomnographic data into one of the stages above and then plotting the whole night's trend as a sleep hynogram as shown in Fig. 1.

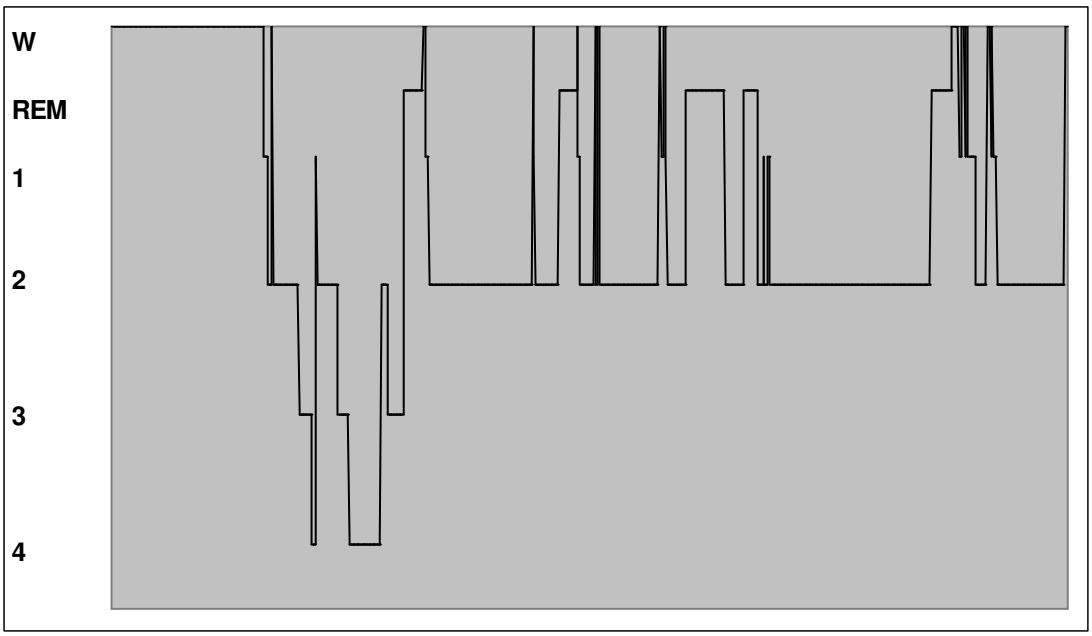


Fig. 1. All night Hypnogram.

C. Motivation

In the 2002 “Sleep In America” poll commissioned by the National Sleep Foundation, 58% reported having at least one of the four symptoms of insomnia at least few nights a week, and 35% had had those symptoms for the past year [5]. About one in eight adults aged 55 to 84 reported being diagnosed with at least one sleep disorder [6]. Physicians have identified around 70 sleep disorders, most of which can be treated to some extent once they are diagnosed [7].

The American Academy of Sleep Medicine lists just 16 accredited Sleep Centers in Texas, and the situation is similar throughout the country [8]. It is clear that the number of sleep centers is drastically less compared to the population of people having sleep disorders. People reporting to these labs have to wait for months to get their sleep analyzed.

Nowadays, the Polysomnographic systems sleep labs use also come with automated sleep scoring systems. Sleep scoring can also be outsourced to special labs for cost reduction [9]. However, these labs still have sleep experts who visually examine the polygraphic records and come up with the sleep scoring, since the results from the software are not in agreement with the sleep scorer. Agreement can be as low as 50% according to a sleep scorer at the College Station Physicians Center.

Hence the need arises for a low cost, flexible sleep scoring system that can be relied upon by the individual sleep scorer.

D. Expectations from System

The system developed should display all the signals together such that they can be easily distinguished from one another, and a scrollback mode should also be available. It should analyze the records and come up with the sleep stage for each of the 30 second epochs, and display a hypnogram of the same. User should be able to change key parameters – e.g. the frequencies of the clinical bands. Options to store the results digitally and print them must also be available.

E. Thesis Overview

This Thesis aims at developing an offline sleep scoring system in Labview. Anonymous data has been made available for this purpose. The data are whole night polysomnographic recordings of seven normal adults without any severe sleep disorders. Data has been sampled at a rate of 128 cycles/s, and the channels recorded are: ECG, chin EMG, LOC and ROC, C3 and O1.

The software has been developed in LabVIEW because of its inherent abilites in processing large amounts of data, its elaborate GUIs, and its data acquisition capabilites and compatible hardware (which will prove extremely useful in case the system has to be turned into an online one).

This system extracts certain features from the data after pre-processing it to remove artifacts, and uses a neural network to classify it into the sleep stages, and finally

displays a hypnogram of the same. It also allows for the results to be stored digitally, and printed if necessary. Thus it satisfies, in part, the expectations from such a system.

CHAPTER II

SLEEP SCORING

A. Electrode Placement and Sampling

The placement of electrodes for recording the polysomnograph determines the data that will be obtained. For EEG signals, the international standard is the 10-20 system of electrode placement [10] as shown in Fig. 2.

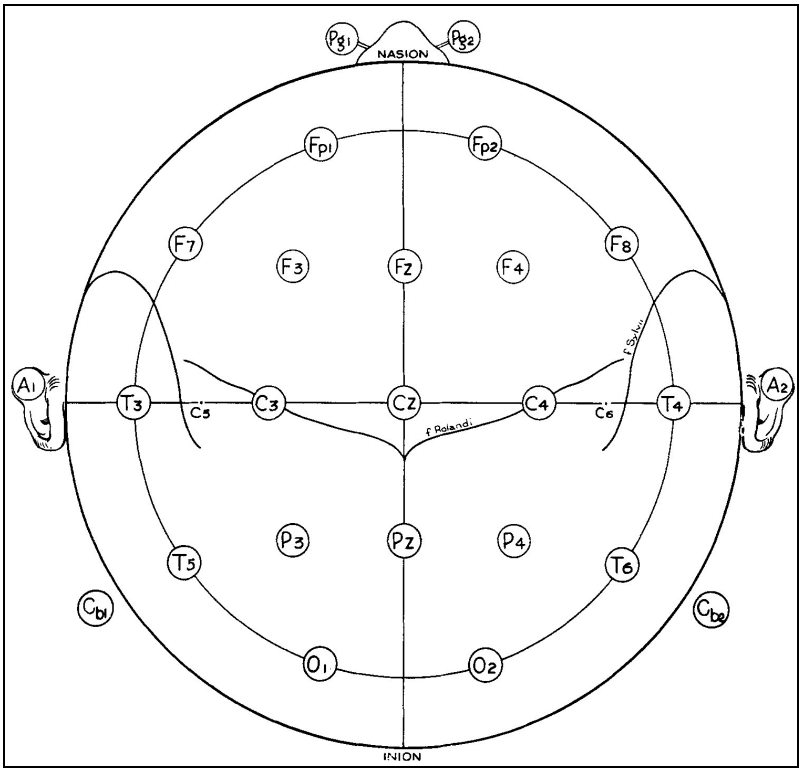


Fig. 2. The International 10-20 System of electrode placement.

The notations used for the brain lobes are F: Frontal, C: Central, T: Temporal, P: Parietal, O: Occipital. In this thesis, the channels provided in the EEG raw data are C3 and O1. Also, the other channels used are Right Oculogram (ROC), Left Oculogram (LOC), Chin Electromyogram (EMG), and a channel of ECG. This Thesis uses all channels except the ECG channel to obtain the sleep stage, as no significant information regarding the stage can be obtained from the ECG.

TABLE I

POLYSOMNOGRAPHY FREQUENCIES AND AMPLITUDES

Signal	Maximum Frequency of Interest (Hz)	Amplitude	Nyquist Sampling Rate (cycles/s)
EEG	~ 40	500 μV	80
EOG	5	5 mV	10
EMG	~ 2K	5 mV	4K
ECG	50 (monitoring)	1 mV	100

To analyze the data with the help of a computer, it has to be digitized. Hasan [11] suggests a sampling rate of 100 Hz, and Penzel and Conradt [12], in their review article, suggest the same sampling rate. TABLE I shows the magnitudes and highest frequencies of interest for the various channels, and also the sampling rates as per Nyquist's theory. The diagnostic frequency for ECG goes upto 250 Hz; however, for monitoring purposes (i.e. heart rate calculation) it is just 50 Hz.

Data for this thesis has been sampled at 128 cycles/s. This will result in undersampling for the EMG channel, but only the power for the channel is to be considered and hence it doesn't affect any calculations. Furthermore, for high resolution in EEG channels i.e. to detect small frequency shifts, a much higer sampling rate will be required, but this thesis does not involve such an analysis.

B. Clinical Bands

To classify the EEG recordings, the EEG frequencies are classified into bands as follows: Delta activity is from 0-4 Hz, Theta activity is from 4-8 Hz, Alpha activity is from 8 – 12 Hz, and Beta1 activity is from 13-22 Hz [13]. Few experiments also include one more band, called Beta2. In this thesis, a Beta2 band has been defined from 22-30 Hz. Also, as shown in Fig. 3, K-complexes are defined as sharp bursts in the EEG activity and Spindles are defined as high frequency activity in the range of 12-14 Hz, both of which last for more than 0.5 seconds [14].

K-complexes and spindles are difficult to be detected programmatically. Spindles usually follow k-complexes.

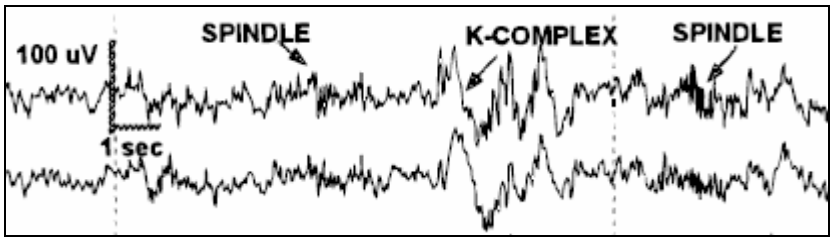


Fig. 3. Spindles and k-complex [15].

C. Sleep Stages and Scoring Criteria

The polysomnographic data is scored by splitting it into 30 second epochs and analyzing each of those epochs. Based on the clinical bands described above, the epochs are classified into stages Wake, Stage 1, Stage 2, Stage 3, Stage 4, and REM.

1) Stage Wakefulness: It is characterized by the presence of occipital alpha activity at 8-12 Hz as show in Fig. 4 and occurs when the subject is awake with eyes closed [14]. This might be accompanied by slow eye movements. In a normal subject, this stage is not dominant, but for subjects with disturbed sleep it might be significant [16].

2) Stage 1: As Fig. 5 shows, it is defined by a low voltage EEG, composed of mixed frequencies predominantly in the range of 2-7 Hz [4]. Also any epoch longer than 3 minutes with low amplitude and absence of spindles and REM is classified as Stage 1.

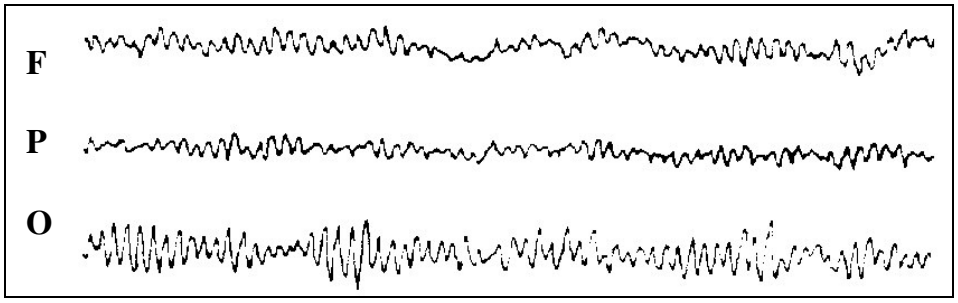


Fig. 4. Stage wakefulness [3].

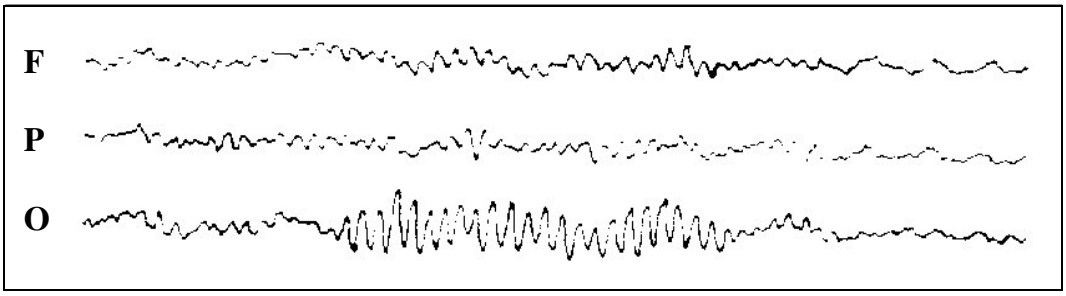
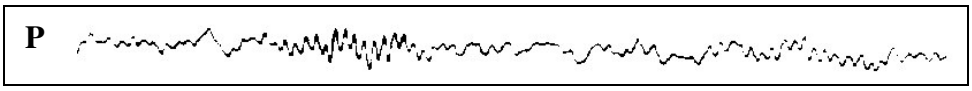


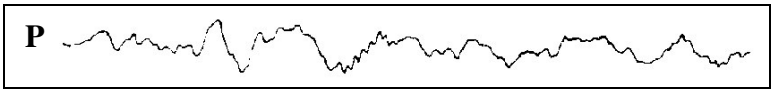
Fig. 5. Stage 1 [3].

Dement et al. [3] defined it as any pattern with absolute lack of spindle activity, and it was more of a transient stage from wakefulness to Stage 2. Thus, this stage does not have any characteristic features which might help easily identify it, and hence any relatively long period which doesn't satisfy the criteria for other stages is classified as Stage 1.

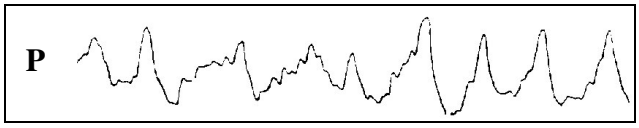
3) Stage 2: As Fig. 6(a) shows, its onset is characterized by the presence of 12-14 Hz spindles and K-complexes with a background of low voltage EEG activity [4]. However, the spindles and K-complexes have to last for atleast 0.5 s before the stage can be classified as Stage 2 [4]. Amzica & Steriade [17] found that K-complexes could last from 0.25 s to more than a second. In a later chapter, a method to detect these short duration complexes, given the relatively low sampling frequency, is described.



(a)



(b)



(c)

Fig. 6. Stages 2, 3, and 4. (a) Stage 2, (b) Stage 3, (c) Stage 4 [3].

4) Stage 3: This stage, as Fig. 6(b) depicts, is characterized by the presence of slow delta waves (0.5 - 2 Hz), exceeding 75 μV in amplitude, which occupy 20-50% of the epoch [4]. REM should also be absent. The frequency cutoffs for this band vary with scorers, and although R&K criteria define the cutoff at 2 Hz, scorers use upto 4 Hz as Delta [18]. This thesis also uses a default of 4 Hz, though it can be varied by the scorer.

5) Stage 4: This is similar to stage 3, except that the slow, high amplitude delta activity occupies more than 50% of the epoch, as shown in Fig. 6(c) [4]. This is a very important sleep stage, as it has been shown to affect emotional behavior and memory, and their regulation is an obligatory function of this stage [19], [20]. Stage 4 could be responsible for clearing unused memories [21]. Selective deprivation of stage four sleep also induces depression and hypochondriacal state [22]. The human growth hormone is also released during this stage.

6) Stage REM: This is defined by a low voltage, mixed frequency EEG activity along with the presence of Rapid-Eye-Movements (REM), and low amplitude EMG. It is similar to stage 1, except for the presence of characteristic saw-tooth waves in the EOG as shown in Fig. 7 [4]. There should also be absence of sleep spindles and K-complexes. The left and right oculograms have the exact opposite phase, since both the eyes always move in the same direction. REM sleep is also important for retention of new memories [21].

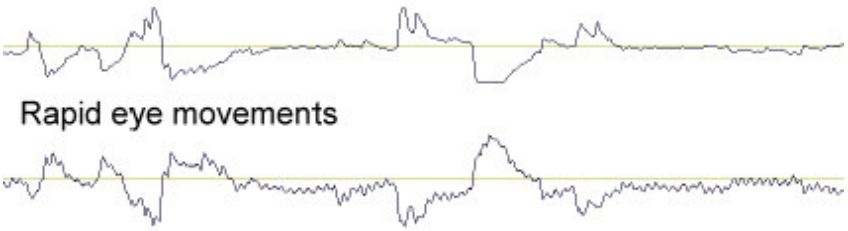


Fig. 7. EOGs during REM state [23].

D. Smoothing and Hypnograms

As mentioned earlier, on a basic level, sleep data are analyzed epoch by epoch according to the rules of scoring mentioned above. But these rules are not applied in isolation to the individual epoch, rather neighboring epochs are also a factor in scoring. R&K rules include a number of smoothing rules for removing apparent sudden shifts in stages. E.g. if there is a continuation of stage 4 and in between that trend are one or two epochs of stage 1, then those epochs should also be scored as stage 4. This is in keeping with the view that physiologically speaking, the stages of sleep would change gradually and not fluctuate rapidly.

With the advent of computerized scoring, epochs are now of a much smaller length – 1 or 2 seconds each. This is done to analyze features that might be missed if larger epochs are considered e.g. spindles and K-complexes. However, even for such epochs, the results are averaged over 30 seconds. Thus the 30 s epoch might be considered as the

main epoch, while the smaller epochs, which can be called sub-epochs, exist merely for calculation purposes and have no physiological significance as such.

After scoring and smoothing, the results for the whole night recording are plotted as a hypnogram. It is this hypnogram that gives a general idea about the subject's sleep.

CHAPTER III

PRESENT METHODOLOGIES

A. Data Preparation

Since the signals being recorded are of such low amplitudes, there is a lot of noise and interference involved. All automated scoring systems have an initial stage of artifact removal. These artifacts range from dc components, power line interference to those caused by movement, change in electrode impedance due to sweat, and mutual interference between channels [15]. Most of these artifacts can be removed by application of simple filters, while others, like mutual intereference, can be eliminated after calculating correlation coefficients. Epochs which have severe artifacts are not considered for scoring.

B. Spectral Analysis

Although there is increasing debate over acceptance of R&K rules as the gold standard, there is no other system which is being put to such extensive use [14]. As a result, most automated scoring systems try to mimic the visual scoring techniques as laid down by R&K. These being based on the frequencies occuring in the EEG, some sort of spectral analysis becomes integral to automated scoring systems. The power spectrum is obtained by implementation of FFT, and then powers or relative powers for each of the clinical bands are calculated.

Earlier on, the whole 30 s epoch used to be analyzed as a whole. Nowadays it is more common to further split the epochs into 1 s or 2 s epochs to address concerns about stationarity of 30 s data, and also to extract detailed features. Based on those features, each 30 s epoch is scored [13], [15]. This method is typically common for preliminary EEG analysis. After this stage however, the scoring methods diverge, and each system has a different approach to interpreting those results, as well analyzing other channels viz. EOG, and EMG. Also, methods differ in further analysis of EEG for detection of spindles and k-complexes.

C. Rule and Case Based Reasoning

Park et al. [15] designed an integrated Rule and Case based system to mimic the human approach to scoring. In a Rule based system, a rigid knowledge base is developed in the form of If-Then rules. Thus, a Rule based scoring system tackles problems by identifying priorities of the rules and then applying the appropriate rule. An example would be that of classifying an epoch as Stage 4. R&K rules state that any epoch with absence of REM and spindles in which slow waves occupy more than 50% of the epoch should be classified as Stage 4. This can easily be implemented as a series of rules – IF NO REM, AND, NO Spindles, AND, No K-complexes, AND Delta Power > 50%, THEN Stage 4.

Any number of rules can be added based on a similar logic. The disadvantage is that they do not adapt based on success and failures. Park et al. bypassed this problem by using a Case based reasoning system.

A case based reasoning system defines cases as a set of problem-solution cases that are extracted from the Signal Processing Unit, and the Rule Based Scoring Unit [15]. The Case based unit is accessed only if the Rule based unit cannot classify the epoch in a reliable manner. This unit is similar to the human experience – it applies the solutions of a similar previous case, and if the solution doesn't match, it revises the solution. However, this revision is limited in nature and a human expert might have to define a new case. The features extracted from the data are as follows.

1) EOG: The parameters calculated are - total power spectral frequency between 0.15 to 0.45 Hz, and 0.5 to 1.2 Hz, and the correlation coefficient between LOC and ROC.

2) EMG: Tone of chin EMG.

3) Wave Segment of EEG: The EEG epochs are divided into sub-epochs of 1 s each, and then after computing a windowed FFT, indexes are calculated to determine dominant clinical band of the sub-epoch. This sub division of epochs is done to include temporal information which can get lost in the 30 s epochs.

This system gives an average agreement rate in normal recordings of 87.5%. However the system suffered from problems in event detection at the Rule Based scoring level [15]. Also, to get a higher performance a much higher number of cases would be

required, which would essentially mean that the system has a large and complicated knowledge base, and that is clearly undesirable.

D. Neural Network

During recent years, neural networks have been finding increasing use to automatic sleep stage scoring [24]. Typically, neural networks are capable of mining through large amounts of data during the learning phase, and come up with weights which decide how much of a given parameter should be considered while making a decision.

Schaltenbrand et al. [13] developed an automatic scoring system using a multilayer feedforward neural network. First, feature extraction is performed on 2 s epochs, and then those features are averaged over 15 such consecutive epochs to obtain information about the 30 s epoch. Note that the 2 s epochs do not overlap. The features are obtained entirely from the power spectrum of the EEG, EOG, and EMG. The features extracted are described below.

1) EEG: Relative powers in the Delta (0-4Hz), Theta (4-8Hz), Alpha (8-13 Hz), Beta1 (13-22 Hz), and Beta2 (22-25Hz) bands; Total power of the EEG (0-35Hz); Ratio of powers Delta/Alpha and Alpha/Theta; Mean frequency of EEG spectral density; Dispersion of EEG spectral density.

2) EOG: Relative power in the Delta (0-4Hz) band; Total power of EOG spectral density; Mean frequency of the EOG spectral density; Dispersion of EOG spectral density.

3) EMG: Total power of the EMG spectral density; Mean frequency of EMG spectral density; Dispersion of EMG spectral density.

The above 17 parameters form the vector of parameters which is then fed to the neural network model. A 3 layer perceptron, with a 17 unit input layer, 10 unit hidden layer, and a 6 unit output layer is built. The 17 unit layer is input for the parameter vector, while the 6 unit layer is output of the 6 sleep stages (W, 1, 2, 3, 4, and REM). A hidden layer has been chosen keeping in view the complexity of the data.

After training and supervision, this system yeilds an 80.6% agreement between the automatic system and the consensus of readers. In this system, the mean frequencies in the three channels seem like unnecessary parameters as they have no actual significance in the R&K scoring process.

CHAPTER IV

PROPOSED APPROACH

A. Epoch Length Selection

To determine the epoch length to be used, a runs test was performed on the data to determine stationarity of various epoch lengths. This has to be done because spectral analysis need to be performed, and for that the data have to be stationary.

TABLE II

RUNS TEST SAMPLE RESULTS

Subject	Day	Signal	No. of Epochs	Epoch Length(s)	Runs	Expected Runs		Stationary
						Lower	Upper	
23	7	EMG	60	30	27	22	39	Y
23	7	LOC	60	30	19			N
23	7	ROC	60	30	21			N
23	7	C3	60	30	25			Y
23	7	O1	60	30	21			N

TABLE II (Continued)

Subject	Day	Signal	No. of Epochs	Epoch Length(s)	Runs	Expected Runs		Stationary
						Lower	Upper	
24	7	EKG	60	20	23	22	39	Y
24	7	EMG	60	20	7			N
24	7	LOC	60	20	5			N
24	7	ROC	60	20	9			N
24	7	C3	60	20	3			N
24	7	O1	60	20	7			N
24	7	EKG	60	10	22			Y
24	7	EMG	60	10	17			N
24	7	LOC	60	10	19			N
24	7	ROC	60	10	15			N
24	7	C3	60	10	25			Y
24	7	O1	60	10	21			N
23	7	EKG	60	2	22			Y
23	7	EMG	60	2	28			Y
23	7	LOC	60	2	23			Y
23	7	ROC	60	2	24			Y
23	7	C3	60	2	26			Y
23	7	O1	60	2	25			Y

Shown above in TABLE II are sample results from the runs test. From these results, it is clear that 2 s data are stationary. Also, the 2 s analysis is preferred because smaller epoch lengths mean better detection of details which is important in stage 2 for e.g. As earlier mentioned, staging has to be performed for the 30 s epochs; thus the 2 s parameters are averaged over 15 such consecutive epochs.

B. Thirty Second Level Epoch Processing

Analysis of data are performed at two levels of epochs – one at the 30 s level, and the other at the 2 s level. The functions performed at the 30 s level are described here. All other functions are to be assumed to be at the 2 s sub-epoch level.

1) Data Reading: This function acquires all the required data for the epoch from the data files. This is the function which necessisated the requirement of the 30 s level. It performs opening of the files, which is a slow processes. If this were to be done at the 2 s level, the program would spend most of its time reading the data from the files. Since the program is an offline one, it makes sense to read as much data as possible in a single iteration, without causing memory problems because all the data read is held in memory till the end of the iteration, when it is cleared.

2) Removing DC Offset: Usually, all the channels are AC coupled and hence there is no dc component present, but since this is not known for a fact, the data is processed to

remove dc offset. This is simply done by removing mean of data from data itself, and this task is also performed at the 30 s level to minimize calculations.

3) Movement Artifacts: This is a step towards pre-processing of data. The user can select thresholds for each of the channels, above which the epoch should be labeled as movement epoch. In an online system, these can have real units (μV or mV), but in an offline case where there is no indication of the gains, units have no meaning. Hence they have to be set in terms of absolute numbers, and may even vary from subject to subject. Data made available for this thesis had not indication about the gain and the actual physical values. However, they were found to have the same thresholds after visual examination and hence thresholds were set to correspond to obvious movement artifacts in the various channels.

The thresholds for this thesis are as follows: ECG – 2500, EMG – 800, LOC and ROC – 1000, and C3 and O1 – 350. (To make this scoring system accept any data, these thresholds need to be standardized.) These are detected by using waveform peak detection algorithm in Labview. A string of consecutive elements (atleast equal to width of peak) which exceed the threshold and then return to the value below it, is defined as a peak.

An epoch is classified as movement epoch if the absolute values of two or more channels exceed their thresholds or if the number of thresholds in a single channel exceeds six. The width of peak is defined as four (changeable by the user), as during visual inspection it was found to be a good compromise between detecting false peaks

and not detecting any real peaks. If epoch is scored as movement epoch, no further analysis is performed on it. This will also decrease the strain on the neural network, since the network won't have to classify extra data. Performing this function at the 30 s level saves a lot of calculation effort.

C. Two Second Level Epoch Processing

After performing the 30 s level analysis, data are further split into 2 s epochs through a For loop with fifteen iterations. This is the stage where the filtering and feature analysis takes place. Everything that goes into this level plays a part in determining the stage of the epoch. After feature extraction for each of the 2 s data, the features are averaged for the fifteen iterations. Functions performed here include filtering (which is the first stage), spectral analysis to extract certain features, and REM and K-complex analysis where some more features are derived. The two parameter extraction processes mentioned above take place simultaneously as they do not pass data to one another and need only the filtered data.

D. Filtering

Before deriving the parameter vector, the data needs to be filtered to eliminate noise and extract data only upto frequencies of interest. The frequencies of interest extend upto 30 Hz for EEG signals. For the rest of the signals here, the frequencies don't count as much, since the total power of those signals is considered. In any case, due to data being

sampled at 128 cycles/s, according to Nyquist's theory, the maximum frequency that can be represented is that of 64 Hz.

Based on these considerations, the requirements of the filter are defined as: Bandpass filter; Lower Cutoff 1 Hz, Upper Cutoff 50 Hz, Lower stopband 0.5 Hz, and Upper Stopband 55 Hz. By choosing the upper cutoff of 50 Hz, there is no need for another notch filter to eliminate the power line interference. The specifications for FIR and IIR are summarized below in TABLE III.

TABLE III

FILTER SPECIFICATIONS

Filter Type	Bandpass
Lower Frequency Cutoff (Hz)	1
Upper Frequency Cutoff (Hz)	50
Lower stopband (Hz)	0.5
Upper Stopband (Hz)	55
Passband Ripple (db)	1
Stopband Attenuation (db)	60

After deciding on these requirements, the type of filter needs to be chosen. The main options are FIR and IIR filters. For FIR filters, the transfer function is a polynomial in z^{-1} as shown below.

(1)

$$H(z) = \sum_{n=0}^{N} h[n] z^{-n}$$

For IIR filters, the transfer function is a real rational function in z^{-1}.

(2)

$$H(z) = \frac{p_0 + p_1 z^{-1} + ... + p_M z^{-M}}{d_0 + d_1 z^{-1} + ... + d_N z^{-N}}$$

The FIR filter can be designed to provide exact linear phase relationship, but for the same specifications, the FIR filter needs more coefficients than an IIR filter. Besides, the IIR filter also has sharp frequency cutoffs, and due to a lower order, the processing also takes lower computational effort and hence IIR is more efficient. Typically the FIR filter requires N_{FIR} computations per output sample (N_{FIR} being the order of the FIR filter), whereas the IIR filter requires $2^* N_{IIR} + 1$ computations per output sample (N_{IIR} being the order of the IIR filter). For the same magnitude response, the ratio N_{FIR}/N_{IIR} is in the order of tens or more [25]. The saving in computational effort is thus obvious.

FIR filters also have to be applied smoothing windows, due to truncation errors which results in Gibb's phenomenon. Choice of windows (Hanning, Hamming, etc) is also allowed, but the Blackman window is recommended.

In the case of this thesis, the phase response is not important, while the frequency cutoff is. Also, the computational effort has to be as minimal as possible for faster processing. So, although the user has a choice of deciding the type of filter to be implemented, the IIR filter is recommended, and set as default.

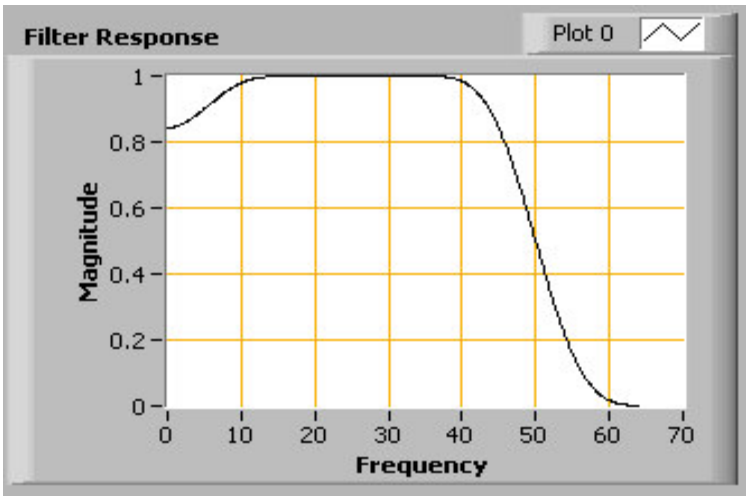


Fig. 8. FIR filter response.

Fig. 8 shows the magnitude response of the FIR filter with 25 coefficients, with the Blackman window applied. The rest of the specifications are the same as mentioned in TABLE III.

Fig. 9 below is the magnitude response of a 3rd order Butterworth Bandpass filter. Note that user can select between various IIR filters too, i.e. Butterworth, Chebyshev, etc. Since powers in individual bands are found out, the passband ripple has to be as low as possible. The Butterworth filters are better in this regard.

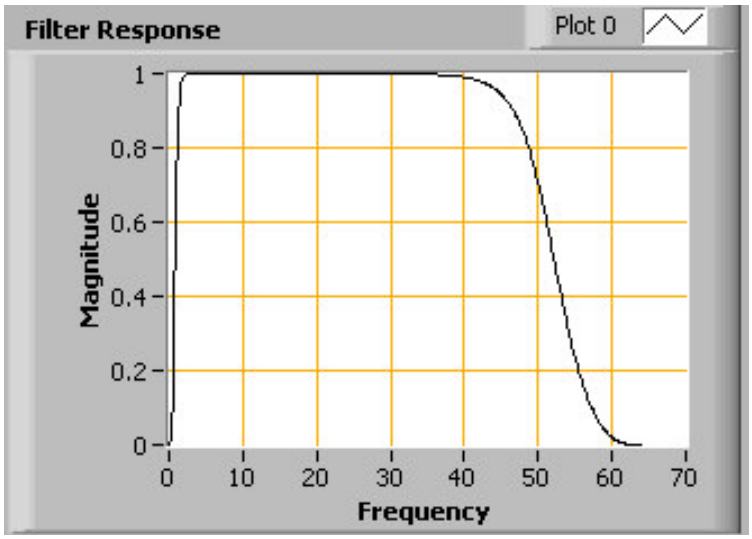


Fig. 9. IIR filter response.

Thus, from the figures, it is clear that even a 3rd order IIR filter has a better response than a 25th order FIR filter. The subVI for filtering takes in data from all the channels, at the 2 s level, and outputs the filtered data for the those channels. The functions that take data from this subVI are the ones for spectra and REM and K-complex detection.

E. Spectra

The classifications of sleep stages (see Chapter II) is based on the frequencies present in the EEG signal, the presence of sharp waves (k-complexes) in the EEG, and the presence of REM in the EOG. In fact, stages 3 and 4 are distinguished from each other solely on the basis of the power content of a single clinical band. Thus, it is only natural that the power content of the signal be studied in relation to the frequency content of the signals. This is achieved by calculating the Auto Power spectrum of the signals (on a 2 s level), which is defined as

(3)

$$Power\ Spectrum = \frac{FFT^{*}(Signal) * FFT(Signal)}{N^{2}}$$

For most of the time, the LOC and ROC will not be in perfect phase with each other. If the phase difference is large, then it implies that there is an error somewhere. Detecting whether the EOGs are in phase is important, because as mentioned later, the parameters for the EOG are averaged, and for that to be done, they have to in phase, else it might lead to false REM.

Detecting if the LOC and ROC are in phase with each other can be performed by finding phase of the Cross Spectrum, defined as

(4)

$$Cross\ Spectrum = \frac{FFT(1^{st}\ Signal) * FFT^{*}(2^{nd}\ Signal)}{N^{2}}$$

F. Features from the Spectra

Before moving onto the features that are extracted from the spectra, the various clinical bands are summarized below in TABLE IV for easy reference. Also note that feature extraction is performed at the 2 s epoch level.

TABLE IV

CLINICAL BANDS

Clinical Band	Minimum Frequency (Hz)	Maximum Frequency (Hz)
Delta	0	4
Theta	4	8
Alpha	8	12
Beta1	12	22
Beta2	22	32

To be able to extract the powers in the various clinical bands from the power spectra, numerical integration has to be performed. Labview performs integration in an incremental way i.e. it finds area under the curve from zero upto each of the frequencies[1]. Thus finding the powers in each of the clinical bands becomes easy. E.g. power in the Theta band can be found by extracting power upto 8 Hz, and subtracting the power of Delta band from it. Note that all the powers of clinical bands are normalized by the total power.

1) EEG: The features extracted from the EEG channels (C3 and O1 in this case) are the relative powers in the various clinical bands and the total power. These are the only features that can be extracted from the spectra that have an influence on the sleep stage (see Chapter II). Thus, the features are – Relative powers in Delta, Theta, Alpha, Beta1, and Beta2 bands, and the Total Power.

The features of each of the channel are considered separate features, as each channel (placed on a different lobe) provides different information.

2)EMG: The only power extracted from this channel is the total power. Large amplitude EMGs correspond to flexing and/or movements, and low amplitude EMGs correspond to low or no movements. The frequency distribution of these is irrelavant in scoring, and hence only total power counts.

[1] If the input spectrum has 10 samples, the integration will also have 10 points. The 1st point would correspond to power upto the 1st point, and so on. The last element would correspond to the total power.

3)EOG: In EOGs, apart from detecting REMs, it is also necessary to know if there are very slow eye movements, which correspond to sleep onset. Hence, EOG power in the delta band is also extracted [13]. Since REM also results in large amplitudes, the total power in the EOG can be used to distinguish REM and NREM states to some extent.

For most part, the EOGs are synchronous since both eyes move at the same time. Hence, to decrease the redundancy and thus number of features, the relative Delta powers and Total powers of the LOC and ROC are averaged.

The averaging done might create small problems, as the EOGs are not exactly in perfect phase all the time. As mentioned earlier, a large difference in phase might indicate an error[2]. To find the extent to which the two signals are out-of-phase, their cross spectrum is calculated and the phase of one relative to the another is found. This calculation will have half the number of the samples in the 2 s data (due to the single sided spectra), and it will also vary from positive to negative, in the same epoch too, since the eyes can move with respect to one another in any direction (this is limited to very small movements). Hence, the standard deviation of the relative phase is found for each 2 s epoch.

[2] As an afterthought, it is interesting to note that even slight twitching of an eye (which usually occurs in a single eye at a time, or more in one than the other), could perhaps cause large sawtooth waves in the EOGs, which might lead to detection of REM. It is not known if such twitching occurs only while dreaming, which would indicate that the sleep stage is indeed REM, or in any stage, which would mean false detection of REM. The detection of relative phase will indicate this possibilty, hence its use as a parameter.

Thus the parameters extracted from the LOC and ROC are relative Delta Power, Total Power, and Phase between the two, meaning the number of features from both the channels combined is three.

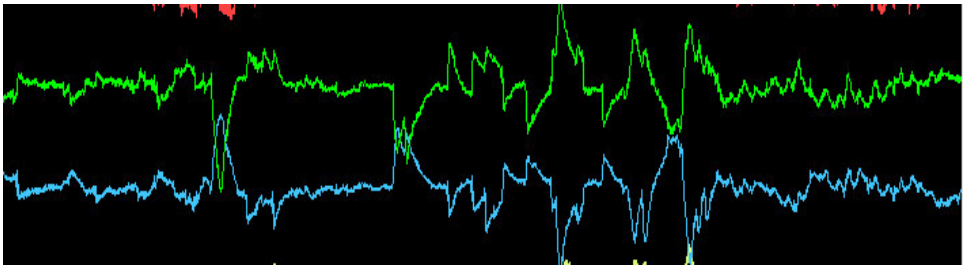
G. K-complex and REM Detection

K-complexes and Spindles are essential for detection of Stage 2, while the rapid eye movements indicate REM stage. Detection of REM is crucial because of the similarity between the EEG waveforms in Stage 1 and Stage REM. Improper detection of K-complexes (Stage 2) leads to a false classification. Schaltenbrand et al. [13] say their program had confusion between Stage 1 and Stage 2, and Stage 4 and Stage 2. The program by Park et al. [15] classified 11% of Stage 1 as REM, and 14% of Stage 1 as Stage 2. Thus detection of these two features become vital. The processing for these is done separately and simultaneously as they are not dependent on the spectral measurements.

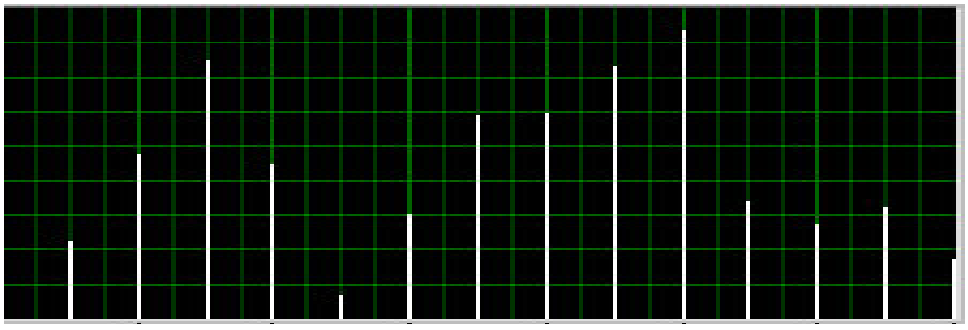
1) REM detection: This stage involves the typical sawtooth waves, usually of high amplitudes, in EOG channels. While this translates into increased power in the EOG spectrum which will also influence parameters in the spectral measurements, additional detection is performed here. The slope of the EOG channels is found out by way of derivatives (another method would be to run a peak detection algorigthm and find the slopes based on the number of samples between peaks and the consecutive valleys).

Since the slope is also calculated point-by-point, the standard deviation for the 2 second epoch is calculated. It is averaged over the 30 s epoch.

Fig. 11(a) shows the LOC and ROC for a 30 s epoch having REM. Fig. 11(b) shows the plot of 15 standard deviations for the same 30 s epoch. It can be seen that the sawtooth waves in EOGs correspond to high values on the plot of standard deviation, although not in direct proportion.

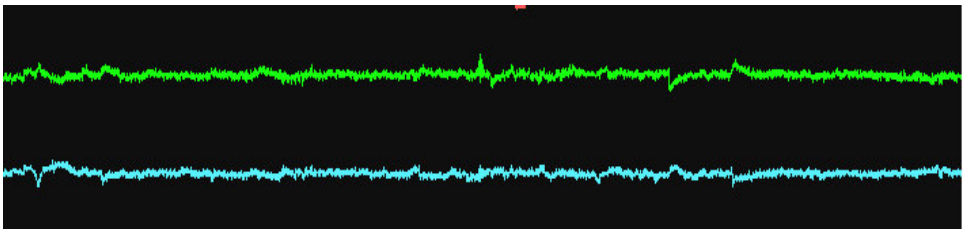


(a)

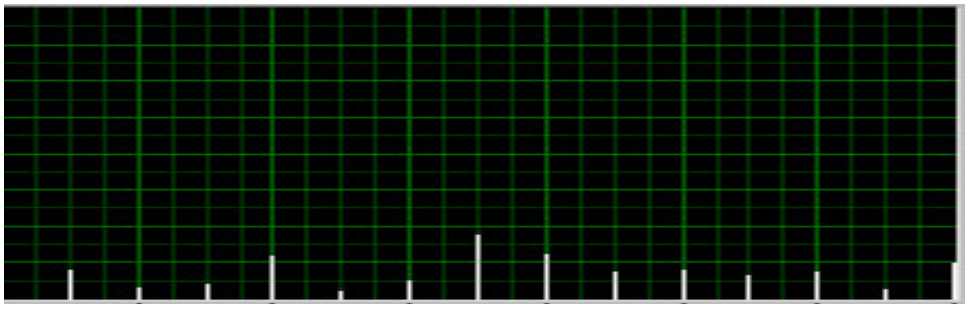


(b)

Fig. 10. REM and plot of the feature for it. (a) LOC and ROC during REM, (b) Plot of Standard Deviation of the slope for the same epoch.



(a)



(b)

Fig. 11. N-REM and plot of the feature for it. (a) LOC and ROC during non-REM, (b) Plot of the Standard deviation of the slope for the same.

Fig. 11(a) shows the LOC and ROC for a 30 s epoch of a non-REM period. Fig. 11(b) shows the plot of the corresponding standard deviations. Note that the EOG scales and Parameter scales Fig. 11 and Fig. 11 are the same.

Similar results were obtained for a large set of readings, and it was concluded that this feature is a good parameter for classification.

2) K-complex & Spindle detection: There are numerous difficulties in detecting a K-complex. It cannot be detected by an auto spectrum, even on the 2 s level. This is because though it has a high magnitude, it has a very short duration, and in the spectra its get overshadowed by the dominant features. Also, the frequency of the K-complex is around 4 Hz, which falls in the portions of the Delta and Theta bands. Hence it becomes impossible to determine if the frequency content comes from the K-complex or if it just part of the background EEG activity.

Quite a few programs do not extract any special features for the K-complex or spindles. E.g. Schaltendbrand et al. [13] perform no separate detection of waveforms, since they use smaller epochs of 2 s. Performing feature extraction at this level does give more information about K-complexes, but it is not significant enough to detect stage 2. As mentioned before, they have problems due to false positives resulting in Stage 2. Park et al. [15] used 1 s sub-epochs and generated indices to indicate the dominant frequency of the sub-epochs.

Neural networks have also been used solely for the purpose of K-complex detection. Bankman et al. [26] used a general model of a k-complex to determine features that could be extracted from it, and fed those features to a neural network. Their method yeilded 8.1% false positives at a sensitivity of 90% and 14.1% false positives at a sensitivity of 95%, which is sufficiently high. But this involves a lot of features for the k-complex itself, and besides there is no data available on the implementation of this system for sleep scoring.

Due to the short time duration of the K-complexes and spindles, the temporal information of these features becomes essential. Since the processing is based on spectral measurements, all the temporal information gets lost. However, with the advent of Short Time Fourier Transform (STFT) and Wavelet Transforms, both of which give joint Time-Domain information, knowledge of spectral as well as temporal information becomes possible. Tang and Ishii [27] used a wavelet transform for detection of K-complexes. (More information could not be found as only the abstract of this paper was available.)

Gorur et al. [28] used an STFT to extract features and then those features were processed using a neural network. They obtained a classification rate of 88.7% with the neural network. Once again, sleep scoring using this technique has not been discussed.

This thesis detects K-complexes based on the fact that they are sharp waves and their power is concentrated below 4 Hz (which falls in the range of Delta and Theta bands) [26]. They are difficult to detect in the presence of similar background activity, and the only feature that can lead to their presence is that they are localized in time. Thus detection by STFT is ideal.

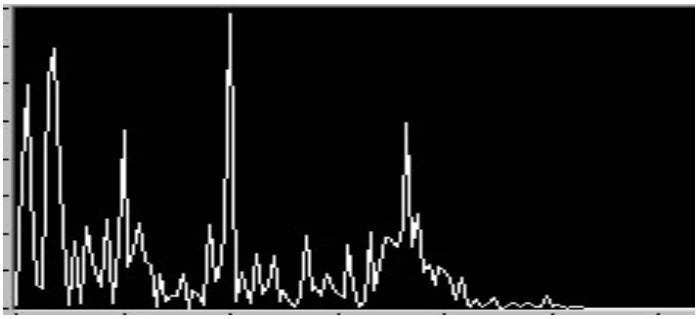


Fig. 12. Auto-Power spectrum of a 2 s epoch of C3.

Consider the Auto-Power spectrum of a 2 s epoch of channel C3 as shown in Fig. 12. This spectrum, which is ultimately based on the FFT, has the disadvantage that all temporal information is lost in the frequency domain (and conversely there is no frequency information in the time domain).

The Short Time Fourier Transform takes a sliding window (say length '*l*') and applies it to the data, takes the Fourier of that windowed data, and then slides the window by a certain time (say 'τ'), and repeats the procedure. The shifting and windowing function, and the resulting STFT can be represented as

(5)

$$g_{w,x}(t) = e^{jwt} w(t-\tau)$$

(6)

$$STFT\{f(w,\tau)\} = \int w(t-\tau) f(t) e^{-jwt} dt$$

The parameters needed for the STFT are the window length '*l*' and the time shift 'τ'. In the case of this thesis, the τ was chosen to be 16 samples i.e. for a sampling rate of 128 cycles/s, it corresponds to 12.5 msec, while the window length '*l*' is chosen to be 32 samples i.e. 25 msec. This provides for an overlap, which will smoothen the results. (Otherwise a spike or short spindle might fall in two different windows and as a result show up in none of them. Hence the overlap is needed to avoid such situations.)

Fig. 13 shows the STFT of the same segment that was used for the spectra in Fig. 12. It is a 3D graph, where the Z-axis represents the amplitude, one axis represents the frequency, and the time axis represents the shift, and not the actual time. E.g. in this case, there are 256 samples, and τ is 16. Therefore, the shift scale will have 256/16 i.e. 16 units on it. The frequency scale has the same range as that of the auto power spectrum.

Fig. 14 shows the projection of the STFT on the frequency axis. It is similar in nature to the FFT with the difference that the STFT performs such calculations for every increment in the time shift τ.

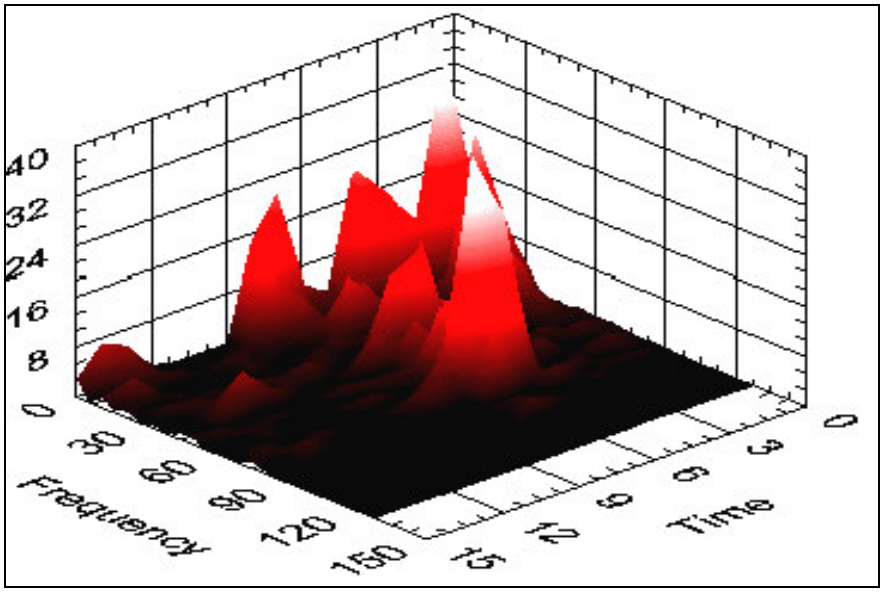


Fig. 13. STFT for a 2 s C3 segment.

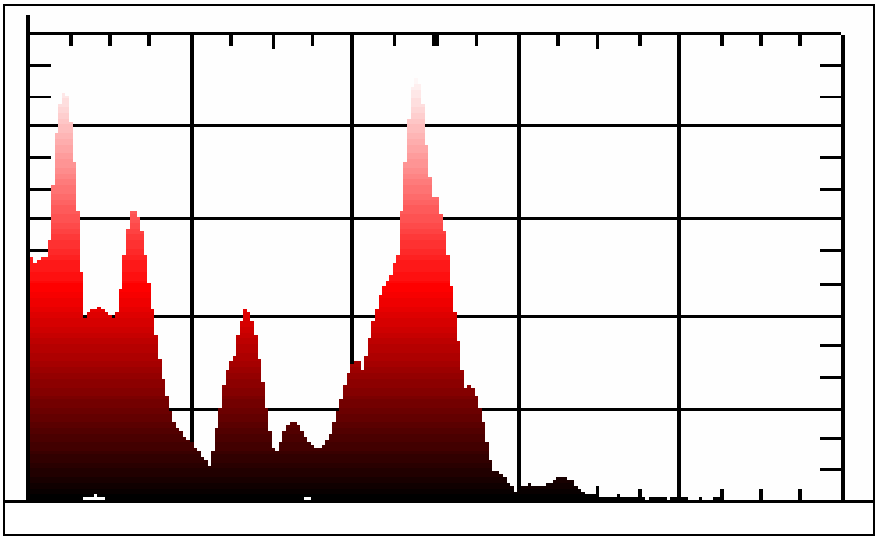
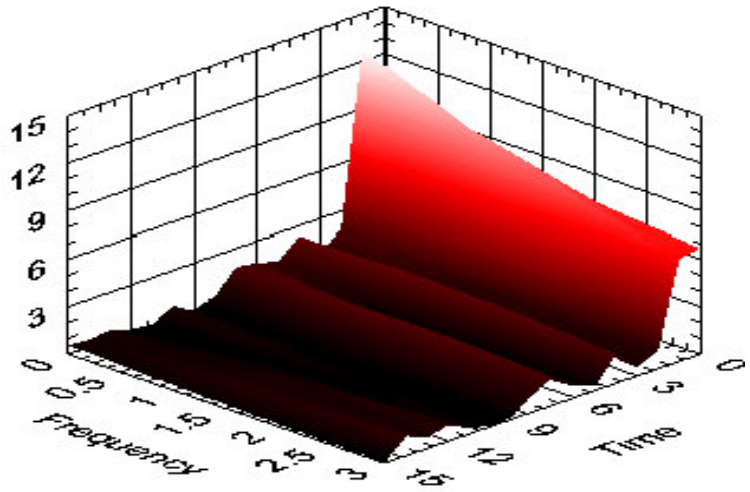


Fig. 14. Projection of STFT on the frequency axis.

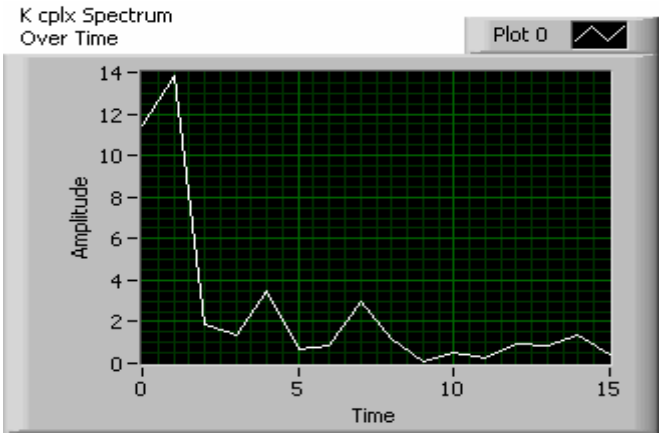
The corresponding portion for the K-complexes (2-5 Hz) from the STFT is extracted as shown in Fig. 15(a). That portion is a 2D array in time & frequency. To get the total power over the frequency range without losing the temporal information, numerical integration is performed for each of the time shifts. This operation can visualized as collapsing the extracted portion of the STFT onto the time shift axis, as shown in Fig. 15(b).

If there are short K-complexes (~ 0.5 s) in a given segment of 2 s, they will occupy only a part of the epoch, and if they are longer (>1 s), they will occupy a larger portion of the epoch. In the first case, the standard deviation of the collapsed portion will be high, while in the latter case the mean will be high. So both the parameters are determined and are part of the feature vector.

Now that all the features have been described, along with the justification for selecting them, they are summarized here in TABLE V.



(a)



(b)

Fig. 15. STFT extracted portion and its power. (a) Extracted portion from 2-5 Hz (b) Its power.

TABLE V

FEATURES FROM VARIOUS CHANNELS

Channel	Feature	Generic/Particular
ECG	-	-
EMG	Total Power	Generic
EOG	Delta Power	Slow eye movement
	Total Power	Generic
	Cross Phase Std. Deviation	LOC & ROC cross phase
	Average slope Std. Deviation	REM detection
EEG	Delta Power	Stage 3 & 4
	Theta Power	Generic
	Alpha Power	Stage Wake
	Beta1 Power	Generic
	Beta2 Power	Generic
	Total Power	Generic
	STFT portion K-complex Standard Deviation	K-complex, Stage 2
	STFT portion K-complex Mean	K-complex, Stage 2

CHAPTER V

RESULTS

A. Data Selection and Experiments

The total raw data available for this study was that of 7 patients over 3 days. Similar sleep scores for testing and validation were also available. However, the scores were not for the entire duration of sleep i.e. the epochs of the raw data and the scores did not match. For most files, the time stamps on the raw data files (which contained the actual channels) and the time stamps for the scores were different. As a result, most of the files could not be used for testing.

The files for Day1 (from raw data and scores) did match, alongwith two files from Day 7 and two from Day 2, and that data was used for training and validation of neural network. The neural network software used was Pathfinder. Parameters for subjects were generated by the Labview program. That data was combined into an Excel file, wherein the data was split into three segment – first part contained 60% of the data, and the latter two contained 30% data each. The first segment was for Training the Neural Network, the second for Testing, and the third for validation.

1) First Sample Experiment: In this case, the valid data from all three days (a total of ten files) was considered. It was split into sections of 6, 2, and 2 files each. There were 6 hidden nodes in this case. Also, the data was *not* modified to have equal number of samples for each of the sleep stages (in the training and testing portions). The number of

stage 2 epochs was much higher than the rest, and as a result the network trained well on this stage, but performed poorly on the others. The classification rate was 62.70%, with stage 2 having a rate of 99.2% and stage 0 (Wake) having a rate of 49.5% as Fig. 16 shows.

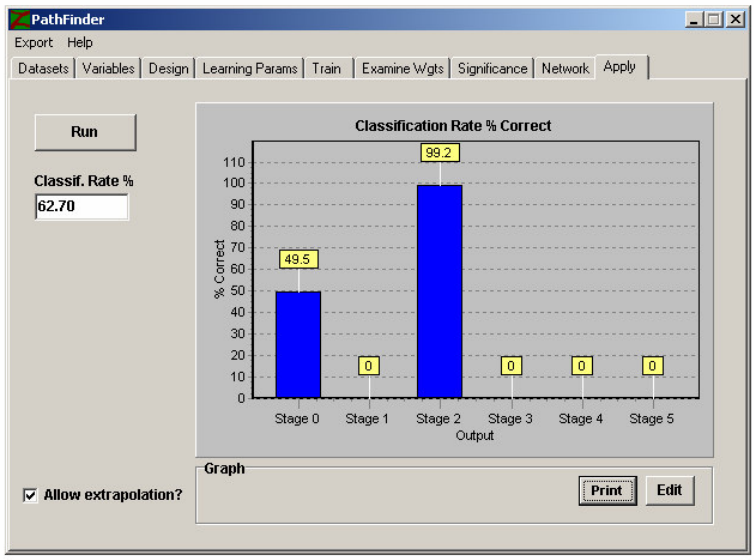


Fig. 16. Classification rates for 6 hidden node perceptron with all valid data.

2) 2nd Sample Experiment: In this case, only data from Day 1 was considered, since this data better matched the sleep score data. Thus, five subjects were considered – three for training, and one for testing and validation each. Only 2 hidden nodes were used, as

in the previous case there was a possibility that the network had learnt the results for Stage 2. Also, similar number of epochs were considered for each of the stages to avoid the possibility of overlearning on a particular stage. Although this data had worse classification rates than before, it classified data into three stages and not just two as in the previous case. The classification rate for stage 2 was still higher than the rest at 73.9%. The overall classification was 50.57% as shown in Fig. 17.

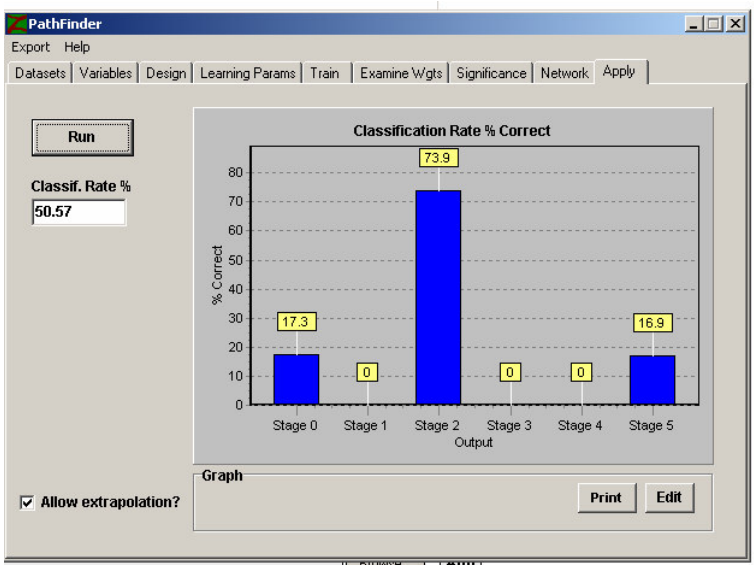


Fig. 17. Classification rates for 2 hidden node perceptron with day1 data.

3) 3rd Sample Experiment: This case was similar to the previous one, except that 3 hidden nodes were used. The classification rate was poorer than previous as Fig. 18 shows.

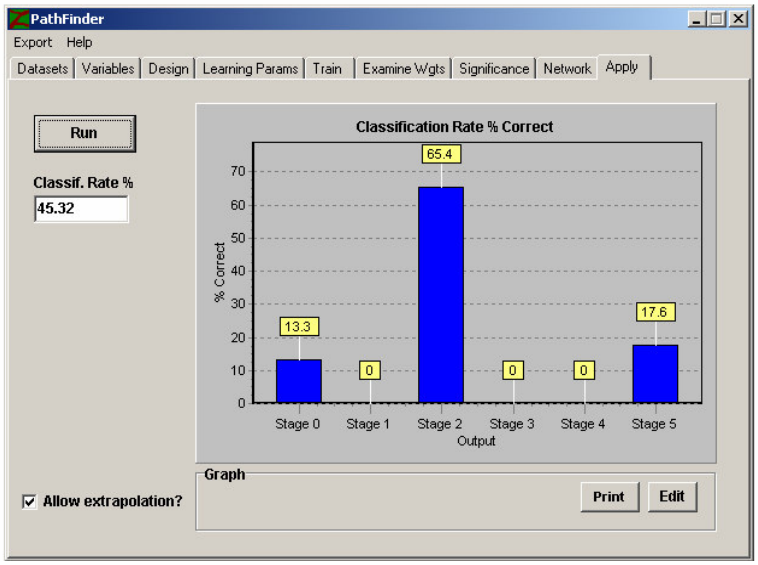


Fig. 18. Classification rates for 3 hidden node perceptron with day1 data.

B. Results

Similar experiements with small variations were performed on the data. However, the classification rates essentially remained the same. The rate for stage 2 remained high,

indicating that the k-complex detection program was giving suitable parameters. Overall, the performance of the system is poor, and as such cannot be used for classification. Whether this poor performance is due to mismatch of the raw data and scored data, or because of unsuitable parameters is not known.

CHAPTER VI

SUMMARY

The objective of this thesis was to develop a reliable sleep scoring system in Labview. The need for such a system was discussed. After defining the characteristics found in polysomnographs, the sleep scoring criteria were defined. Previous attempts using two different approaches were studied.

Discussion on data preparation was followed by the features to be extracted along with the justification for their extraction. The results from the neural network based on those parameters were shown.

Based on the results obtained, it is clear that at its present state the system is unable to reliably classify data, and thus it is inadequate to perform its objective. Training of neural network on reliable data will yield better results.

REFERENCES

[1] R. C. Rosen, R. Zozula, E. G. Jahn, and J. L. Carson, "Low rates of recognition of sleep disorders in primary care: Comparison of a community-based versus clinical academic setting," *Sleep Medicine*, vol. 2, pp 47-55, 2001.

[2] W. C. Dement and M. M. Mitler, "Waking up to the importance of sleep disorders," *Journal of American Medical Association*, vol. 269, pp 47- 55, 1993.

[3] W. C. Dement and N. Kleitman, "Cyclic variations in EEG during sleep and their relation to eye movements, body motility and dreaming," *Electroencephalography and Clinical Neurophysiology*, vol. 9, pp 673-690, 1957.

[4] *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*, A. Rechtschaffen and A. Kales, Eds., Brain Inform. Service/Brain Res. Inst., Univ. California, Los Angeles, 1968.

[5] "Sleep in America" Poll, commissioned by the National Sleep Foundation, 2002, accessed from www.sleepfoundation.org on January 25, 2005.

[6] "Sleep in America" Poll, commissioned by the National Sleep Foundation, 2003, accessed from www.sleepfoundation.org on January 25, 2005.

[7] National Institute of Neurological Disorders and Stroke, "Brain Basics: Understanding Sleep", accessed from http://www.ninds.nih.gov/disorders/brain_basics/understanding_sleep_brain_basic.htm on January 23, 2005.

[8] American Academy of Sleep Medicine, "Locate a Sleep Center Near You," accessed from http://www.aasmnet.org/statetemplate.asp?state=TX on January 23, 2005.

[9] T. Whalen, "Organizing your outsourcing," accessed from http://sleep-medicine.advanceweb.com/common/editorial/editorial.aspx?CC=21059 on January 25, 2005.

[10] H. H. Jasper, "Reports of the committee on methods of clinical examination in electroencephalography," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp 370-375, 1958.

[11] J. Hasan, "Past and future of computer-assisted sleep analysis and drowsiness assessment," *Journal of Clinical Neurophysiology*, vol. 13, pp 295-313, 1996.

[12] T. Penzel and R. Conradt, "Computer based sleep recording and analysis," *Sleep Medicine Reviews*, vol. 4, pp 131-148, 2000.

[13] N. Schaltenbrand, R. Lengelle, and J. Macher, "Neural network model: Application to automatic analysis of human sleep," *Computers and Biomedical Research*, vol. 26, pp 157-171, 1993.

[14] S. Himanen and J. Hasan, "Limitations of Rechtschaffen and Kales," *Sleep Medicine Reviews*, vol. 4, pp 149-167, 2000.

[15] 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.

[16] J. Verbraecken, W. De Backer, W. De Cock, W. Wittesaele, and P. Heyning, "The usefulness of an alternative EEG scoring system in obstructive sleep apnea," *Acta Otorhinolaryngol Belg*, vol. 48, pp 27-35, 1994.

[17] F. Amzica and M. Steriade, "The K-complex: Its slow (<1-Hz) rhythmicity and relation to delta waves," *Neurology*, vol. 49, pp 952-959, 1997.

[18] IFSECN, "A glossary of terms most commonly used by clinical electroencephalographers," *Electroencephalography and Clinical Neurophysiology*, vol. 37, pp. 538-548, 1974.

[19] C. Zi-Jian, "An integrative analysis to sleep functions," *Behavioural Brain Research*, vol. 69, pp. 187-194, 1995.

[20] C. Zi-Jian, "The functions of sleep: Further analysis," *Physiology and Behavior*, vol. 50, pp. 53-60, 1991.

[21] M. V. Ambrosini, A. G. Sadile, G. Carnevale, M. Mattiaccio, and A. Giuditta, "The sequential hypothesis on sleep function I. Evidence that the structure of sleep depends on the nature of the previous waking experience," *Physiology and Behavior*, vol. 43, pp. 325-337, 1988.

[22] H. W. Agnew, W. B. Webb, and R. L. Williams, "Comparison of stage four and I-REM sleep deprivation," *Perceptual and Motor Skills*, vol. 24, pp 851-858, 1967.

[23] Sleepstudy.org, "Scoring sleep stages from the EEG," accessed from http://www.sleepstudy.org/review/sleepEEG.htm#REM on January 23, 2005.

[24] C. Robert, C. Guilpin, and A. Limoge, "Review of neural network applications in sleep research," *Journal of Neuroscience Methods*, vol. 79, pp. 187-193, 1998.

[25] L. R. Rabiner and R. W. Schafer, *Theory and Application of Digital Signal Processing*, Englewood Cliffs NJ: Prentice Hall, 1975.

[26] I. N. Bankman, V. G. Sigillito, R. A. Wise, and P. Smith, "Feature based detection of the K-complex wave in the human electroencephalogram using neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 39, pp. 1305-1310, 1992.

[27] Z. Tang and N. Ishii, "Detection of the K-complex using a new method of recognizing waveform based on the discrete wavelet transform," *IEICE Transactions on Information and Systems*," vol. 78, pp. 77-85, 1995.

[28] D. Gorur, U. Halici, H. Aydin, G. Ongun, F. Ozgen, and K. Leblebicioglu, "Sleep spindles detection using short time fourier transform and neural networks," *IEEE Proceedings on the 2002 International Joint Conference on Neural Networks*, vol. 2, pp. 1631-1636, 2002.

VITA

Name: Parikshit Bapusaheb Deshpande

Address: Texas A&M University, Department of Electrical Engineering, College Station, TX 77843

Email Address: parikshit_deshpande@neo.tamu.edu

Education: B.E., Electrical Engineering, Government College of Engineering, Pune, India, 2001

M.S., Electrical Engineering, Texas A&M University, College Station, 2005