



Human Sleep Disorders:
Apnea and its diagnosis using the ECG

An Interactive Qualifying Project
submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfilment of the requirements for the
degree of Bachelor of Science

by

Colbert Sesanker
K Zin Thet
Jonathan Wong

Date: 10/07/2010

Report Submitted to:

Professor Carolina Ruiz, Project Advisor
Department of Computer Science
WPI

Table of Contents

Abstract	3
Acknowledgements	4
1. Introduction	5
2. Background on Human Sleep	6
2.1 Sleep Stages	6
2.2 Sleep Deprivation	7
2.3 Sleep Disorders	8
2.4 Sleep Apnea	8
2.5 Central Sleep Apnea (CSA)	11
2.6 Smoking and Sleep	13
2.7 Alcohol and Sleep	13
2.8 Depression and Sleep	14
2.9 Restless Leg Syndrome	14
2.10 Weight, Age and Sleep	14
3. Sleep Tools, Analysis and Databases	15
3.1 Sleep tools	15
3.2 Sleep Data Analysis	17
4. Materials and Methods	24
4.1 Materials	24
4.2 Methodology	26
5. Results	33
5.1 Test Results of patient a01	35
5.2 Test Results of patient c01	39
6. Discussion	43
7. Conclusions and Future Work	45
Bibliography	48
Appendix A – Our Apnea Detection Code	51

Abstract

Exploratory research was done on human sleep. From the knowledge gathered, we implemented a Matlab program that analyzes ElectroCardiogram (ECG) signals of human patients to diagnose Sleep Apnea. Our method was refined over time through experimentation and is based on algorithms pioneered by other researchers. All data for our experiments was collected from publically available online databases.

Acknowledgements

We would like to thank our advisor, Professor Ruiz, for her constant updates comments, suggestions and support, Professor Sergio A. Alvarez, of Boston College, and Amro Khasawneh (of WPI) for offering numerous helpful suggestions on the methodology and algorithm. Also, we would like to thank Billy Zhong (of WPI) for extensive help with implementing the code and the KDDRG group for aiding us in choosing a project.

1.Introduction

We researched the fundamentals of human sleep, and explored ongoing sleep research; we placed particular emphasis on how human sleep impacts society. While doing background research, we searched for and found freely available databases online, provided by sleep research professionals and medical teams. We explored how smoking affects sleep quality, how sleep deprivation increases motor vehicle crashes, sleep deprivation in adolescents, and sleep disorders. We finally decided to base the project on a single problem, sleep apnea. The main objective was to replicate a computer algorithm process that uses Electrocardiogram (ECG) recordings of human subjects to determine whether a person has sleep apnea or not (1) (2).A major advantage of using ECG to reliably diagnose sleep apnea is that ECG is relatively cheap and easy to apply to a patient, in contrast with other diagnostic methods. Our biggest challenge was to implement the appropriate algorithm for the ECG data so that the prediction results were accurate. Therefore, various data and signal analysis methods were explored and different software programs were tested. We chose Matlab for our implementation, and obtained conclusive results by the end of the project.

2. Background on Human Sleep

Although its exact function is unknown, sleep is absolutely essential for human health and constitutes about one third of our lives. Sleeping is generally defined as ‘a state of unconsciousness from which a person can be aroused’ (3). This definition is deceiving as sleeping is a very active and dynamic process, composed of various stages and cycles, each with its own characteristic mental and metabolic activity. The Electroencephalogram (EEG) is one of the more popular devices used to capture the characteristic stages of sleep in the form of electrical signals called brainwaves.

2.1 Sleep Stages

Sleep consists of periodic cycles, each lasts ninety minutes to two hours and repeats itself 4-5 times each night (3). Each cycle has two stages: Rapid Eye Movement (REM) and NON-REM (abbreviated NREM). REM sleep is a single stage while NREM sleep is broken up into the four stages described below.

In the first stage of NREM one is just beginning to fall asleep and can be easily awoken. Muscles begin to relax with occasional muscle twitches, and brain waves begin to slow down; from alpha (8-13 Hz) waves to theta waves (4-7 Hz). This stage only lasts for about 5-10% of total sleep time (4).

In the second stage, brain waves slow down even further with occasional bursts of rapid brain waves ranging from 11-14 Hz. Breathing and heart rate slow down and body temperature decreases slightly. Eye movements stop during this stage. This stage lasts for about 40-50% of total sleep time.

In the third stage, very slow brain waves called delta waves (0.5 - 2 Hz, 70% slower than a normal active person's brain) begin to appear. Only delta waves are present in the fourth stage; all eye and muscle movement cease in this stage. Stages three and four are described as deep sleep and differ only by the percentage of delta waves present. Together, these 2 stages make up about 20% of the total sleep time. Moreover, the immune system is restored and hormones are released during this stage (4).

REM sleep is the stage where dreaming occurs and is more related to processing and organizing mental content into a logical framework. REM sleep is characterized by a low voltage rapid EEG, and accounts for about 20-25% of total sleep. Paralysis sets in during REM sleep to help us from acting out our dreams (5).

2.2 Sleep Deprivation

If rats are deprived of sleep, they die within 3 weeks instead of living to their full 3 year life expectancy (3). For humans, when one is severely sleep deprived, brain activity and metabolism slows while judgment becomes substantially impaired (coordination may be worse than when intoxicated); also one's risk for chronic diseases like type two diabetes dramatically increases (3). Memory loss, immune system deficiencies, weight gain and mood swings are just some other side effects of sleep loss. Social interactions are also adversely affected by sleep loss. Sleep deprivation and insomnia are dangerous and require medical attention.

2.3 Sleep Disorders

Digressing from the normal cycles of NREM and REM sleep, we can examine various sleep disorders and how they affect sleep; also we can look at the effects that different substances, habits and individual traits have on sleep.

2.4 Sleep Apnea

Sleep apnea is a common, but often undiagnosed, disorder defined by the presence of one or more pauses in breathing or shallow breaths in one's sleep. Pauses in breathing can last from a few seconds to a few minutes (5). Sleep apnea can lead to daytime tiredness, as one often moves from deep sleep to light sleep when breathing pauses or becomes shallow (5). Aside from poor sleep quality and daytime tiredness, sleep apnea can lead to more serious problems; irregular heartbeat, hypertension, memory loss and even death are all products of untreated sleep apnea (5).

2.4.1 Obstructive Sleep Apnea (OSA)

Obstructive sleep apnea (OSA) is the most common form of apnea; and is usually caused by the over relaxation of throat and tongue muscles during a person's sleep, blocking the air passageway into the lungs. When a person is awake, these muscles help keep the airways stiff and open so that normal airflow is supplied to the lung. But when a person falls asleep these muscles are relaxed, though in normal cases this would not cause the blockage or narrowing of the airways. Obstructive sleep apnea is when the airways are blocked or narrowed during sleep (3).

2.4.2 Causes

Many factors contribute to the obstruction of the airways. Aside from the relaxation of throat muscles, abnormal biological structures in one's breathing system, e.g., bones and tissues can obstruct breathing in one's sleep. Weight is another factor; those overweight, have extra fatty tissue that could thicken the wall of the windpipe and narrow the airways. Age is a factor as older people have a less efficient way of sustaining stiffness of their throat and tongue muscles due to their physiological conditions (6).

2.4.3 Effects

OSA disrupts the normal airflow into the lung and decreases a person's oxygen intake. The oxygen level of the person will continue to decrease with time until it reaches a critical point where the brain would send signals to disturb the person's sleep to wake the person up. This biological mechanism helps to maintain oxygen level in one's body and thus keeping a person alive. However, with OSA the person will have frequent apnea episodes in a night and thus resulting in many disruptions to his/her sleep. Therefore there is a significant decrease of sleep quality and time for the patient. Repeated drops in oxygen levels and disrupted sleeping time would in turn cause the body to produce stress hormones and these hormones are linked to increased heart related problems.; Not only does OSA cause one to suffer from the consequences of sleep deprivation, but it also raises ones risk for heart disease (6).

2.4.4 Signs and Symptoms

Loud and chronic snoring is a common sign seen in most OSA patients, but not everyone who snores has sleep apnea . Choking or gasping is also seen in severe OSA cases where airway becomes completely blocked, causing one to gasp for air (3)

Many with obstructive sleep apnea don't know they have it, and a family members or bed partner has to tell them.

Daytime sleepiness is also another sign of sleep apnea. The symptoms are similar to people with long term sleep deprivation coupled with heart problems.

2.4.5 Treatments

Behavioral changes are the simplest treatment for mild OSA, e.g., sleeping on the side instead of lying flat on the back. Losing weight tends to be the best treatment for the obese, but it is extremely difficult due to the cause and effect cycle of OSA and obesity (the lethargy from OSA makes it hard to exercise). In severe cases, the only available option is to surgically remove/restore the complication of the air ways (6).

Medications may help with nasal airway blockages, and obesity (indirectly helping with the OSA conditions). Other useful medications include heart and respiratory medicines that help with other underlying conditions of OSA.

The most effective and popular method of treatment is using the Continuous Positive Airway Pressure (CPAP) device because it can be used for any level of OSA severity and is non-surgical. It is a nasal or face-mask device that provides continuous pressure to the airways to keep them open. However, the device is hard to get used to as it is difficult for a person to fall asleep with a continuous blow of air into the nose and the noise produced. If a patient is unable to adapt to the CPAP device, a Bi-Level Positive

Airway Pressure device (BIPAP) can be used. It is designed to modify the higher levels of airway pressure created by CPAP and it changes to a lower air pressure during exhaling and therefore removing the difficulty to breathe out against a high pressure air. The Auto-titrating CPAP machine is the most effective device among the CPAP machine types, as it automatically adjusts to the different pressure changes during a person's sleep: If the airway pressure changes due to apnea episodes, sleep positions or different sleep stages, the machine automatically detects the change and adjusts itself to allow for breathing with open airways (with minimal effort). However, each device has its own advantages and drawbacks depending on the patient's needs and apnea severity; therefore it is best to seek advice from a specialist and get tested for one's personally most effective treatment option (7).

2.5 Central Sleep Apnea (CSA)

Central sleep apnea, or CSA (much less common than OSA), is a form of apnea caused by the failure of the central nervous system to initiate the signals that trigger breathing. In CSA one makes no effort to breathe, there is no struggle. After a period of no breathing, there is a critical point where one gasps for air in an attempt to make up for the oxygen lost and the buildup of waste gases; in total, it accounts for less than 5% of all sleep apnea patients (3).

2.5.1 Causes

The primary causes of CSA are heart, respiratory and lung diseases, although complications in the brain stem which control breathing also cause CSA. People with long term heart, lung and/or respiratory problems develop abnormal carbon dioxide (CO₂) levels in their bloodstream; the brain starts to adapt to the high CO₂ levels in the blood stream, and becomes slower to respond to the increased CO₂ levels over time. If these underlying conditions become worse, and elevated CO₂ levels rise, the brain's CO₂ response becomes even slower and the CSA gets worse. As a result, short pauses in breathing occur during the person's sleep; sometimes these CSA episodes are severe enough to stop breathing for more than 20 seconds.

Malfunctions of or injuries to the brain stem could also cause the brain to not send the signals that initiate breathing, thus causing CSA.

2.5.2 Effects and Symptoms

CSA is less prevalent than OSA, but has the same physiological effects as OSA. The only difference is that with CSA, is the problem is more centralized in the central nervous system; thus it is generally harder to treat than OSA. (8).

2.5.3 Treatments

Treatments for CSA is similar to that of mild OSA, however the most effective treatment would be to treat the underlying conditions causing the brain function abnormalities (7). The CPAP and BIPAP devices are used to treat central sleep apnea

with less success. They are other devices to treat CSA (most work by altering CO₂ levels), e.g., inhaled CO₂ is used to increase the CO₂ above the apnea threshold (8).

2.6 Smoking and Sleep

According to Dr. Naresh M. Punjabi of John Hopkins University, although much research has been done on the effects of smoking on various medical conditions, little has been done on smoking and EEG (electroencephalogram) activity (9). According to Dr. Punjabi, the spectral analysis of brain-wave activity showed that compared with nonsmokers, smokers on average had more alpha waves (15.6% vs. 12.5%) and less delta waves (59.7% vs. 62.6%) during sleep. As delta waves correspond with deep sleep, and alpha waves correspond with wakeful relaxation, smokers may sleep lighter than they should (9).

2.7 Alcohol and Sleep

Alcohol consumption can alter the time required to fall asleep as well as total sleep time by disrupting the sequence, duration and efficiency of sleep states (10). Alcohol may decrease the time required to fall asleep because of its sedating effect; this is why many people consume alcohol to fall asleep. Alcohol consumed within an hour of bedtime seems to disrupt the second half of sleep, causing awakenings and disrupted dreams (10). Moderate Alcohol, consumed as much as six hours before sleep, can disrupt sleep even if there is no alcohol in the body, suggesting more fundamental interactions with sleep mechanisms (10). Alcohol also appears to make OSA symptoms worse and trigger episodes in those who normally do not present symptoms of OSA (10).

2.8 Depression and Sleep

Depression can range from moderate to severe depending on the presence and severity of certain symptoms. Depression's effect on sleep can range from mild to severe, but generally consists of sleeping too much or sleeping too little. Two of the major signs of depression are insomnia and disturbed sleep (11).

2.9 Restless Leg Syndrome

Restless leg syndrome may be linked with depression and is characterized by an uncomfortable sensation in the legs coupled with an uncontrollable urge to move the legs when at rest. These sensations can be very uncomfortable and cause movement and kicking while asleep (12). Restless leg syndrome is very similar to periodic limb movement disorder (PLMD) characterized by involuntary twitching or jerky movements of limbs while relaxed or asleep.

2.10 Weight, Age and Sleep

Obesity and being overweight may contribute to sleep disorders such as central sleep apnea (13). "Researchers found that people with severe sleep apnea who lost the recommended amount of weight were three times more likely to experience a complete remission of sleep apnea symptoms compared with people who did not lose weight" (13). The amount of sleep you need decreases with age. A newborn baby might sleep 20 hours a day. By age four, the average is 12 hours a day. By age 10, the average falls to 10 hours a day, and by adult hood decreases to 7-8 hours. Senior citizens need 7-8 hours, but often sleep lighter, awaken more frequently and generally spend less than 7-8 hours asleep (3).

3. Sleep Tools, Analysis and Databases

3.1 Sleep tools

Polysomnography is a study of physiological states during one's sleep using a multiple diagnostic test. A full polysomnogram (PSG) uses different monitors to record and convert 'limb, arm and eye movements, activity of the brain, rate of breathing, oxygen level and heart rate into time series data' (14). The Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG) and Electrocardiogram (ECG) are used to measure brain activity, eye movement, muscle tension and leg movements, and electrical activity of the heart respectively.

3.1.1 Electrocardiogram (ECG)

The sinus node (also known as the natural pacemaker) in the human heart releases electrical signals that cause the heart to contract and pump blood (14). The Electrocardiogram (ECG) detects these electric signals and records them. During a sleep study the ECG is primarily used to record changes in the activity of the heart over time.

3.1.2 Electromyogram (EMG)

The EMG uses several needle probes inserted into muscle tissue to detect the electrical potentials across the muscles produced by contraction; these potentials can be displaced visually on an oscilloscope or displayed audibly using a speaker. Throughout sleep, the EMG can monitor changes in muscle tension that serve to characterize events.

3.1.3 Electrooculogram (EOG)

EOG measures the electric potentials created when there is a change in the location of retina, and thus gives a measure of eye movement (15).

3.1.4 The Electroencephalogram (EEG)

The EEG measures brain activity using electrodes to measure electric potentials produced by the brain across the scalp. We can record these potentials as they change with time and visually see wave like patterns that we can use to define various brain activity (14).

3.2 Sleep Data Analysis

3.2.1 Brain waves

In general, EEG signal analysis and classification consists of three steps: Signal processing that involves filtering the important part of the signal essential for classification, “feature extraction used to extract unique features suitable for characterization purposes”, and indentifying optimum boundaries for the various classes of signals (16).

There are two general stages for these types of approaches. The first extracts the signals from the raw EEG signal, while the second uses a technique (e.g., statistical) to derive ‘a single number from the raw data’ (16). The Fourier transform works well when analyzing stationary signals whose statistical properties do not change with time. Non-stationary signals that change constantly with time are usually analyzed using wavelets or some other technique (16).

EEG signals are recorded after digitizing and applying the Fourier transforms to the detected brain potential (17). Primary recorded signals usually have noise; the signals are band pass filtered by using the FFT (Fast Fourier Transform) to get a succession of EEG spectra in frequency domain (18). By analyzing the primary signal in the frequency domain, we can categorize brain activity and sleep stages.

The brainwaves are separated into four ranges in order of frequency. The first range is from 15Hz to 40Hz. This range corresponds to the so called beta waves. Beta waves are emitted when a person is working or actively thinking. As the brain is actively concentrating, the frequency rate of beta waves is high and the amplitude of that is low. The brainwaves between the frequency ranges of 9Hz to 14Hz are alpha waves. Alpha

waves are produced during the first stage of NREM sleep when a person is beginning to fall asleep but can be easily awoken. The frequency of alpha waves is less but the amplitude is higher. Human brain produces theta waves during the second stage of NREM sleep when human's breathing and heart rate slow down. Theta waves have greater amplitude and slower frequency. The frequency range of theta waves is between 5Hz and 8Hz. When theta waves are produced temperature decreases slightly and eye movements stop. The final range, from 1.5 to 4 Hz, is delta waves. Delta waves are emitted during the deep dreamless sleep which is also called as the fourth stage of NREM sleep. As a person's brain is not thinking or working actively during the dreamless sleep stage, the frequency rate of beta waves is very low (bounded below by zero) and the amplitude of these waves is high (19).

3.2.2 Sleep Databases

Sleep data is stored in a database, or a well-structured repository of *data* possibly taking up many gigabytes or terabytes of disk space. A database is a series of data stored in *tables*, whose columns are headed by *attributes* that describe the data in each column. The rows or *tuples* in each table contain data, whose properties are identified by the attributes defined in the tables design. A good database should allow the user to *query* or ask questions about the data; an example of a query using a language such as SQL, is selecting an attribute then selecting the tuples with that attribute under some constraint (20). A good database should capture the relation between entities and their describing attributes. Also the database should allow for exploratory analysis such as finding trends or patterns, though patterns need not be apparent.

Sleep databases generally consist of two main types of data: The time-series information collected overnight from each patient is known as the micro dataset, while the patient background information like medical history, lifestyle, weight and perceived well being (gleaned from surveys and questionnaires) is known as the macro dataset. In The Sleep Disorder Center database (Center at Day Kimball Hospital, CT), the micro dataset was constructed by converting the signals collected from the full PSG into time series data (15). Other databases may not have a full PSG and use other tools to measure something less or more specific.

The time series information is used to construct various definitions that reveal certain characteristics of sleep. *Events* are “patterns in data that capture time-related occurrences of interest” (15). Sleep stages, body position, and heart rate are all examples of events. The REMbrandt software (used by The Sleep Disorder to provide the macro and micro dataset for the WPI sleep database) partitions time into discrete thirty second intervals known as *epochs*. Events are displayed every epoch, and the Polysomnogram signals are classified into sleep stages on an epoch by epoch basis (15). The four sleep stages are characterized epoch by epoch based on the event reading.

The macro dataset in the database from ‘The Sleep Disorder Center’ is divided into the following categories: a log of any communication history with the patient, an analysis report written by the interpreting physician, patient medical history, presleep study form, survey questions, post sleep study form, print out of technical summary, technician notes, and other miscellaneous information (15). Demographic data may include information such as age, height, weight and smoking habits. An Epworth Sleep

scale is used to determine the daytime sleepiness of the patient and a Beck Depression inventory is used to measure depression [16].

3.2.3 Existing sleep databases

Physiobank is an online source that provides public access to sleep databases (21). Different sleep databases vary by the tools used, the characteristics of the subjects, the database format and the objectives. The Apnea-ECG Database is an online database that aims to find a diagnostic for sleep apnea only using the ECG (a challenge by Physionet); it uses the ECG and other tools directed toward measuring breathing and diagnosing apnea; its subjects consisted of men and women 27 to 63 years of age (1). The Sleep Heart Health Study Polysomnogram Database consists of 1000 overnight polysomnograms and is aimed at finding a relationship between sleep disordered breathing and cardiovascular disease (15).

3.2.4 MIT-BIH Polysomnographic Database

The MIT-BIH Polysomnographic Database is multi-parameter sleep screening for obstructive sleep apnea from Boston's Beth Israel Deaconess Medical Center Sleep Laboratory. The study tests the effects of constant positive airway pressure (CPAP). CPAP usually 'prevents or significantly reduces airway obstruction' in these patients (22). The database contains over eighty hours of recordings, each of which includes an ECG signal, an invasive pressure signal, an EEG signal and a respiration signal (22). The six and seven channel recordings also include a respiratory effort signal; some include an EOG signal and an EMG signal while the remainder includes a cardiac stroke volume signal and earlobe oximeter signal (22).

The Santa Fe Time Series Competition Data Set B was extracted from the MIT-BIH Polysomnographic Database. The three samples in the dataset are heart rate, chest volume and blood oxygen concentration (1). The file is only about four hours and thirty minutes, significantly shorter than the full eighty hour recording.

3.2.5 The Sleep-EDF Database

The Sleep-EDF Database was compiled the *European Data format*, or EDF. EDF is a simple and ‘flexible format’ for exchange and storage of physical signals. It was developed by European ‘medical engineers’ and allows for a comparative analysis between researchers who choose to use the simple and now common format. In 1992, EDF became the standard for EEG and PSG recordings in commercial equipment and multicenter research projects; a standard which holds today (23).

Each EDF file in the database has a header specifying the patient along with recordings and signals including their calibration. The recordings were obtained from Caucasian males and females (21-35 years old) without any medication. The recordings contain horizontal EOG, FpzCz and PzOz EEG each sampled at 100 Hz. Recordings of oro-nasal airflow, rectal body temperature and submental-EMG envelope are also included. EDF files can be converted to and from Waveform Database format (WFDB), another database format used for further analysis (23).

3.2.6 The Sleep Heart Health Study Polysomnography Database

The Sleep Heart Health study (SHHS) Polysomnography Database investigates the relation between sleep, disordered breathing and cardiovascular disease. The study ‘recruited people from nine existing epidemiological studies where data on

cardiovascular risk factors had been collected previously' (24). For inclusion in the study participants had to age 40 years or older, have no previous treatment for sleep apnea, no tracheotomy and no current oxygen therapy (24). In all, 61441 individuals participated in the study (quite substantive) (24). 'Snorers were oversampled to increase the overall prevalence of sleep disordered breathing' (24). Studies were conducted in an 'unattended setting' such as the participants homes and attended by a trained technician (25). EEGs, EOGs EMGs, THOR/ABOR (for abdominal excursions), nasal-oral thermocouple, pulse oximetry, ECG, body position sensor and ambient light detectors comprised the recording montage.

Over 1000 polysomnograms were contributed to PhysioBank, but only polysomnograms with 'passing grades' were included. Grades are administered by the Reading Center at the SHHS and determined by the overall quality of the recordings. Each record is accompanied by a summary report, detailing overall quality, overall grade and specific grades (24).

3.2.7 St. Vincent's University Hospital / University College Dublin Sleep Apnea Database

This database is comprised of 25 full overnight polysomnograms from adults suspected of sleep disordered breathing. Subjects were selected from the Sleep Disorders Clinic at St Vincent's University Hospital, Dublin, 'for possible diagnosis of obstructive sleep apnea, central sleep apnea or primary snoring' (25). The 'Subjects had to be above 18 years of age, with no known cardiac disease, autonomic dysfunction, and not on medication known to interfere with heart rate' (26). The following signals were recorded: ' EEG (C3-A2), EEG (C4-A1), left EOG, right EOG, submental EMG, ECG (modified

lead V2), oro-nasal airflow (thermistor), ribcage movements, abdomen movements (uncalibrated strain gauges), oxygen saturation (finger pulse oximeter), snoring (tracheal microphone) and body position' (25).

4. Materials and Methods

4.1 Materials

In 2000, Physiobank challenged biomedical researchers to detect sleep apnea using only the ECG. This challenge was intended to ‘advance the state of art in this clinically significant problem’ and foster ‘wide-ranging collaborations’ (1). A \$500 prize was awarded to the research team that developed the best algorithm for detecting sleep apnea (there were no cases of central sleep apnea, all cases of apnea are obstructive or mixed) on a patient by patient basis. The second is a minute by minute classifier, where apnea is confirmed or denied each minute of a patient’s recording. They were 30 patients tested, 20 with apnea and 10 normal and 5 borderline (not clear whether they have apnea or not, and these are not included in the 30 tested). The score for the subject classifier ranges from 0 to 30 based on the number of subjects correctly diagnosed, while the score for the minute by minute classifier ranges from 0 to 17268 as they’re a total of 17268 minutes in the 30 files.

We choose to replicate the subject classifier for the paper with the highest scoring algorithm for the subject classifier. (They score 30/30 for the subject classifier and place fifth in the minute by minute classifier) [(27), (1)]. This subject classifier is pioneered by MR Jarvis (of California Institute of Technology) and PP Mitra (of Bell Laboratories) in a paper entitled Apnea Patients Characterized by 0.02 Hz Peak in the Multitaper Spectrogram of Electrocardiogram Signals (2).

The dataset for this challenge was provided by Doctor Thomas Penzel of Philips-University, Marburg, Germany (1). This dataset is referred to as the 'CinC Challenge 2000 dataset' and is publically available online at <http://www.physionet.org/physiobank/database/apnea-ecg/> (28). The CinC Challenge 2000 dataset contains 70 records (a total of 583 megabytes disk space) and is divided equally into a learning set and a test set. The learning set is used to train a predictive model using data mining techniques, while the test set is used to evaluate the predictive accuracy of the resulting model. Both the learning and the test set contain 35 minute by minute annotated files: In the learning set, files range from: a01 through a20 (apnea patients), b01 through b05 (borderline patients), and c01 through c10 (normal patients); in the test set, files range from x01 to x35 (28). Each file (varying in length from 7 to 10 hours) contains: 'a continuous digitized ECG signal, a set of apnea annotations (one for each minute of the recording, derived by human experts on the basis of simultaneously recorded respiration and related signals), and a set of machine-generated QRS (these are the largest spikes as seen in Figure 5.2 and Figure 5.7) annotations (28). 'Eight of the recordings in the learning set include three respiration signals (oronasal airflow measured using nasal thermistors, and chest and abdominal respiratory effort measured using inductive plethysmography) each digitized at 20 Hz, and an oxygen saturation signal digitized at 1 Hz' (28). These eight files serve as reference material for how the apnea annotations were made and may provide insight into relationships between the respiration and ECG signals.

Like the competitors in the Sleep apnea 2000 challenge, our data for our subject classifier was acquired from the CinC Challenge 2000 dataset. Physiobank ATM (online tool that supports conversion) was used to convert the learning set (in **.dat** format) into **.mat** files available for analysis in Matlab. Implementation was done exclusively in Matlab. In **.mat** format, each signal (assigned variable name 'val') is a 1 by 1000000 array, with sampling frequency of 100hz and time interval of .01 seconds (obtained from Physiobank ATM) yielding 10000 seconds per signal. An **.info** file and a **.m** file is needed in the Matlab path for proper analysis (provided by Physiobank ATM).

4.2 Methodology

4.2.1 Jarvis and Mitra's methods

First we describe how Jarvis and Mitra classify apnea, then we describe their subject classifier (2). We limit ourselves to a simplified version their subject classifier; in our spectral analysis methods, we don't use Slepian sequences and other advanced spectral analysis techniques.

A low frequency peak at .02 Hz is apparent in all apnea patients and no normal patients; thus this low frequency peak is used to classify apnea. This peak was also present in the point process spectrogram (a modified spectrogram, used in the minute by minute classifier, were a series of delta functions represent heartbeat peaks), suggesting the peak is more than an artifact, and truly captures something fundamental. (15).

Also, heartbeat appears less periodic during apnea periods (apparent from higher order frequencies) and breathing signals show variation, but overall breathing and

periodicity have proven to be poor indicators of apnea. Thus the .02Hz peak is their (as well as most other participants in the challenge) characteristic feature of choice.

Their subject classifier makes use of the 0.02Hz apnea signature. Without use of a sliding time window, the peak in the 0.02 Hz band may be overwhelmed by the power contributed to this band during normal breathing; thus a spectrogram rather than a frequency spectrum was used. Spectrograms are computed for each minute, minutes whose spectra at 0.03Hz and .1Hz simultaneously exceed a given threshold D are eliminated from consideration. The exact method they used to compute D is not specified. Now, after all minutes whose spectra are above the threshold D are eliminated, a peak in the frequency range of .005-.035Hz (an attempt to capture the characteristic .02Hz peak in apnea patients) is detected in the mean spectrum of the remaining minutes by a zero crossing of the derivative. Now, we measure the difference between the peak and the mean (denote this difference by A) relative to the background spectrum power (in the .005-.035Hz range). Apnea is defined when A exceeds 10% of the total mean spectrum power in this frequency range. If A is less than 10% of the total mean spectrum power in this frequency range, D is lowered and more minutes are considered for elimination; this process is iterated until either the patient is diagnosed with apnea, or 80% or more of the minutes are eliminated, where the patient is declared normal (2).

4.2.2 Our Subject Classifier

Like Mitra and Jarvis, we eliminate minutes based on their power on the .1Hz and .035Hz frequency bands; we do this by calculating the power at .03Hz and .1Hz for each minute, and define our threshold using the minutes in the top 2% of both frequencies. Our procedure is almost the same as Mitra and Jarvis's methods, except we don't use advanced spectral analysis techniques to calculate the spectrogram (we use Matlab's spectrogram function) and we use our own method to determine the threshold. Also, we use Matlab's maximum function to calculate the peak (the difference of the peak relative to the mean) in the .005-.035 frequency range and Matlab's average function (rather than integration) to calculate the mean spectrum in the .005-.035 frequency range. Numerous other modifications deviating from Mitra and Jarvis's methods were fine tuned over time based on the accuracy of our experimental methods and are detailed in the 6. Discussion.

A. Compute arrays with powers on .1 and .03 Hz for each minute

The first 1200 seconds of the ECG signal are eliminated as they have much larger amplitude than the rest of the signal. This is common practice in sleep studies as the first few minutes of a study tend to consist of noisy data. We divided the rest of the signal into minutes; one by one, we computed a spectrogram for each minute.

We assign the spectrogram to the following (to extract data in the form of matrices and arrays from the spectrogram):

```
[pspec fspec tspec] = spectrogram(temp, [], [], [], fs);
```

We extract the matrix `'pspec'`, whose rows are indexed by frequency, and columns indexed by time. In Matlab's spectrogram, the frequency is automatically normalized, so we scale the frequency array, `fspec`, by factor of $50/\pi$. However, if we passed the sampling frequency parameter, `fs`, into the function, the `fspec` is no longer normalized. Furthermore, on the above command each empty bracket, `[]`, will tell MATLAB to take the default value for those parameters. The frequency array has 1025 entries (un-normalized frequency ranges from 0-50 Hz) equally spaced by .0488 increments. Now, we compute the power at .03Hz and .1 Hz and store the powers in two separate arrays (`power_03` for .03 Hz, and `power0_1` for .1 Hz); the *k*th entry of each array is the *k*th power for the *k*th minute (power for .03Hz in `power_03` array and power for .1 Hz in `power0_1` array). As the frequency array is equally spaced by .0488 increments, the indices corresponding to .03Hz and .1 Hz in `fspec` are 1 and 2 respectively. We sum `pspec` across time at these two indices, for each minute, yielding 2 powers for each minute corresponding to .1Hz and .03 Hz (Namely, $\sum_{t=0}^1 \|\widehat{S}_f(t)\|^2$ where \widehat{S}_f is the amplitude of a given frequency band, but $\widehat{S}_f(t)$ can take complex values, thus we take the modulus ('abs' in Matlab)).

B. Removing noisy minutes

Now we compute two thresholds: `thres0_3` for `power_03` and `thes0_1` for `power0_1` as the smallest element in the top 2% of the arrays:

```
thres0_03 = quantile(temp_power0_03,percentile);
```

```
thres0_1 = quantile(temp_power0_01,percentile);
```

We use .98 for the `percentile`, as we are interested in the top 2 percent of the minutes. For each minute, we simultaneously compare the powers from .1 and .03Hz to `thres0_3` and `thres0_1`, and zero those minutes whose powers on .1 and .03 Hz simultaneously exceed both thresholds (this is just the intersection of minutes in the top two percent of both power arrays).

```
for k=1:num_min
    if ((power0_03(k) > thres0_03) && (power0_1(k) >
thres0_1))
        stemp_index = round(fs*start_time);
        s1_filt(stemp_index+1:(stemp_index+len_time)) = 0;
    end
    start_time = start_time + 60; % increment to the next
minute
end
percentile = percentile - 0.01; % decrement the percentile
by 0.01
```

After applying the threshold, we test for a certain peak in the signal as described in the next subsection below. If we find this peak, we declare apnea; if not, we lower the percentage until 80% or more of the minutes are discarded by the filter described above, where we declare the patient normal.

C. Test for Peak and Determine Whether to Declare Apnea, or Lower the Threshold and Retest the Peak

Now, we store the remaining minutes in a matrix (we denote this matrix by 'sumP' $\text{sumP}(a, b) = \text{sum}(\text{abs}(\text{pspec}(b, :)))$) with rows indexed by the remaining minutes and 1025 columns indexed by frequency (the entries along columns are the powers at the respective frequencies). By default, MATLAB will use 2048 points for the frequency vector; however since in the frequency response it contains both a positive and negative frequency, the 2048 is divided by half to only include the positive side for the spectrogram. 2048 divide by 2 is only 1024, so we have 1025 columns is because it also includes the zero frequency. Now, we sum sumP across rows and divide by the number of rows yielding an average for the minutes, denoted 'avgP'. We find those indices corresponding to .005 and .035 Hz in avgP but arrive at a problem: The columns (frequency) are indexed by .0488, so both .005 and .035 correspond to the first index. We resolve this by using the first index of avgP for .005Hz and the third index of avgP for .003Hz. We test for a peak in this range using Matlab's maximum function ($\text{peak} = \text{max}(\text{avgP}(1:3))$). Now, we test for the average power in this range using Matlab's mean function ($P = \text{mean}(\text{avgP}(1:3))$). Jarvis and Mitra's subject classifier measures the mean power in avgP as:

$$\frac{1}{.03} \int_{.005}^{.035} |Power(freq)| d(freq) , \text{ but we feel that the mean between the first}$$

and third indices (that is, the average of the powers at indices 1, 2, and 3) will suffice. The difference between the peak and the mean is measured relative to the mean. If this difference is greater than ten percent of the average, we declare apnea; otherwise we recompute the thresholds for the minutes in the top 2 percent of the remaining minutes

and recalculate the peak for the average of these minutes and measure it relative to the mean as before. We keep recalculating the thresholds as we eliminate more and more minutes. We stop when 80% or more of the minutes have been eliminated using the threshold procedure explained above, where we diagnose the patient as normal.

Namely:

```
    peak = max(avgP(1:3));
    P = mean(avgP(1:3));
    diff = peak - P;

    prob_sig = (1-
(length(find(s1_filt==0))/length(s1_filt)))*100;

    if ((prob_sig <= 20))
        is_Apnea = 0;
        break;
    else
        if ((diff > 0.1*P))
            is_Apnea = 1;
            break;
        end
    end
end
if (is_Apnea == 1)
    disp([temp_str ' has Apnea.']);
    apnea_count = apnea_count + 1;
else
    disp([temp_str ' is normal.']);
    norm_count = norm_count + 1;
end
end
end
```

For the code in its entirety, please see Appendix A – Our Apnea Detection Code.

5. Results

We tested the accuracy of our algorithm using the test set from the CinC Challenge 2000 database (28). We implemented an automated test function to verify the performance of our algorithm. Our test function allows us to input all the patients' individual ECG signals at one time and output the calculated accuracy of our algorithm. This makes it a lot easier for future implementation as we can modify the parameters of the algorithm, and then run it through our test function, which is far more practical than manually testing each patient at a time. We use a different threshold in this two patient demonstration. Also we calculate the peak using a Fourier Transform on the averaged minutes rather than a spectrogram; hence when we zoom into the .005-.035 Hz range we see many points, not just three. We compared our prediction results against the actual results provided in the files as shown in Table 5.1.

Patients	Does our algorithm predict that the patient has apnea?	Does the patient actually suffer from apnea?
a01	Yes	Yes
a02	Yes	Yes
a03	No	Yes
a04	No	Yes
a05	Yes	Yes
a06	Yes	Yes
a07	Yes	Yes
a08	Yes	Yes
a09	No	Yes
a10	Yes	Yes
a11	Yes	Yes
a12	No	Yes

a13	Yes	Yes
a14	Yes	Yes
a15	Yes	Yes
a16	No	Yes
a17	No	Yes
a18	Yes	Yes
a19	Yes	Yes
a20	Yes	Yes
c01	Yes	No
c02	No	No
c03	No	No
c04	Yes	No
c05	Yes	No
c06	No	No
c07	No	No
c08	Yes	No
c09	No	No
c10	Yes	No

Table 5.1 Performance of the algorithm

According to our results, shown above, our algorithm's predictions are correct for 14 patients out of 20 patients who have apnea and 5 patients out of 10 patients who are normal. Therefore, we can conclude that our algorithm has 63.33% accuracy.

In order to illustrate the process followed by the algorithm we implemented, we show below how the algorithm works on two given patients, one with apnea (a01) and one without apnea (c01).

5.1 Test Results of patient a01

5.1.1 Original Signal

Figure 5.1 shows the original ECG signal of patient a01.

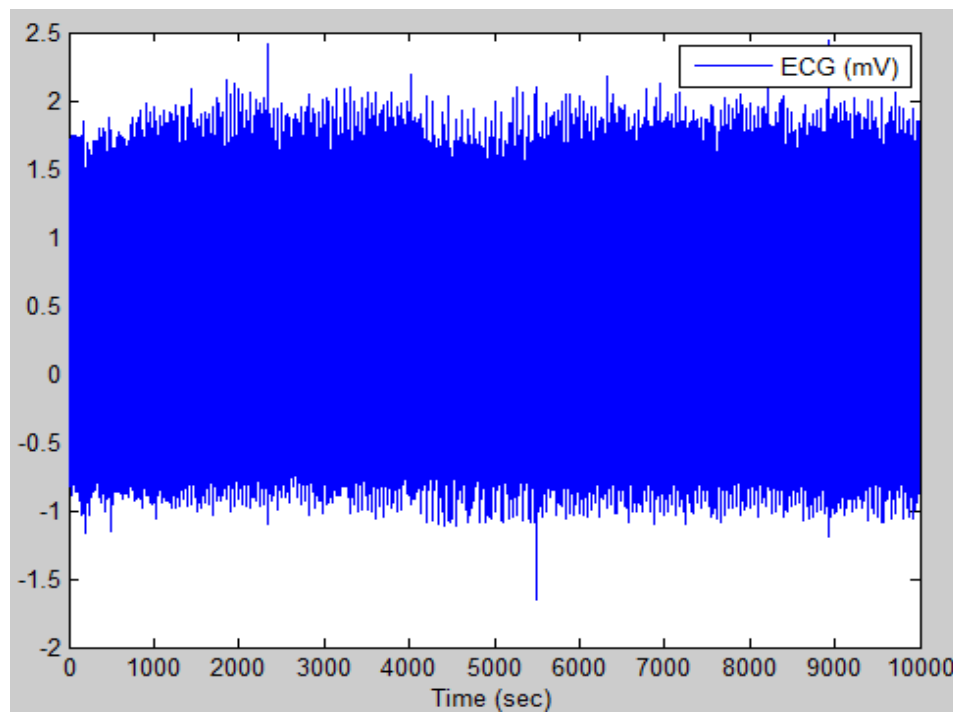


Figure 5.1 - Original ECG signal of patient a01

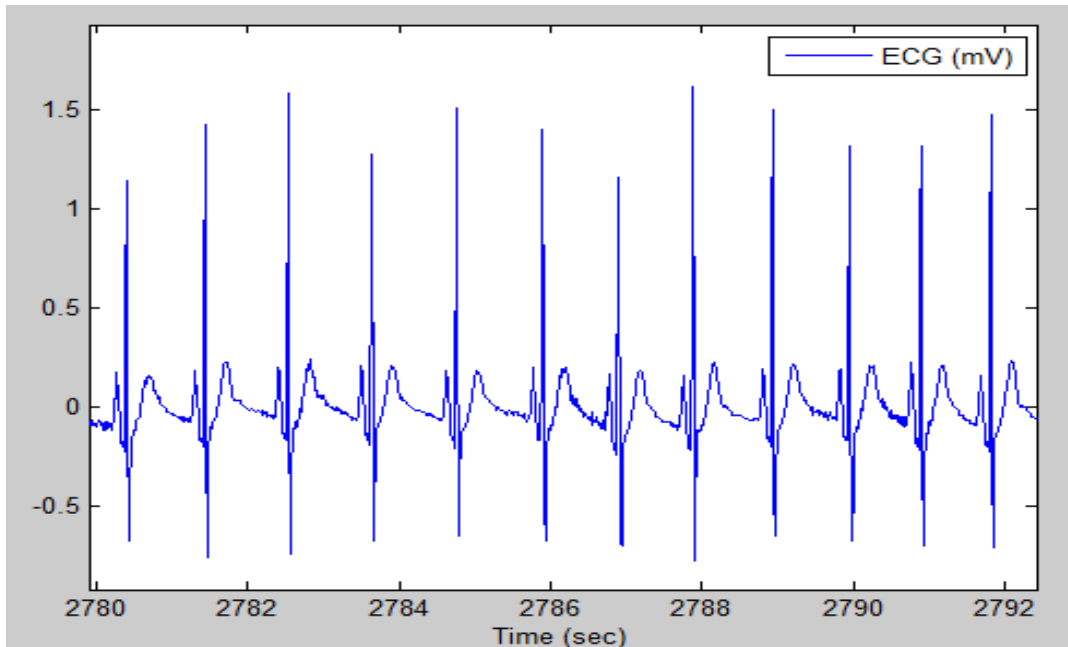


Figure 5.2 – Original ECG signal of patient a01 from 2780sec to 2792sec

As you can see from Figure 5.2, there are pulses running at different frequencies over a period of time intervals. By analyzing the amplitudes and the frequencies of these pulses, we can identify whether this patient has apnea or not.

5.1.2 Signal after discarding minutes

First we filtered out minute-long intervals by calculating a threshold to eliminate the top 2% of frequencies 0.1 Hz and 0.03Hz, as explained in the Methods section. We eliminated some minutes of the signal using our threshold procedure. Figure 5.3 shows the remaining ECG signal of patient a01 after discarding the minutes and Figure 5.4 shows the average over the Fourier Transform of each of the remaining minutes.

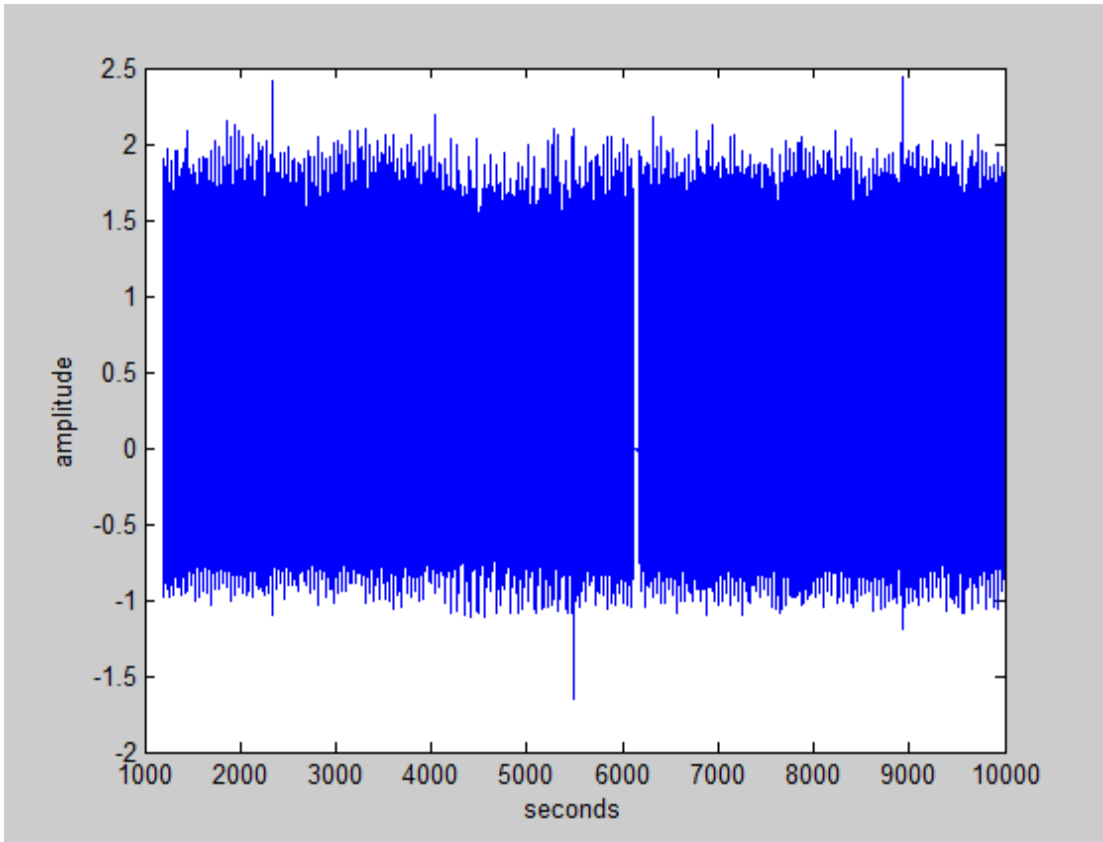


Figure 5.3 - Remaining ECG Signal of patient a01 after discarding some minutes

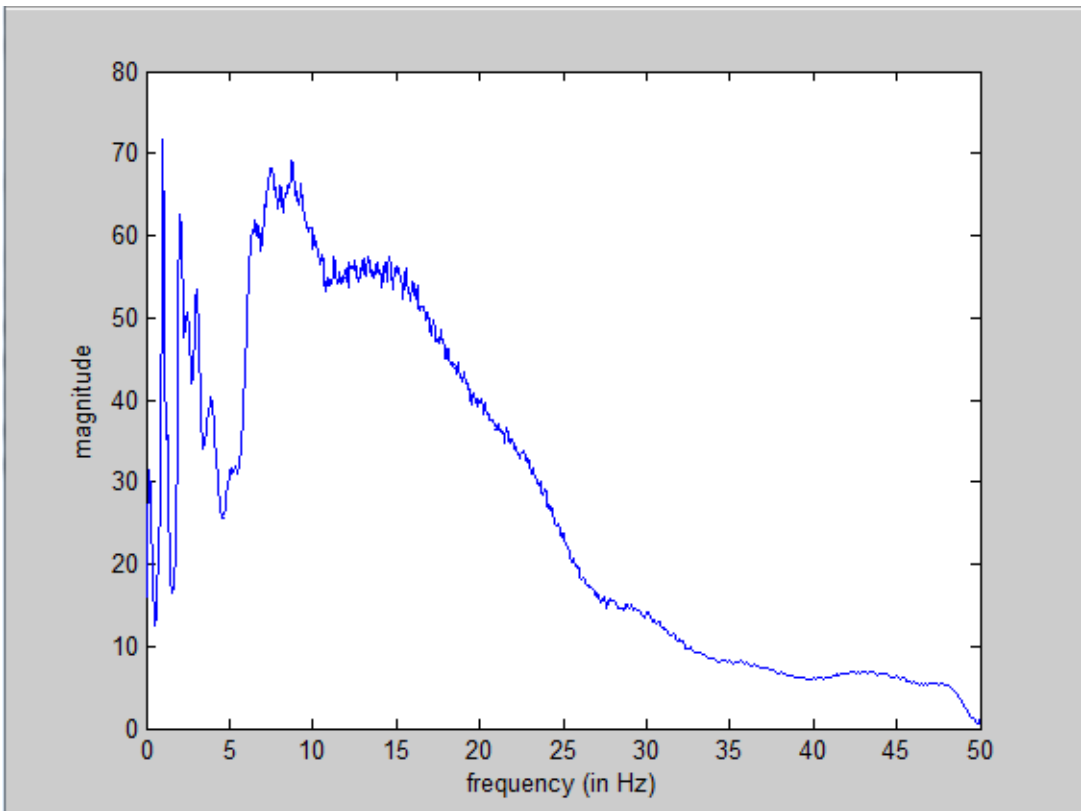


Figure 5.4 - Average Fourier Transform of the Remaining Signal for each minute

5.1.3 Finding Peak and Average Power in the range of Frequencies between 0.005Hz and 0.035Hz

After filtering the minutes by using the threshold method mentioned earlier, we investigated the average power and peak power. The average power between 0.005Hz and 0.035Hz is 198.5624, and the peak is 278.8562. From these two values, the difference is 80.2938, is much larger than the 10% of the average power, which is 27.8856. Hence, we verified that patient a01 has apnea by analyzing this patient's ECG signal.

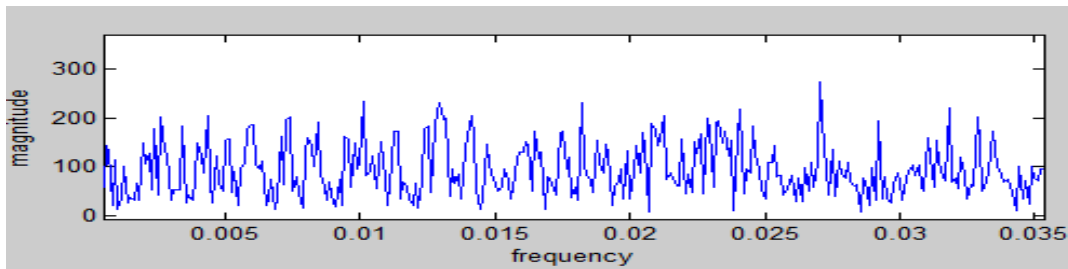


Figure 5.5 – Average Fourier Transform of the Signal at each minute between 0.005Hz and 0.035Hz

5.2 Test Results of patient c01

5.2.1 Original Signal

Figure 5.6 shows the original ECG signal of patient c01.

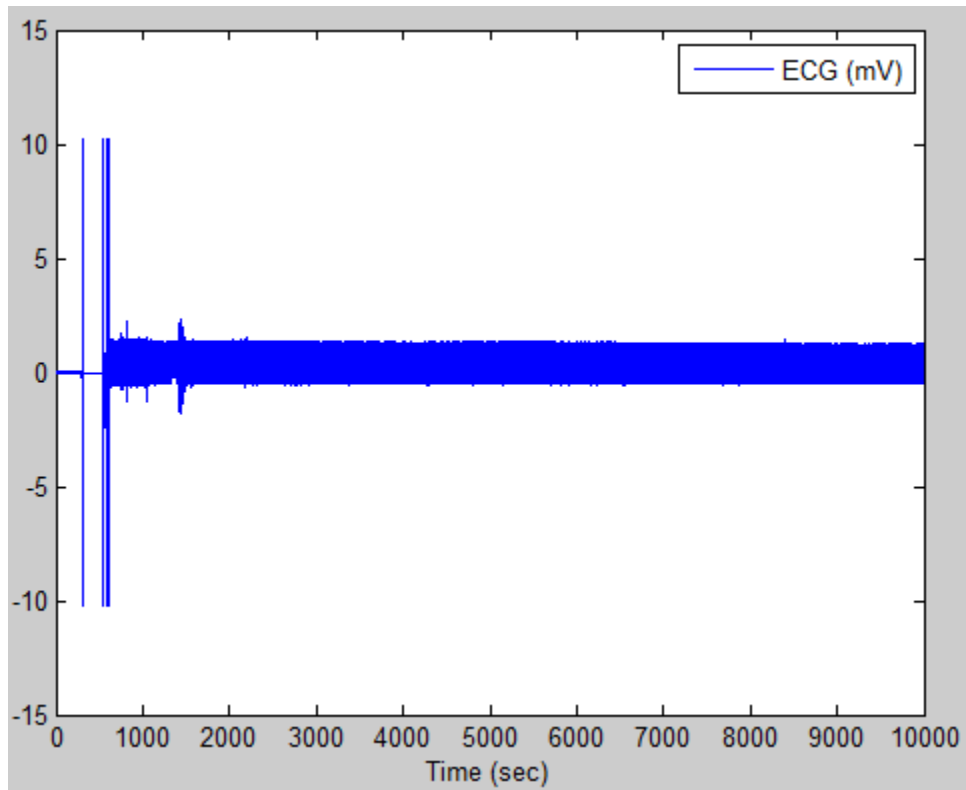


Figure 5.6 Original ECG signal of patient c01

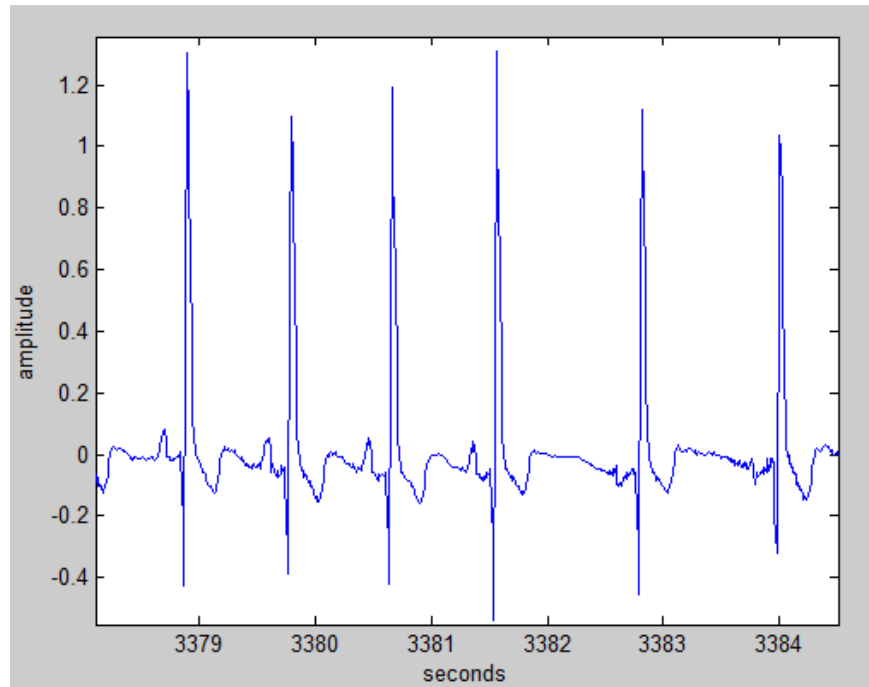


Figure 5.7 – Original ECG signal of patient c01 (zoomed in)

5.2.2 Signal after discarding some minutes

As explained in the previous section as well as the methods section, we first filter each minute interval using the threshold. The threshold is calculated as the minimum of the top 2% of the powers at the frequencies, 0.1Hz and 0.03Hz. Once we detected the power at 0.1Hz and 0.03Hz that is greater than the threshold, we can eliminate that part in the signal. Any minute in the original signal that is greater than both thresholds is rejected. After filtering parts of the signal out, we achieved the following signal, shown below. The first Figure 5.8 illustrates the ECG signal of patient c01 in the time domain. As can be seen on Figure 5.8, there are marks showing the rejections. The second figure, Figure 5.9, shows the average fourier transform of the averaged remaining minutes

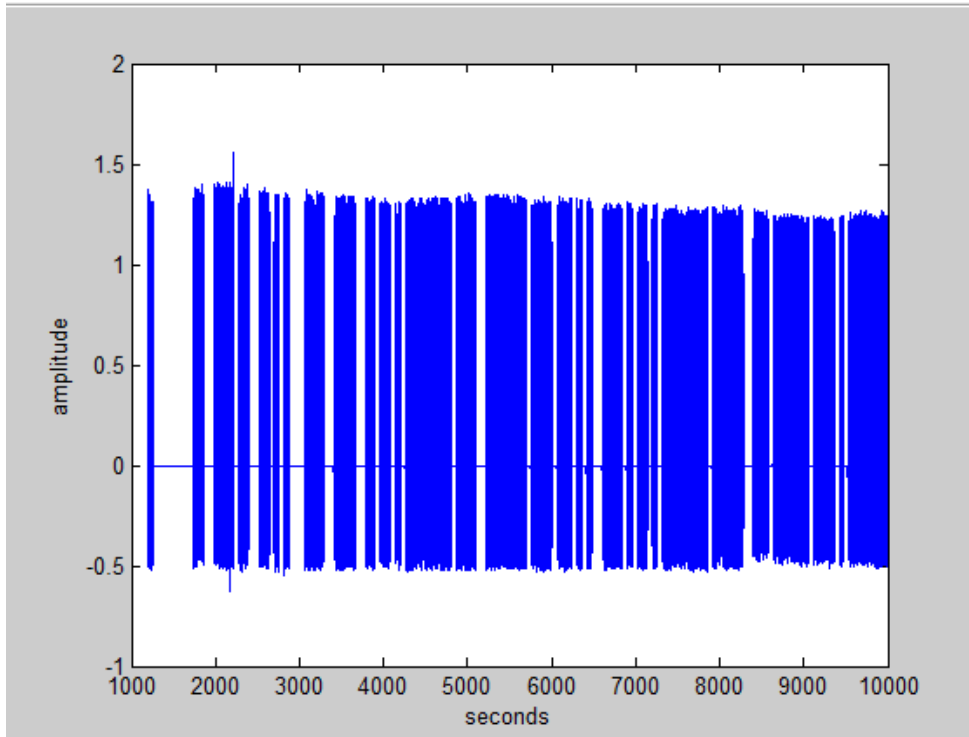


Figure 5.8 - Remaining ECG Signal of patient c01 after discarding some minutes

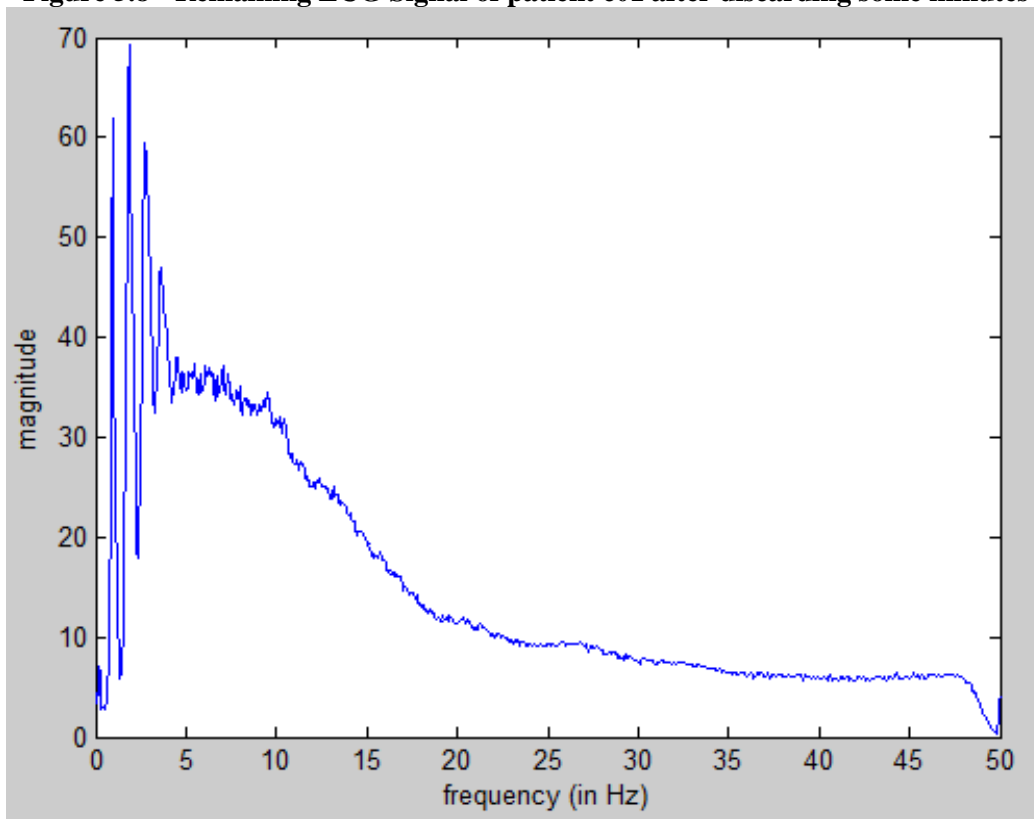


Figure 5.9 - Average Fourier Transform of the Remaining Signal for each minute

5.2.3 Finding Peak and Average Power in the range of Frequencies between 0.005Hz and 0.035Hz

Once the unused minutes are discarded, we investigated the average power and peak power in the frequency range, 0.005Hz and 0.035Hz. We found the average power to be 24.9784 and the peak to be 27.4213 for this patient. Since the difference between the peak and the average power displayed a smaller value than that of the 10% of the average power, we can confirm that this patient does not have apnea. The difference value is 2.4429, and the 10% of the average power is 2.4978. Furthermore, the signal remaining is also less than 20% of the original signal. Hence, we confirmed that patient, c01 does not display signs of apnea. The Fourier Transform of the signal of patient c01 can be seen in the figure as shown in Figure 5.10.

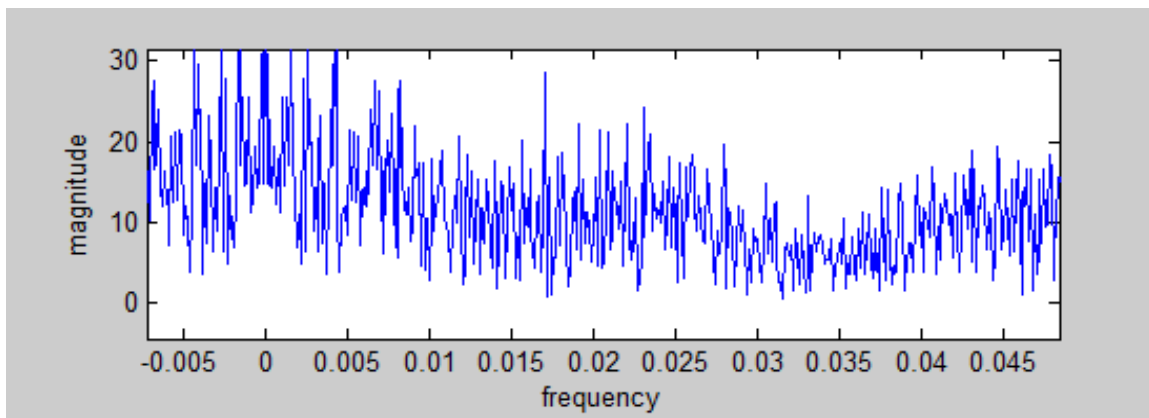


Figure 5.10 - Fourier Transform of the Signal between 0.005Hz and 0.035Hz

6. Discussion

We were not able to achieve the exact results as the ones shown in the Jarvis and Mitra's paper because we implemented a modified version of their method. In their paper, they mention the use of multitaper spectral estimation and trapper as the Slepian sequences (2). This method can predicate more accurate results (2). However, we did not implement the multitaper spectral estimation in our algorithm due to its complexity. Without Slepian sequences, our results are less accurate; this most likely accounts for the discrepancy between our results and the ones mentioned in the paper.

Furthermore, when we set the threshold to eliminate the minutes with powers at 0.1Hz and 0.03Hz that are greater than the threshold, we tried our best to replicate the subject classification procedures in MR Jarvis and PP Mitra's paper; we limited ourselves to simplified procedures that we progressively improved with experimentation. First, we found two maximum powers at the frequencies, 0.1Hz and 0.03Hz. Then, we set the minimum of these two powers (scaled down by 10%) as the threshold. Our results yielded too many false positives, thus we were inclined to define a new, stricter threshold. We defined our new threshold to eliminate minutes in the top 2% quartile of the powers at 0.1Hz and 0.03Hz; this gave us much better results.

Another cause for concern is our spectrogram was limited in its sampling of low frequencies as both .005Hz and .035Hz correspond to the same index. Basing our

algorithm on the greatest of three points instead of a continuum of points is a great cause for concern.

We experimented by varying the percentage of the peak many times to get better results. Certain factors biased most of the patients' diagnoses toward Apnea. Our best results were achieved when we measured the power of the difference between the peak and the mean to 10% of the mean spectrum (we experimented with various percentages).

Moreover, different methods of determining the power arrays at frequencies of 0.1Hz and 0.03Hz (for each minute to filter the noise and measure the peak) can also contribute to the discrepancy in the results. We brainstormed several methods to determine the power arrays and detect the peak. The first method was to use a spectrogram to find the power arrays and a frequency response to find the peak. The second method was to use a frequency response of the ECG signal to find the power arrays for noise cancellation and the peak for Apnea test. The final method was to use spectrogram to determine both the power arrays and the peak. We tested our algorithm with all methods as described above. In conclusion, we observed that our algorithm provided the best results when we use the spectrogram to filter out the noise and perform the Apnea test.

Our results are largely based on how we determined the power arrays at frequencies of 0.1Hz and 0.03Hz as well as the peak. These results can be affected by the threshold that we set to filter noise, and the percentage of the peak that we used to test for apnea. Currently, our results indicate an accuracy of 63.33%.

7. Conclusions and Future Work

This project was both a learning experience and a challenge. We started the project with the general objectives of learning about human sleep, learning about the impact of sleep in society, searching for human sleep data available online, and performing exploratory analysis of that data. More precise objectives on what data to look for and what type of analysis to perform was left open at the beginning and our group decided on these issues during the course of the project. We did lots of research and learned many things while deciding what we wanted to do. First, we spent our time researching human sleep and databases. We examined the fundamental stages of human sleep and their corresponding EEG patterns. We investigated the relationship between health and human sleep which eventually made way for our further research into sleep disorders. We learned many surprising, not necessarily intuitive, facts about human sleep, such as how suppressing certain sleep stages (REM) can alleviate depression and that people who sleep seven hours per day, on average, live longer.

The data mining skills we learned from examining real data from real patients in online databases are quite valuable. Now we are familiar with navigating publically available databases (e.g., Physiobank) and know the foundations of analyzing the data. Also, writing our code familiarized us with Matlab and we feel prepared for other more demanding tasks in Matlab.

We learned about various sleep disorders (sleep apnea, depression, restless leg syndrome), sleep-altering-substances (Alcohol, tobacco) and sleep-altering-conditions (Obesity). For the sleep disorders, we examined symptoms, diagnosis and treatment; for sleep-altering-substances we looked at which sleep stages they affected and their effect

on brainwaves while asleep, and for obesity we examined its relation to sleep apnea and its effect on overall sleep quality.

Initially we were caught between three topics: smoking and sleep, car accidents and sleep, and detecting central sleep apnea with the EEG. In our uncertainty we learned about each one (especially from the presentations we did on each one in the KDDRG seminar) but eventually (after our KDDRG presentation) settled on detecting central sleep apnea with the EEG. As central sleep apnea is more localized in the central nervous system, we hypothesized an EEG, rather than a ECG would be a more appropriate tool of diagnosis.

Constantly exploring alternatives, reevaluating ideas, finding novel ways of interacting as a team (e.g., use of a discussion board on myWPI), dividing the work according to our strengths and weaknesses and looking out for each other when we had other responsibilities, was a genuine learning experience; a written document summarizing our results is far from representative of the totality of our learning experience.

When our analysis objective became to detect apnea from patients' sleep signals, we decided to follow the approach presented by Jarvis and Mitra (2). They created an algorithm based on complex signal processing methods that produced excellent results at detecting sleep apnea from ECG signals. Their research paper presents their general approach but it is very vague in terms of the concrete details of their algorithm. It is hard to say whether the ambiguity of Jarvis and Mitra's method description detracted or contributed to the project experience. At first glance, one may think their methods ambiguity made it harder to learn the process, but our constant reevaluation between how

each of us interpreted the methods, drove our novel experimentation and produced understanding. If the paper had a clear methodology, we may have blindly parroted a prescribed procedure.

As for future work, one could continue fine-tuning the code developed in this project to improve its classification accuracy. Also, one could consider performing minute-by-minute detection of sleep apnea in the patients' signals, not just an overall patient classification. Such method could be based for instance on the approach presented in Jarvis and Mitra's paper. Other interesting future work includes looking into diagnosis of sleep Apnea using the EEG. We could use similar methods to the ones explored in our current methodology, but we would need to find a characteristic feature of Apnea in the EEG analogous to the peak at .02 Hz in the frequency spectrum for the ECG. We could look into new spectral analysis methods in our procedure like the Slepian sequences in Jarvis and Mitra's paper.

Bibliography

1. **Physiobank.** Physiobank Challenge 2000. *Physiobank*. [Online] April 27, 2009. [Cited: December 12, 2009.] <http://www.physionet.org/challenge/2000/>.
2. **Jarvis M. R, Mitra P. P.** *Apnea Patients Characterized by 0.2 Hz Peak in the Multitaper Spectrogram of Electrocardiogram Signals.* : IEEEExplore, Computers in Cardiology 2000, Sept. 24-27, 2000. Vols. 27:769-772. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=00898638&tag=1>
3. **Russo, M.B.** Sleep: Understanding the Basics. *EmedicineHealth*. [Online] October 31, 2005. [Cited: November 8, 2009.] http://www.emedicinehealth.com/sleep_understanding_the_basics/page11_em.htm.
4. *The Visual Scoring of Sleep in Adults.* **Silber.** s.l. : Clin Sleep Med 2007, Vols. 3(2); 121-31.
5. Sleep Apnea, What is it. *National Heart Lung and Blood Institute*. [Online] May 2009. [Cited: November 8, 2009.] http://www.nhlbi.nih.gov/health/dci/Diseases/SleepApnea/SleepApnea_WhatIs.html.
6. **Nabili, Siamak.** Sleep Apnea. *Medicinenet*. [Online] November 6, 2009. [Cited: April 20, 2010.] : http://www.medicinenet.com/sleep_apnea/article.htm.
7. **Staff, Mayo Clinic.** Central Sleep Apnea. *Mayo Clinic*. [Online] June 16, 2009. [Cited: April 20, 2010.] <http://www.mayoclinic.com/health/central-sleep-apnea/DS00995>.
8. **D.J, Eckert.** Central Sleep Apnea. *National Library of Medicine*. [Online] September 21, 2009. [Cited: April 20, 2010.] <http://www.nlm.nih.gov/medlineplus/ency/article/003997.htm>.
9. **Lie, D.** Cigarette Smoking Causes Sleep Disturbances. *MedscapeCME*. [Online] February 11, 2008. [Cited: November 8, 2009.] <http://cme.medscape.com/viewarticle/569973>.
10. **Gordis, E.** Alcohol and Sleep - Alcohol Alert No.41 1988. *National Institute of Alcohol Abuse and Alcoholism*. [Online] July 1988. [Cited: November 8, 2009.] <http://pubs.niaaa.nih.gov/publications/aa41.htm>.
11. **Fieve, R.** What is Depression. *WebMD*. [Online] September 12, 2009. [Cited: November 8, 2009.] <http://www.webmd.com/depression/guide/what-is-depression?page=3>.
12. **NINDS.** Restless Legs Syndrome. *National Institute of Neurological Disorders and Stroke*. [Online] March 30, 2010. [Cited: November 8, 2009.] http://www.ninds.nih.gov/disorders/restless_legs/detail_restless_legs.htm.

13. **Warner, J.** Weight Loss Helps Sleep Apnea. *WebMD*. [Online] September 28, 2009. [Cited: November 8, 2009.] <http://www.webmd.com/sleep-disorders/sleep-apnea/news/20090928/weight-loss-helps-sleep-apnea>.
14. **Cohen, J. S.** Electrocardiogram. *eMedicineHealth*. [Online] October 5, 2005. [Cited: November 8, 2009.] http://www.emedicinehealth.com/electrocardiogram_ecg/page11_em.htm.
15. **Shivin, Misra.** *A database for exploratory analysis of human sleep*. s.l. : M.S. Thesis. Department of Computer Science. Worcester Polytechnic Institute., 2008, May.
16. **Web-US.** What is The Function of Various Brainwaves? *Web-US*. [Online] 2005. [Cited: November 28, 2009.] http://www.scholarpedia.org/article/Hilbert_transform_for_brain_waves.
17. **Staff, M.** EEG. *Mayo Clinic*. [Online] May 21, 2009. [Cited: 23 2009, November.]
18. **Freeman, W. J.** Hilbert Transforms for Brainwaves. *Scholarpedia*. [Online] May 9, 2008. [Cited: November 28, 2009.] http://www.scholarpedia.org/article/Hilbert_transform_for_brain_waves.
19. *Spectral Analysis of the EEG in Meditation*. **Banquet, J.P.** 35, Netherlands : Elsevier, January 10, 1973, Vols. 143-151.
20. **Jeffrey D. Ullman, J.W.** *A First Course In Database Systems, Second Edition*. s.l. : Prentice Hall, 2002.
21. **Physiobank.** Physiobank. *Physiobank*. [Online] April 26, 2010. [Cited: 11 2010, May.] <http://www.physionet.org/physiobank/>.
22. *Components of a New Research Resource for Complex Physiologic Signals*. **Goldberger.** s.l. : Circulation 2000, 2000, June 13, Vols. 101(23):e215-e220.
23. **Kemp, Bob.** Sleep Recordings and Hypnograms in European Data Format. *Sleep Recordings and Hypnograms in European Data Format*. [Online] Physionet, September 18, 2007. [Cited: January 20, 2010.] <http://physionet.mit.edu/physiobank/database/sleep-edf/>.
24. **Robbins J, et al.** Sleep Heart Health Study Polysomnography Database. *Sleep Heart Health Study*. [Online] November 19, 2009. [Cited: January 20, 2010.] <http://www.jhucct.com/shhs/index.asp>.
25. **Heneghan, Conor.** St. Vincent's University Hospital/ University College Dublin Sleep Apnea Database. *Physionet*. [Online] February 8, 2008. [Cited: January 20, 2010.] <http://physionet.mit.edu/pn3/ucddb/>.

26. **Physiobank**. Physiobank Archive Index. *Physiobank*. [Online] April 26, 2010. [Cited: May 11, 2010.] <http://www.physionet.org/physiobank/database/>.
27. *Systematic comparison of different algorithms for apnea detection based on electrocardiogram recordings*. **Penzel T, McNames J, de Chazal P, Raymond B, Murray A, Moody G**. s.l. : Medical & Biological Engineering & Computing, 2002, Vols. 40:402-407.
28. **Physiobank**. CinC Challenge 2000. *Physiobank*. [Online] April 26, 2010. [Cited: May 11, 2010.] <http://www.physionet.org/physiobank//database/apnea-ecg/>.
29. **Rhodes, M**. Electroencephalogram. *eMedicineHealth*. [Online] July 29, 2008. [Cited: November 23, 2009.] <http://www.webmd.com/epilepsy/electroencephalogram-ecg-21508>.
30. *Electroencephalogram (EEG) Brain-Wave Feature extraction Using Short-Time Fourier Transform*. **Tamil, Emran Mohd**. Malaya : University of Malaya.
31. **Physiobank**. Apnea ECG Database. *Physiobank*. [Online] April 26, 2010. [Cited: May 12, 2010.] <http://www.physionet.org/physiobank/database/apnea-ecg/>.

Appendix A – Our Apnea Detection Code

```
%% Apnea_Test_Func1
% This function takes the inputs:
% type => type of the patient, a or c
% counter => number of patients
% From these inputs, it determines whether that patient has Apnea or
% not using spectrograms to analyze the power spectrum density. It
% also prints to the command window which patient has Apnea and which
% patient is normal. Finally, it outputs:
% apnea_count => number of patients who have apnea
% norm_count => number of patients who does not have apnea

function [apnea_count norm_count] = Apnea_Test_Func1(counter,type)
apnea_count = 0; % initialize counter for apnea patients
norm_count = 0; % initialize counter for normal patients
for a_count = 1:counter % number of patients
    skip_time = 1200; % amount of time to skip in the signal

    if a_count < 10 % adjust strings for single digit numbers
        temp_count = ['0' int2str(a_count)];
    else
        temp_count = int2str(a_count);
    end

    temp_str = [type temp_count 'm.mat']; % combine "mat" file string
    temp_str_2 = [type '01m.info']; % combine "info" file string
    [time val] = plotATM(temp_str, temp_str_2); % use plotATM to get signal & time vector

    fs = 100; % sampling frequency
    val = val((skip_time*fs+1):end); % new signal after skipping first 20 minutes
    time = time((skip_time*fs+1):end); % new time after skipping first 20 minutes
    Ts = 1/fs; % spacing between each sample
    totaltime = max(time); % find total time
    num_min = floor((totaltime-skip_time)/60); % find number of minute
    power0_03 = zeros(1,num_min); % initialize power array for 0.03 Hz
    power0_1 = zeros(1,num_min); % initialize power array for 0.1 Hz
    start_time = 0; % initialize start time
    end_time = 60; % initialize end time
    time_interval = [start_time+Ts:Ts:end_time]; % find the time interval in 1 minute
    len_time = length(time_interval); % find the length of time interval in 1
    % minute

    %% Find power at 0.1 Hz and 0.03 Hz
    for a=1:(num_min)
        index_start = round(fs*start_time); % find index of start time
        temp = val(index_start+1:index_start+len_time); % signal at 1 specific minute, need +1
        % because index for array starts at 1
        [pspec fspec tspec] = spectrogram(temp,[],[],[],fs); % spectrogram of an 1 minute
        % signal, [] => take default
        % values
        index0_1 = max(find(fspect<=0.1)); % index of frequency closest to 0.1 Hz
        index0_03 = max(find(fspect<=0.03)); % index of frequency closest to 0.03 Hz
        start_time = start_time + 60; % increment start time
        power0_03(a) = sum(abs(pspec(index0_03,:))); % filling up 0.03 Hz power array
        power0_1(a) = sum(abs(pspec(index0_1,:))); % filling up 0.1 Hz power array
    end

    %% Filter the Signal to 20% of total time
    s1_filt = val(1:end);
    percentile = .98; % initialize percentile to .98
    while (1)
        start_time = 0;
        thres0_03 = quantile(power0_03,percentile); % find threshold of power array for 0.03Hz
        thres0_1 = quantile(power0_1,percentile); % find threshold of power array for 0.1Hz
        for k=1:num_min
            if ((power0_03(k) > thres0_03) && (power0_1(k) > thres0_1))
                stemp_index = round(fs*start_time); % find the starting index of a specific
                % minute
            end
        end
    end
end
```

```

    s1_filt(stemp_index+1:(stemp_index+len_time)) = 0; % zero the signal when powers are
                                                % greater than the threshold
end
start_time = start_time + 60;                % increment to the next minute
end
percentile = percentile - 0.01;              % decrement the percentile by 0.01

%% Recompute Power in every frequency
temp_size = 1025;                            % size of frequency vector, achieved value
                                                % through experimentation

start_time = 0;
sumP = zeros(num_min,temp_size);
for a=1:num_min
    index_start = round(fs*start_time);
    temp = s1_filt(index_start+1:index_start+len_time); % signal of a specific minute
    [pspec fspec tspec] = spectrogram(temp,[],[],[],fs); % spectrogram of an 1 minute
                                                % signal, [] => take default
                                                % values
    start_time = start_time + 60;            % increment to next minute
    for b =1:length(pspec)
        sumP(a,b) = sum(abs(pspec(b,:)));    % two dimension array to store all the
                                                % power values. The dimension correspond
                                                % to time x freq.
    end
end

%% Find Avg Power
dim = size(sumP);                            % get dimension of sumP
avgP = zeros(1,dim(2));                     % initialize array to hold the average
                                                % power
for i=1:dim(2)
    total_temp = 0;                          % reset the total power to zero
    for k=1:dim(1)
        total_temp = total_temp + sumP(k,i); % calculate the total power
    end
    avgP(i) = total_temp/dim(1);             % calculate the average power using the
                                                % total power and the number of minutes
end

%% Find Index at f = 0.005Hz and 0.035Hz
index0_005 = max(find(fspect<=0.005));     % find index closest to 0.005 Hz
index0_035 = max(find(fspect<=0.035));     % find index closest to 0.035 Hz
if (index0_005 == index0_035)              % if indexes are equal increment one of
                                                % them by 2
    index0_035 = index0_035 + 2;
end
peak = max(avgP(index0_005:index0_035));    % find peak from 0.005Hz to 0.035Hz
P = mean(avgP(index0_005:index0_035));     % calculate average from 0.005Hz to
                                                % 0.035Hz
diff = peak - P;                           % find the difference between peak and
                                                % average
prob_sig = (1-(length(find(s1_filt==0))/length(s1_filt)))*100; % find the amount of
                                                % signal remaining after discarding
                                                % minutes
if ((prob_sig <= 20))                      % if amount of signal remaining is less
                                                % than or equal to 20% then patient does
                                                % not have Apnea
    is_Apnea = 0;
    break;
else
    if ((diff > 0.1*P))                    % if difference is greater than 10% of
                                                % average, then patient has Apnea
        is_Apnea = 1;
        break;
    end
end
end

% Display status of patient
if (is_Apnea == 1)
    disp([temp_str ' has Apnea.']);
end

```

```
    apnea_count = apnea_count + 1;
else
    disp([temp_str ' is normal.']);
    norm_count = norm_count + 1;
end
end
end
```