May 2010

Analysis on Societal and Technical Aspects of Human Sleep

Soe San Win

Worcester Polytechnic Institute

Follow this and additional works at: https://digitalcommons.wpi.edu/iqp-all

Repository Citation

This Unrestricted is brought to you for free and open access by the Interactive Qualifying Projects at Digital WPI. It has been accepted for inclusion in Interactive Qualifying Projects (All Years) by an authorized administrator of Digital WPI. For more information, please contact digitalwpi@wpi.edu.
Analysis on Societal and Technical Aspects of Human Sleep

An Interactive Qualifying Project Report

Submitted to the Faculty of
Worcester Polytechnic Institute

in partial fulfillment of the requirements for
Bachelor of Science Degree

By:

Soe San Win

Date: May 3, 2010

Approved:

__________________________________________

Professor Carolina Ruiz, Advisor
Department of Computer Science,
Worcester Polytechnic Institute (WPI)

Professor Sergio A. Alvarez, Co-Advisor
Department of Computer Science,
Boston College
ACKNOWLEDGEMENTS

Firstly, I sincerely would like to thank Professor Carolina Ruiz and Professor Sergio Alvarez for their guidance, support, careful feedback, and patience throughout the course of the project. I would also like to thank Mr. Amro Khasawneh, a Master’s student at Knowledge Discovery and Data Mining Research Group (KDDRG) in Worcester Polytechnic Institute (WPI) Computer Science department who gave me priceless feedback and help concerning this project. Last but not least, my lines of thank cannot be concluded without mentioning the past members of KDDRG group: Mr. Parameshvya Laxminarayan and Mr. Shivin Misra whose work provided me with a lot of information needed for my project.
ABSTRACT

Sleep is an essential metabolic process of human life. Thus, sleep disorders can produce disastrous consequences to individuals as well as to the dynamics of society. Throughout this Interactive Qualifying Project, I investigated the societal and technical aspects of human sleep in an exploratory manner.
EXECUTIVE SUMMARY

In this Interactive Qualifying Project, an exploratory study of societal and technical aspects of human sleep was performed.

Firstly, I examined the societal aspects of human sleep concerning the impacts of sleep on the society. This phase of the study includes study of the biological facts about sleep, and their correlation to the dynamics of the society.

Secondly, I explored the technical aspects of human sleep by studying contemporary research on human sleep in Computer Science, Psychology and Medicine, determined the representation of human sleep study data as can be found in contemporary sleep studies, researched the publicly available human sleep data repositories, and documented current rules and regulations of determining the sleep stages.

Thirdly, I investigated and identified sleep data that will support the research work of other graduate and undergraduate students in the Knowledge Discovery and Data Mining Research Group (KDDRG) at WPI. I extracted the desired data from over a 1000 human sleep studies a data repository maintained by KDDRG.
Table of Contents

Chapter 1........................................................................................................................................... 1
Introduction ............................................................................................................................................. 1
  1.1 Overview ....................................................................................................................................... 1
  1.2 Problem Statement ...................................................................................................................... 2
  1.3 Related Work ............................................................................................................................... 2

Chapter 2............................................................................................................................................... 5
Societal Aspects of Sleep ..................................................................................................................... 5
  2.1 Importance of Human Sleep ......................................................................................................... 5
  2.2 Mechanisms of Sleep .................................................................................................................. 6
  2.3 Sleep and Modern Society .......................................................................................................... 8
  2.4 Consequences of Inadequate Sleep ............................................................................................ 10

Chapter 3............................................................................................................................................. 12
Technical Aspects of Human Sleep .................................................................................................... 12
  3.1 Physiological Signal Information ............................................................................................... 12
    3.1.1 Time-series Signals .............................................................................................................. 12
    3.1.2 Electrode Positions ............................................................................................................ 13
    3.1.3 Methods of Physiological Data Acquisition .................................................................... 15
    3.1.4 European Data Format (EDF) ......................................................................................... 17
  3.2 Sleep Stage Characteristics ......................................................................................................... 19
    3.2.1 Standardized Procedure for Sleep Stage Identification ..................................................... 19
    3.2.2 Sleep Stage Scoring with R&K Model ............................................................................... 20
    3.2.3 Application of the R&K Model .......................................................................................... 23
    3.2.4 Wakeful Stage .................................................................................................................... 23
    3.2.5 NREM Stage 1 ................................................................................................................... 24
    3.2.6 NREM Stage 2 ................................................................................................................... 25
    3.2.7 NREM Stage 3 ................................................................................................................... 26
Chapter 1

Introduction

This chapter will describe the introductory material for the project. This includes: the nature of the project, the problem statement, rationale, and related work for the project.

1.1 Overview

Sleep is an essential and common metabolic feature of human beings. Therefore, irregularities and deficiencies in sleep could be detrimental to an individual’s life. If a significant number of individuals in a society suffer from sleep related problems, the society will suffer from great losses in productivity and progress. Hence, sleep related problems and sleep disorders have serious impact on the dynamics of a society. Due to the serious negative impact of sleep disorders, there has been ongoing research concerning sleep irregularities and their diagnosis.

In this Interactive Qualifying Project, we focus on identifying and acquiring sleep study data to be used in data mining algorithm development for sleep disorder diagnosis. The idea for this approach started with the thesis research work of (1) at KDDRG. Through his thesis research, Parameshvyas Laxminarayan worked on an algorithm that validates presence of certain attributes in individuals (e.g., sleep disorders) via the analysis of other observable attributes (e.g., personal traits and/or particular characteristics of sleep data). This Interactive Qualifying Project aids the work of other current and future members of KDDRG by investigating in depth human sleep stages, and by making available to them in an easy to use format sleep stage data
downloaded from a large data repository maintained by KDDRG. Currently, the project directly aids the research work of Mr. Amro Khasawneh whose research is to discover sleep structure types using expectation maximization via the use of Machine Learning Models (2).

1.2 Problem Statement

At present, KDDRG has already compiled a huge set of sleep data from 1,046 patients suffering from sleep disorders. The compilation and implementation of this database is documented in (3). The database is on Terabyte scale, and is one of the largest sleep study databases available. Currently, KDDRG is aiming to further develop data mining algorithms on this set of data to gain insight into the nature of sleep disorders. To contribute to this aim, the Interactive Qualifying Project was designed to investigate societal and technical issues revolving around an in-depth understanding of the sleep data collected.

The problem this Interactive Qualifying Project seeks to solve is to gain an in-depth understanding of the sleep study data as used in sleep research, the rules and regulations for determining the sleep stages, the existence of similar sleep study databases around the world, and to identify and extract the required sleep data out of the sleep data repositories to which we have access.

1.3 Related Work

There are two major research projects that preceded this Interactive Qualifying Project. The first research work was done by Mr. Parameshvyas Laxminarayan as part of his Master’s
Thesis (1). In the thesis, he proposed the Window-based Association Rule data mining algorithm which is a modified and extended version of the Set-and-Sequences Association Rule data mining algorithm developed at Worcester Polytechnic Institute (WPI) to support the mining association rules from complex data types (4). Laxminarayan’s research data comprised of subjective data obtained via survey questionnaires and objective data obtained via sleep studies of patients at UMass Medical Center in Worcester, Massachusetts and Day Kimball Hospital in Putnam, Connecticut. The research focused on both medical and statistical significance of the associations discovered by the algorithm. This research also led to the development of predictive classification models. The details of the project is documented in (1).

The second research work was done by Mr. Shivin Misra as part of his Master’s Thesis (3). In the thesis, Mr. Misra focused on the design and development of a database repository of both subjective and objective human sleep data. The subjective data for the database consists of 70 attributes such as demographic data, smoking, drinking, exercise habits, depression, daytime sleepiness, etc. The objective data consists of 50-55 time-series, each 6-8 hour long, per patient comprising different physiological signals tracking heart rate, blood oxygen level, blood pressure, snoring, body position and limb movements. 350 additional attributes summarizing the sleep stages, arousals and respiratory disturbance are also included. The research work focused on the design of a database system that can facilitate the effective handling of this huge data size and complexity in clinically meaningful terms, and the discovery of patterns by machine learning algorithms. The data came from 1,046 patients at the Sleep Disorder Center at Day Kimball Hospital in Putnam, Connecticut. Detailed information of the research work is documented in (3).
Another group doing research work on data mining of sleep data is part of the Institute of Psychophysiology and Rehabilitation at University at Kaunas in Kaunas, Lithuania. The group states that the process of obtaining sleep study data for data mining is complicated, time-consuming and expensive, since it involves over-night stays of patients in the sleep centers and laboratories at the hospitals, and usually a manual scoring of the sleep stages by sleep experts. 

Because the medical doctors specializing in sleep related problems can derive the sleep structure heuristically from the heart rate data, the group proposes to utilize Heart Inter-beat Interval (RR) and Stroke Volume (SV) sequences as the data for mining algorithms. The primary purpose of this approach is the economical advantage of data gathering since harvesting RR and SV data is much cheaper than acquiring sleep study data. Their data mining analysis employs bioinformatics algorithms for estimating parameters of nonlinear dynamics as well as methods of empirical mode decomposition and progressive fluctuation analysis. On the other hand, their classification system includes Discriminant Analysis, Artificial Neural Networks (ANN) and Vector Support Machines (VSM). Comprehensive information for the research work being done at their research center can be found in (5).
Chapter 2

Societal Aspects of Sleep

This chapter will describe societal aspects of human sleep I researched and documented for the purpose of the project.

2.1 Importance of Human Sleep

Sleep is a basic necessity of life. Sleep is as important to health and well-being of human beings as air, food and water. A good night’s sleep will give a person a refreshed and alert start of the day to face life’s daily challenges. However, without having required sleep hours, a person will not be prepared for his/her daily life both physically and mentally. Everything revolving around a sleep deprived individual will therefore be put at risk.

Individuals who get enough sleep will most likely have positive attitudes and demonstrate high performance on the next day, and sleep deprived individuals will most probably manifest negative impact on the surrounding community. Therefore, it is very important to notice that getting enough sleep has a significant effect on the overall quality of life in the society.

According to (6), quality sleep is defined as the amount of sleep hours required for an individual to be refreshed and alert on the next day. The amount of hours vary from person to person even though most sleep experts say seven or nine hours of sleep is generally enough to be defined as the quality sleep. (6) also states that sleep hours and patterns usually change with age. For instance, infants and toddlers require a lot of sleep, and they will require less sleep as they
grow old (6). The following chart as found in (6) breaks down the number of sleep that could be defined as quality sleep for each age range:

<table>
<thead>
<tr>
<th>Age</th>
<th>Sleep Needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newborns (1-2 months)</td>
<td>10.5-18 hours</td>
</tr>
<tr>
<td>Infants (3-11 months)</td>
<td>9-12 hours during night and 30-minute to two-hour naps, one to four times a day</td>
</tr>
<tr>
<td>Toddlers (1-3 years)</td>
<td>12-14 hours</td>
</tr>
<tr>
<td>Preschoolers (3-5 years)</td>
<td>11-13 hours</td>
</tr>
<tr>
<td>School-aged Children (5-12 years)</td>
<td>10-11 hours</td>
</tr>
<tr>
<td>Teens (11-17)</td>
<td>8.5-9.25 hours</td>
</tr>
<tr>
<td>Adults</td>
<td>7-9 hours</td>
</tr>
<tr>
<td>Older Adults</td>
<td>7-9 hours</td>
</tr>
</tbody>
</table>

**Figure 2.1.1: How Much Sleep Do You Really Need? (Taken From (6))**

Quality sleep also means that it is continuous and uninterrupted. Therefore, establishing a regular bed and wake schedule and achieving continuous sleep will help a person to flow along with his/her internal biological circadian clock, and go through all of the necessary sleep stages to obtain the restorative, energizing and revitalizing benefits of sleep. Even though the sleep patterns change, it is important to remember that the need for sleep remains the same for all ages.

### 2.2 Mechanisms of Sleep

Biologically, sleep is regulated by two brain processes, namely the restorative process and the sleep timing process. (6).

The restorative process can be modeled as the drive to sleep. This process naturally works by responding to the hours a person is awake. The longer the person is awake, the stronger the drive to sleep as the restorative process will respond to the lack of sleep.
On the other hand, the second process can be modeled as the sleep timer. The process controls the timing of sleep and wakefulness during the day-night cycle. This timing is regulated by the circadian biological clock in our brain. This part of the brain called SupraChiasmatic Nucleus (SCN) responds to light, and this is why people naturally tend to get sleepy at night when it is dark and there is not much light.

The circadian biological clock also controls the timing of other biological functions making sure that appropriate levels of metabolic processes occur at the right time. For example, important hormones are secreted, blood pressure is lowered and kidney functions change when a person sleeps and vice versa. Furthermore, research even indicates that memory is consolidated during sleep. Therefore, one can deduce from these facts that sleep deprivation which puts the circadian rhythm in off balance state, will have bad consequences on one’s health and life.

This circadian clock in the brain runs on a 24-hour period cycle. During this 24-hour cycle, the drive to sleep is the most intense around 2:00 AM - 4:00 AM along with a brief period of drive to sleep between 1:00 PM - 3:00 PM. A summary of the sleep mechanism can be found in following figure found at (6).

![Figure 2.2.1: Sleep Mechanism According to Circadian Clock (Taken From (6))](image)
Knowing the sleep mechanism and the circadian pattern is of societal importance in a dynamic society because a healthy society must have inhabitants whose sleeping habits mostly matching the circadian rhythm. The society will then have positive impact from the inhabitants as well as a great productivity.

2.3 Sleep and Modern Society

In reality, no matter how great the idea of getting quality sleep may be, people of modern society rarely get the quality sleep due to the competitiveness and demands of the modern society. With advancement in technology, lives of people as well as the structure of the society have become more complex, and increased level of demands and competition is introduced into the modern society.

With increased level of demands and competition comes the increased level of stress and number of hours from sleep allotted to meet the demands of the society. In addition, stress is found to be the number one cause of sleeping difficulties (6). The top triggers of stress in the present day society are found to be school-, college-or job-related pressures as well as family or marriage problems. We can clearly see that the top triggers of stress, which are the main causes of sleep deprivation and disorders, correlates to the societal issues.

Surveys conducted by the National Science Foundation (NSF) in the period (1999-2004) revealed that at least 40 million Americans suffer from over 70 different sleep disorders and 60 percent of adults report having sleep problems a few nights a week or more. In addition, more than 40 percent of adults experience daytime sleepiness severe enough to interfere with their daily activities at least a few days each month - with 20 percent reporting problem sleepiness a
few days a week or more. Furthermore, 69 percent of children experience one or more sleep problems a few nights or more during a week.

Groups that are at particular risk for sleep deprivation include night shift workers, medical physicians and residents, time-critical workers, parents and college students. Usually, sleep problems can be easily resolved if they are handled with care from the beginning. However, usually, they are not addressed promptly and effectively, leading to detrimental sleep related societal issues.

Shift workers comprise 20 percent of employees in the United States, and sleep is particularly elusive for them. Shift work forces people to try to sleep when there are activities around as well as when the biological circadian rhythm forces them to be awake. According to the study by (6), shift workers are two to five times more likely to fall asleep on their other daily routines than employees with regular, daytime work hours.

Similarly, medical physicians and residents who have to work all day round for possible emergencies are another group of individuals in the modern society whose sleep hours are minimized. Statistics show that average sleep hours for medical physicians and the residents revolve around five hours.

Also, workers in the round the clock industries working to beat the competition that has widespread use of nonstop automated systems to communicate and an increase in shift work will have difficulties finding the quality sleep required.

Furthermore, in the modern day families where people take care of their own children without the help of other family members, the role of parents can take a toll on sleep hours since people will need to pay attention to both their children and the daily professional work.
Lastly, surveys show that college students are the next group of individuals at risk due to the life style they lead. This usually results from the unusual schedules they have to accommodate for academic classes, social activities and work.

2.4 Consequences of Inadequate Sleep

If a sleep-deprived person does not take care of and protect himself or herself from sleep deprivation, the person may then start to experience worse consequences such as apathy, slowed speech and flattened emotional responses, impaired memory and an inability to be novel or multitask. Then, the individual will eventually fall into the state of sleep disorders.

In the August 2004 issue of the journal *Sleep*, Dr. Timothy Roehrs, the Director of research at the Sleep Disorders and Research Center at Henry Ford Hospital in Detroit, published one of the first studies to measure the effect of sleepiness on decision making and risk taking. He found that sleepiness does take a toll on effective decision making. Cited in the October 12, *New York Times Science* section, Dr. Roehrs and his colleagues paid sleepy and fully alert subjects to complete a series of computer tasks. At random times, they were given a choice to take their money and stop. Or they could forge ahead with the potential of either earning more money or losing it all if their work was not completed within an unknown remainder of time. Dr. Roehrs found that the alert people were very sensitive to the amount of work they needed to do to finish the tasks and understood the risk of losing their money if they didn't. But the sleepy subjects chose to quit the tasks prematurely or they risked losing everything by trying to finish the task for more money even when it was 100 percent likely that they would be unable to finish, said Dr. Roehrs. (7)
According to the National Commission on Sleep Disorders Research (1998) and reports from the National Highway Safety Administration (NHSA), (2002), high-profile accidents can partly be attributed to people suffering from a severe lack of sleep. (7)

Each year the cost of sleep disorders, sleep deprivation and sleepiness, according to the NCSDR, is estimated to be $15.9 million in direct costs and $50 to $100 billion a year in indirect and related costs. (7)

According to the NHSA, falling asleep while driving is responsible for at least 100,000 crashes, 71,000 injuries and 1,550 deaths each year in the United States. Young people in their teens and twenties, who are particularly susceptible to the effects of chronic sleep loss, are involved in more than half of the fall-asleep crashes on the nation's highways each year. According to the Department of Transportation (DOT), one to four percent of all highway crashes are due to sleepiness, especially in rural areas and four percent of these crashes are fatal.

Sleep loss also interferes with the learning of young people in our nation's schools, with 60 percent of grade school and high school children reporting that they are tired during the daytime and 15 percent of them admitting to falling asleep in class. (7) Moreover, sleep deprivation problems have been common among the college students for quite some time. (8) states that loss of sleep hours among college students due to their life style will lead to lower performance at academics and social life, and serious health problems such as weight gain, stroke, seizures and heart attack.

A society would not be healthy with its inhabitants facing problems with their lives. Therefore, the need for resolving the sleep related problems of the society, which are quite common in the modern day society, become important.
Chapter 3

Technical Aspects of Human Sleep

This chapter will describe the technical aspects of human sleep I researched and documented for the purpose of utilizing in the future research work done at KDDRG group.

3.1 Physiological Signal Information

To study human sleep for possible diagnosis of sleep disorders, we are required to have the means to quantify the biological events that happen during the sleep. This information is conveyed via the physiological signals that change during human sleep. This section will be dedicated to documenting the information about the signals involved in the sleep study. In doing so, we will thoroughly investigate the nature of each signal along with the specific information, such as electrode positions, methods of data acquirement, and sleep data format.

3.1.1 Time-series Signals

Common physiological time-series signals required for performing sleep study are: Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG) and Electrocardiogram (ECG). Sometimes, some or all of the signals monitoring snoring, blood oxygen level, and bodily movements are also included.

EEG signals measure the electrical potential of the brain. They are the first and foremost signals used to study sleep and determine the sleep stage characteristics. It was noted in (1) that sleep stages other than REM can be identified solely by the EEG signal information.
EOG signals measure the electric potential between the front end and the rear end of the eye to keep track of the eye movement. Only with the advent of the EOG signals were the researchers able to identify the REM stage of sleep, which involves rapid bursts of eye movements.

EMG signals measure the electric potential of the muscles (i.e., muscle tension) at chin, and several other places of the body. EMG signals are regarded as the more remarkable markers of the REM sleep stage in (1).

ECG signals measure the electrical activity of the heart with every single heart beat in micro-Volt scale. Also, the heart beat rate is measured and recorded in Beats per Minute (BPM).

### 3.1.2 Electrode Positions

The following figure taken from (9) shows the EEG electrode positions plus two reference electrode positions on the human cranium. Electrode positions for EOG, EMG and ECG are on other parts of the human body apart from the cranium. The article documented the comprehensive directions for correctly applying the electrodes for PSG study according to the International 10-20 System of Electrode Placement.

![Figure 3.1.1: Electrode Positions on Human Cranium (Taken from (9))](image-url)
There are two reference electrodes placed on the cranium as they can be seen from Figure 2.1. To be specific, (9) stated that A1 and A2 are placed on the bony areas behind the ears namely right outer canthis and left outer canthis respectively. The electrode positions Fz, which is 10 percent up from the nasion denoted as “Front” in Figure 2.1, and Cz, which is the center of the cranium, are usually used as ground electrode positions depending on the preference of the particular sleep technician. Usually, the PSG signals are identified by the corresponding electrode and reference electrode pairs.

For EEG signal measurements, electrode positions C3, C4, O1 and O2 are used. C3 and C4 are situated 20 percent left and right of the Cz position respectively. The positions O1 and O2 are placed 10 percent left and right of the position Oz, which is 10 percent up from the inion denoted as “Back” in Figure 2.1. According to (10), either one of C3-A2 (left-central) or C4-A1 (right-central) EEG signals from C3 or C4 electrode position is sufficient for the study of sleep stage characteristics. O1 and O2 electrode positions, O1-A2 (left-occipital) and O2-A1 (right-occipital) EEG signals respectively, are for measuring the occipital EEG for detecting and accessing sleep onsets or arousals during sleep.

For EOG measurements, electrodes are placed slightly above the outer canthus of the right eye, and slightly below the outer canthus of the left eye. The electrode positions are thus named ROC and LOC respectively. Two variations of the EOG signals are usually used in sleep stage characteristics study: LOC-A1, ROC-A1 pair and ROC-A1, LOC-A2 pair. The latter pair is usually used for the purpose of maximizing the amplitudes for the both EOG signals and minimizing the amplitude of pen deflections for conjugate eye movements as stated in (10). Additional EOG signals commonly used are XFlow, XSum, RMI, Phase and RR as stated by (3).
For EMG measurements, electrodes are mainly placed over the chin overlying the mentalis/submentalis muscles. Usually, EMG electrodes are also placed on the legs and arms to track the muscle tension and movements. The left and right anterior tibias for leg EMG recordings can be found one to two inch down from the knee cap and lateral one inch as stated in (9). EMG signals are usually named CHIN1, L LEG2, R LEG3 and ARMS4 according to (3).

For ECG measurements, electrodes are placed on the chest, located mid-clavicular in the third intercostals space. The usual ECG signals are named HEARTRATE and EKG8 according to (3).

The snoring sensor electrode is placed slightly lateral to the trachea, and the signal is named PSNORE. The bodily movements are recorded by the thoracic and abdominal belts placed over the respective positions, and the signals are named CHEST and ABDM. The blood oxygen concentration is monitored by the Oxymeter and the electrode is usually placed on the finger. The signal name for the blood oxygen level is usually SaO2. All of the above information is referred from (3).

3.1.3 Methods of Physiological Data Acquisition

Sleep study data has been gathered via sleep centers and laboratories at hospitals and research universities. The process of harvesting the data involves the patients staying over-night at those centers and laboratories, and the time-series data as well as sleep questionnaires are gathered during that time. The sleep study keeps track of the brain activity, eye movement, muscle activity, and heart activity, and records them into the time-series data as described in the previous section. The questionnaires acquire information about the individuals’ demographic
data such as age, height, weight, body mass index, collar size, medical history, sleep habits and cigarettes and alcohol consumption habits.

However, some researchers have been pointing out lately that this typical process includes unnecessary burdens on both the patients and the technicians. This can lead to inefficiency and decreased productivity along the way of sleep research and diagnosis of sleep irregularities. For example, limited availability of sleep centers and laboratories will result in long wait-lists, delaying data acquirement. Moreover, the process is expensive for the technical complexity as well as transportation and hosting costs for the patients. Lastly, due to aforementioned reasons, sleep study on the normal, healthy people for the purpose of research has been extremely difficult process. (11) shows a study done for the cost and effectiveness analysis of sleep data gathering processes in Europe. The study is based on the enquiry sent to the 500 sleep medicine providers in Europe. The results from the study show that the actual standard mean cost of a sleep study at sleep center or laboratory (i.e., 500 EURO) is more than two times of the standard mean cost of a sleep study being done remotely with ambulatory equipment (i.e., 238 EURO), and there have been delays of more than 10 years for diagnosis of sleep disorders in 25 percent of the patients as well as up to five physicians visits before referral to a sleep lab.

From this study, we can predict that even though the data gathering process for sleep study relies on the sleep centers and laboratories, the process in the future will depend on the ambulatory equipment that can remotely monitor the sleep data and gather the sleep data away from the sleep centers. There are several researchers and engineers working on developing such ambulatory monitoring equipment and process.
Two remarkable and novel approaches will be documented here for the purpose of completeness. Firstly, (12) proposed the non-obtrusive method for gathering the sleep data. The method they described measures the heartbeat, respiration, snoring, and body movement by means of embedding a thin air cushion, attached to ultra-sensitive pressure sensor, between the bed, and the mattress. In this way, any slight movements of the subject will be recorded via the pressure made on the air cushion. The frequencies and the manners of the pressure occurrences are different, and thus, are differentiated by AGC filters and envelop detection circuits. The bandwidth of each activity is determined: respiration 0.1-0.5 Hertz, heartbeat 5-10 Hertz, snoring 100-500 Hertz. Body movement is detected by heartbeat AGC filter. Heartbeat and respiration signals are recalculated every minute, and this is acquired by applying Fast Fourier Transform Algorithms and estimate the heartbeat, and respiration rate.

Secondly, (13) proposed a similar non-obtrusive approach. Their approach consists of textile sensors integrated into the bed, consisting of a large pillow case and a foot mat electrode made of woven yarn consisting of stainless steel fibers, a miniaturized electric module for pre-processing and storage of the data, and dedicated software for data analysis.

3.1.4 European Data Format (EDF)

European Data Format, which is more widely known as EDF, is a simple and flexible data format for handling multi-channel biological and physical signals. It was developed by the European biomedical engineers gathered at 1987 International Sleep Congress in Copenhagen. These engineers aimed for a universal data format to facilitate the comparison of results from their own respective sleep analysis algorithms. Therefore, they agreed on, and developed a very simple common data format on April, 1990. This data format has come to be known as the EDF
later on. EDF standards were first published in Electroencephalography and Clinical Neurophysiology Volume 82, pages 391-393, 1992. Since then, EDF became the standard for EEG recordings in commercial equipment and research projects on physiology. Moreover, an extension of EDF called EDF+ was developed in 2002. EDF+ has extra capacity to contain interrupted recordings, annotations, stimuli, events, and automatic or manual analysis results. Also, the stricter specifications of EDF+ enable the automatic calibration and localization of electrodes. EDF+ was published in Clinical Neurophysiology Volume 114, pages 1755-1761, 2003. Since then, EDF+ had a rapid growth in Clinical Neurophysiology, Sleep and Cardiology.

Both EDF and EDF+ are freely available. All information, specifications, publications and software about EDF are documented and available at (14).

Subsequent chapter’s sections will document the publicly available sleep database repositories. Most of the publicly available sleep study databases can be found at PhysioBank, the online repository for physiological time-series signal databases, software tools used for the analysis of those signals, growing collection of research papers, tutorials and reference materials for the biomedical research community. PhysioBank is maintained by a group of computer scientists, physicists, mathematicians, biomedical researchers, clinicians and educators at MIT, the Beth Israel Deaconess Medical Center, Harvard Medical School, Boston University and McGill University. Details about PhysioBank as well as PhysioNet are documented in (15).
3.2 Sleep Stage Characteristics

The nature of sleep stages is important in the field of sleep research. Understanding of sleep stages and their characteristics is the foundation to further research on sleep disorders, and the sleep stage scoring is also the focus of recent research. In this section, we will document the fundamental information on the sleep stages, and their characteristics.

3.2.1 Standardized Procedure for Sleep Stage Identification

With developments in the accuracy of electronic equipment, researchers were able to record high precision sleep data. However, there were no specific rules governing the devices and recordings, the results obtained were not standardized, and the efforts of the researchers to investigate the intricacies of sleep were not fruitful as they should have been. Therefore, the Association for Psychophysiological Study for Sleep (APSS) appointed a committee of sleep researchers to develop a standard for sleep stage characteristics. Two sleep researchers, Rechtschaffen and Kales brought forth a standardized model of sleep stages based on the specific characteristics observed from the sleep readings recorded through the night. This model is henceforth called R&K model. (16)

The R&K model breaks down the sleep progress characteristics into six sleep stages. They are wakeful stage, four stages of Non-Rapid Eye Movement Sleep (NREM Stage 1 – 4) and Rapid Eye Movement Stage (REM). EEG signals are essential in identifying the NREM Stages 1 through 4. However, only with EOG and EMG can one clearly identify the REM stage.

In addition, two important concepts defined for sleep stage scoring are Events and Epochs. Events are defined as the patterns in data that capture the time-related occurrences of
interest. The most common events are class of sleep stages. Epochs are generally defined as thirty-second intervals of sleep recordings.

3.2.2 Sleep Stage Scoring with R&K Model

Sleep stage scoring involves identifying the sleep stages on epoch by epoch basis. Sleep staging is guided by the frequency bands of the sleep signals recorded. Among the sleep signals, the R&K model place most emphasis on the EEG signal as stated in (16). The sleep frequency bands used for sleep staging are listed as below.

- Alpha Rhythm: 8 – 13 cycles per second or cps
- Beta Rhythm: more than 13 cps
- Delta Rhythm: less than 4 cps
- Theta Rhythm: 4 – 7 cps

Following figure in (17) explicitly shows the form of EEG signal in relation to its frequency band.

![Figure 3.2.1: Examples of EEG Signals for Corresponding Frequency Band (Taken from (17))](image)
The frequency band activity relating to the individual sleep stages are as follows: alpha band activities are dominant during the wakeful stage, theta band activities are prominent during NREM Stage 1, a mix of alpha band and beta band activities in NREM Stage 2, delta band activities are present during NREM Stage 3 and NREM Stage 4, and theta band and slow alpha activities in REM Stage. NREM Stage 2 is identified by mixed frequency band activity and occurrences of special wave forms called Sleep Spindles and K Complexes. The frequency band and waveform relationship is explicitly shown in figure found in (18) as follows.

![Diagram of EEG signals for corresponding frequency bands](image)

**Figure 3.2.2: Examples of EEG Signals for Corresponding Frequency Band (Taken from (18))**

According to (19) and (18), NREM Stage 3 and NREM Stage 4 are defined to be the deepest sleep stages. Together, they are usually referred to as Slow Wave Sleep (SWS) or Delta Sleep. According to (1), in the first half of a night’s sleep, a person will eventually goes through NREM Stage 1 through NREM Stage 4, and then reach to the first REM Stage. The first REM Stage is arrived at usually after ninety minutes once a person fall asleep. This very first REM Stage will last for about 10 minutes. Then, in the second half of sleep, NREM Stage 2 and REM stage will alternate. The following figure in (1) graphically depicts the normal progression of sleep in terms of a hypnogram.
Finally, following table summarizes the characteristics of sleep stages according to R&K model. The table is the abridged version of the one stated in (10).

### Table 3.2.1: Summary of R&K Sleep Stage Scoring (Adapted from (16))

<table>
<thead>
<tr>
<th>Stage/Stage</th>
<th>EEG</th>
<th>EOG</th>
<th>EMG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relaxed Wakefulness</td>
<td>Eyes closed: alpha (8-13cps)&lt;br&gt; Eyes open: lower voltage, mixed frequency</td>
<td>Voluntary control; REMs or none; blinks; SEMs* when drowsy</td>
<td>Voluntary movement; tonic activity relatively high</td>
</tr>
<tr>
<td>Stage 1</td>
<td>Relatively low voltage, mixed frequency&lt;br&gt; Maybe theta (3-7 cps) activity with greater amplitude</td>
<td>SEMs</td>
<td>Tonic activity, maybe slight decrease from waking</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Relatively low voltage, mixed frequency. Presence of sinusoidal waves (sleep spindles) 12-14 cps. Presence of negative sharp wave followed by slow positive component (K-complex) for time $\geq 0.5$ sec</td>
<td>Occasional SEMs</td>
<td>Tonic activity, low level</td>
</tr>
<tr>
<td>Stage 3</td>
<td>$\geq 20 \leq 50%$ high amplitude ($&gt;75$ $\mu$V), slow frequency ($\leq 2$cps)</td>
<td>None, picks up EEG</td>
<td>Tonic activity, low level</td>
</tr>
<tr>
<td>Stage 4</td>
<td>$&gt;50%$ high amplitude, slow frequency</td>
<td>None, picks up EEG</td>
<td>Tonic activity, low level</td>
</tr>
<tr>
<td>REM</td>
<td>Relatively low voltage, mixed frequency, sawtooth waves, theta activity; slow alpha</td>
<td>Phasic REMs</td>
<td>Tonic suppression</td>
</tr>
</tbody>
</table>

SEM – Slow Eye Movements
3.2.3 Application of the R&K Model

As part of the project, we practiced application of the R&K model by analyzing the real-world sleep signals. The sleep signals for this purpose are obtained from Sleep-EDF database (20). We took the EDF file of absolutely healthy subject, “sc4002e0.rec” for this verification.

As for the software tool to view the sleep data inside of the EDF file, we used the jEDF developed by Nizar Kerkeni. jEDF is a cross-platform software tool for visualization manipulation of data in EDF files. The software tool includes hypnogram manual scoring, fast Fourier Transform Analysis and EDF to text conversion. Amongst the publicly available software tools, we found that jEDF has the most useful features, and thus the choice was made to use it as the exploratory software tool in practicing R&K model. The developmental information and download of the software tool is available in (21).

Subsequent sections will show the analysis of the EEG signals from “sc4002e0.rec” according to R&K model.

3.2.4 Wakeful Stage

According to R&K Model, Wakeful Stage shows high frequency alpha band activity. When relaxed with eyes closed, human adults show alpha range. This pattern attenuates with attention as well as eyes being open. But for an extremely sleepy person, the same alpha rhythm may be present even when the eyes are open, and the pattern attenuates with intrusion of NREM Stage 1. The EEG signal we obtained from “sc4002e0.rec” showed this frequency behavior in the wakeful epochs as shown below.
3.2.5 NREM Stage 1

According to R&K Model, NREM Stage 1 will have slower theta frequency band activity and greater amplitude. Vertex sharp waves are common at this stage. Activity with highest amplitude is in the frequency band of theta range. The EEG signal we obtained showed this change in frequency and amplitude as we moved onto the NREM Stage 1 epoch.

![Wakeful Stage EEG Signal from sc4002e0.rec](image1)

**Figure 3.2.4: Wakeful Stage EEG Signal from sc4002e0.rec**

![NREM Stage 1 EEG Signal from sc4002e0.rec](image2)

**Figure 3.2.5: NREM Stage 1 EEG Signal from sc4002e0.rec**
3.2.6 NREM Stage 2

According to R&K Model, NREM Stage 2 will have mixed frequency activity and specific waveforms namely Sleep Spindles and K Complexes.

According to (10), Sleep Spindles are defined to be the bursts of EEG signal lasting for 0.5 seconds – 1.5 seconds during NREM Stage 2. Waves are not identified as spindles if they are not at least 0.5 seconds long. They consist of waxing and waning spindle shaped waves within 12 to 14 cps, and they show the onset of the stage. Sleep Spindles are also called, “Sigma Waves.” They may or may not be synchronous in both channels of EEG signals but they should be symmetrical and bilateral.

On the other hand, K Complexes are defined to be high voltage peaks in the EEG signal, usually higher than 100 micro-Volts and lasting longer than 0.5 seconds. They are paroxysmal waves with delineating negative sharp wave followed by positive component. K Complexes are usually followed by the bursts of Sleep Spindles, and they can be invoked by auditory stimuli.

The shape of the Sleep Spindles and K Complexes are defined in the following figure taken from (22).

![Figure 3.2.6: Examples of Sleep Spindle and K Complex (Taken from (22))](image-url)
Following figure shows the NREM Stage 2 of the EEG signal we obtained. As expected, we can see the Sleep Spindles, K Complexes as well as the mixed signal activity. The red, thin-lined circles denote the K Complexes, and the blue, thick-lined circles denote the Sleep Spindles.

![Figure 3.2.7: NREM Stage 2 EEG Signal from sc4002e0.rec](image)

### 3.2.7 NREM Stage 3

According to the R&K Model, NREM Stage 3 shows Delta frequency band activity. The EEG signals will show the slow frequency and high amplitude (i.e., peak to peak voltage of 75 micro-Volts) behavior. For NREM Stage 3, this waveform behavior will manifest between twenty percent and fifty percent of the time. The following figure we obtained shows the NREM Stage 3 specifications.

![Figure 3.2.7: NREM Stage 3 EEG Signal from sc4002e0.rec](image)
3.2.8 NREM Stage 4

According to the R&K Model, NREM Stage 4 shows Delta frequency band activity. The EEG waveform behavior of NREM Stage 4 is similar to that of NREM Stage 3. The only difference is that the manifestation of the waveform behavior will be more frequent than in NREM Stage 3, exceeding fifty percent of the time. The following figure we obtained showed this remarkable NREM Stage 4 signal.

![Figure 3.2.8: NREM Stage 4 EEG Signal from sc4002e0.rec](image)

3.2.9 REM Stage

According to the R&K Model, REM Stage is characterized by saw-tooth waves and slow theta frequency band activity. Some parts of the waves have highest vertexes, and operate in relatively slower alpha range. The waveform we obtained showed these features of REM Stage waveforms as in following figure.
3.2.9 Updates on the R&K Model

According to American Association for Sleep Medicine (AASM), there arises a need for the development of a new sleep stage scoring manual because even though the R&K model has been serving as the bible of sleep stage scoring for decades, there are still issues with the model. Several researchers have pointed out that R&K model relies solely on the central leads, and some of its rules are too broad, too narrow or too complex. Besides, there are a lot of different scoring rules extensions from R&K model lately in the field of sleep research. Therefore, there is an urgent need for a standardized universal scoring rule again. Due to these needs, AASM has developed a new sleep scoring manual, and the manual has been adopted since July 2008. Changes are added to technical requirements, electrode positions, sleep stage scoring rules, cardiac rules, respiratory rules and movement rules. (23) (24)

The new technical requirements are minimal and desirable sampling rate, low and high frequency filter settings, method of measuring actual individual impedance against a reference and minimal electrode impedances, and separate 50/60 Hz filter control for each channel.
The new electrode position requirements include EEG montages including frontal derivations combined with the existing central and occipital derivations, frontal derivations for clearer identification K Complexes and slow wave activity, revision of EOG electrode placement, clear definition of EMG electrode positions, and lead 2 placements for ECG.

The new sleep stage scoring rules has differences to the R&K model. The Wakeful Stage, NREM Stage 1, NREM Stage 2, and REM Stage are all remodeled as W, N1, N2, and R Stages respectively. The NREM Stage 3 and NREM Stage 4 are combined as a single stage, and is denoted as N3 Stage. Each stage will have more definitions and procedural notes. The “three-minute rule” is abolished so the new criteria for scoring N2 stage are defined. N2 sleep will be scored as soon as one or more K Complexes unassociated with arousal or one or more trains of sleep spindle occur in the first half of the epoch or the last half of the previous epoch. The major body movements will be either scored as W Stage or the Stage from the previous epoch.

The new cardiac rules include specifications of sinus tachycardia, bradycardia, asystole, wide and narrow complex tachycardia, and atrial fibrillation.

The new respiratory rules include the use of both the oronasal thermal sensor and nasal air pressure transducer for airflow detection and esophageal manometry, calibrated or uncalibrated inductance plethysmography for detection of respiratory effort, and development of Cheyne-Stokes definitions as well as a definition for the optional RERA events both the adult and pediatric population.

---

2 Once in NREM Stage 2, that score is maintained unless a reason to exit presents. One such reason to exit is described as the 3-minute rule. If no specific NREM stage 2 indicators appear, and in the absence of arousals and muscle tone changes that would alter the staging, continue to score all epochs as NREM Stage 2 for up to 3 minutes. At 3 minutes, if no specific indicators for NREM Stage 2 have occurred, scroll back 3 minutes and score those epochs as NREM Stage 1.
Finally, the new movement rules include maximum duration of a limb movement for ten seconds, minimum amplitude of 8 micor-Volts increase in EMG voltage above resting EMG, and scoring criteria for bruxism, REM behavior disorder (RBD), and sleep feature of rhythmic disorder.

All of above information is documented at (24).
Chapter 4

Sleep Data Extraction

This chapter will describe the core of our work for the project, including the identification of the types of sleep data required by the data mining algorithms development at KDDRG, research on how to efficiently access the required type of data, and extraction of the desired data.

4.1 Current Research at KDDRG Group at WPI

This section will briefly describe the current data mining algorithm development at the KDDRG group at WPI to identify the existence of the sleep disorder via the study of the PSG sleep study data from patients.

The current interactive qualifying project is going jointly with the thesis research project of Amro Khasawneh, a Masters student at KDDRG group, whose research is on developing the data mining algorithms for analyzing sleep disorders data.

In his research, he is interested in the transitioning between the sleep stages, and inputs to his algorithms will be the stages of sleep as recorded in the PSG sleep study data file saved in the EDF format, which is the common format of saving the PSG sleep study data.

My responsibility of the joint effort of projects is to identify the sleep data repositories we will have access to for obtaining more PSG sleep study data, identifying the software that will enable us to extract the sleep stage information out of the night long PSG sleep study data of patients, and extract the data for Mr. Khasawneh’s use in his research.
4.2 Sleep Data Repositories

This section will document the special data type for storing sleep data, namely European Data Format as well as the existing sleep database repositories publicly available for the sleep research, apart from the database at WPI by (3).

4.2.1 Sleep-EDF Database

This database can be found at (20). This database consists of publicly available sleep data for eight sleep studies in the years 1989 and 1994. The data is maintained in EDF files that consist of seven signals for the studies done in 1989, and five signals for those done in 1994. The subjects are Caucasian male and female, ranging from twenty-one years to thirty-five years of age, without any medication. The signals involve in the study are: horizontal EOG, FpzCz EEG, PzOz EEG, sub-mental-EMG envelope, oro-nasal airflow, rectal body temperature and an event marker, and EMG. Sampling frequencies of 100 Hertz was used for the first three signals, and 1 Hertz was used for the rest of the signals. The sleep data is scored manually according to R&K Model (16). Four sleep data instances denoted “sc” were from absolutely healthy subjects, and the other four sleep data instances denoted “st” were from subjects having mild difficulties sleeping.

4.2.2 Sleep Heart Health Study Polysomnography Database

This database can be found at (25). This database consists of the sleep studies conducted on subjects 40 years or older, who have no record of treatment for sleep apnea or tracheotomy or home oxygen therapy, to study the correlation of sleep disordered breathing and cardiovascular
disease. Participants were recruited from nine existing epidemiological studies in which data on cardiovascular risk factors had been collected previously. The complete SHHS database consists of 9736 polysomnograms and additional covariate data available via the approval of Sleep Heart Health Study Organization.

The sleep data is maintained in the form of EDF files as well as annotation files. These data was obtained in unattended setting, mostly at the subjects’ home, by the trained technicians. 11 signals were known to be recorded during the study. Those signals along with respective sampling frequencies are as follows: C3-A2 and EEGs at 250 Hertz, right and left EOGs at 50 Hertz, bipolar sub-mental EMG at 125 Hertz, thoracic and abdominal excursions (THOR and ABDO) at 10 Hertz, "airflow" detected by a nasal-oral thermocouple at 10 Hertz, finger-tip pulse oximetry at 1 Hertz, ECG from a bipolar lead at 125 Hertz and 250 Hertz for SHHS-1 and SHHS-2 respectively, Heart rate (PR) at 1 Hertz, body position, and ambient light. SHHS-1 is the initial phase of sleep study conducted between November 1995 and January 1998, and SHHS-2 is the second phase of sleep study conducted between January 2001 and June 2003.

The Sleep Heart Health Study is supported by National Heart, Lung and Blood Institute cooperative agreements of University of Washington, Boston University, University of Arizona, University of California, Davis, University of Minnesota, New York University, Johns Hopkins University, Case Western Reserve University, and Missouri Breaks Research.
4.2.3 St. Vincent University Hospital / University College Dublin Sleep Apnea Database

This database can be found at (26). This database consists of 25 full night sleep studies of subjects 18 years or older with probable sleep-disorder in breathing, i.e. obstructive sleep apnea, central sleep apnea or primary snoring. The subjects were known to be with no cardiac disease, autonomic dysfunction, and not on medication known to interfere with heart rate. The signals involved in the study are C3-A2 and C4-A1 EEGs, left and right EOGs, sub-mental EMG, ECG (modified lead V2), oro-nasal airflow (thermistor), ribcage movements, abdomen movements (un-calibrated strain gauges), oxygen saturation (finger pulse oximeter), snoring (tracheal microphone) and body position, and they are recorded via Jaeger-Toennies system (Erich Jaeger GmbH, Germany). In addition, three-channel Holter ECGs (V5, CC5 and V5R) were recorded using a Reynolds Lifecard CF system (Reynolds Medical, UK) and included in the EDF file.

4.2.4 MIT-BIH Polysomnographic Database

This database can be found at (27). This database is the collection of recordings of multiple physiologic signals during sleep for 18 subjects monitored in Boston Beth Israel Hospital Sleep Laboratory for probable diagnostic of Obstructive Sleep Apnea, and testing of effects of Constant Positive airway Pressure.

The database consists of four-, six- and seven-channel sleep study recordings with ECG signal annotated beat-by-beat, and EEG and respiration signals annotated with respect to sleep stages, and apnea, totaling in eighty hours worth of sleep data.
4.2.5 Polysomnography PhysioRepository

This database can be found at (28). This database is a contribution to the Sleep Heart Health Study Organization by Sleep and Epidemiology Research Center (SERC) at Case Western Reserve University, Cleveland, Ohio. The data is publicly available on the terms and conditions set by SHHS.

The initial lunch of the repository consists of 1000 sleep studies collected during the period 2001 – 2003. The repository query is attribute-based to ease desired data extraction for researchers. The signals studied are four to eight hours long sleep signals including the EEG, heart rate, and respiratory data.

4.3 Data Extraction

This section will describe the process of identifying the usable data according to Mr. Khasawneh’s specifications. To this end, Mr. Khasawneh and I have to identify the inputs into his algorithm, and research on the sleep data repositories that might have grant access to more PSG sleep study data apart from the data of 1046 patients maintained at the KDDRG group’s database.

4.3.1 Identifying the Required Data

We identified that Khasawneh’s research will be looking at the transitioning among the sleep stages, and identify the sleep disorder from the certain pattern of sleep stage transitioning. Therefore, we decided that we will need to extract the sleep stage sequence data out of the PSG sleep study data stored in the form of EDF data format.
Here, we find some complications. By this point in time, we know that the EDF data format contains different kinds of physiological data as described in Chapter 3. However, we were not sure if there will be explicit sleep stage information also stored in the sleep study EDF files or we would be required to define and write a program that will determine the sleep stages using one or combination of the physiological signals as found in the EDF sleep study files. We eventually found a solution to this obstacle, and it is reported in Section 4.3 as it can be found later in this Chapter.

4.3.2 Choosing the Repository for the Study

After identifying the data specifications we need for the joint project, we started looking out for the sleep data repositories to which we would have access for significant number of more EDF data files of sleep study apart from the database at the KDDRG group at WPI.

For this task, we tried to look into the means of obtaining the sleep study data from the sleep repositories as described in previous section. After some amount of research, we found that we have access to the repositories on PhysioBank, and that we do not have access to other repositories. However, unfortunately, we were faced with the situation that most of the sleep data on the PhysioBank were not meeting our data specifications, that is, they do not have all of the types of the physiological signals or they do not have the staging information in them. On the other hand, the saving grace is that after careful checking, we became sure that the sleep study data in the KDDRG database meets the requirements specifications we had.

Therefore, we decided that we will just extract the data from our own KDDRG database. To this end, we will need to look into the software, and the methods required for the process. Moreover, there are some special qualifications that are required to be met before we can start
dealing with the data because the process involves with the real world data that calls for privacy issues.

### 4.4 Extraction of Data

This section will briefly describe about the process we took to extract the data out of the KDDRG group database. The process deals with the special qualification required for dealing with private patient data, and utilizing special software to extract our desired data out of the EDF sleep study files.

#### 4.4.1 National Institute of Health (NIH) Protecting Human Research Participants

National Institute of Health (NIH) requires by law that any academic or industrial researcher who is working with the human subjects must pass the qualification course to be certified to deal with human subjects in the research.

NIH requires this to maintain the privacy and integrity of the participating human subjects directly or indirectly. I took the online course hosted at the NIH website, and pass the qualification test to become certified to legally work with the sleep data.

More information on the NIH policy on human research participants, can be found on the NIH website (29).

#### 4.4.2 Rembrandt Software

We started looking into different software candidates for extracting the data. We have encountered a lot of free as well as commercial software along the course of the project. As we
would expect, there are pros and cons of free software versus the commercial software. As for the free software, the pros are that they are free, and usually not bound by any license or fees. The cons with the free software are that they are usually not cut out with the functionality we are looking for, and their results weren’t guaranteed to be correct. On the other hand, the commercial software has the pros of being trustworthy and provides with a full range of functionalities that every sleep study research would be looking for. However, the cons are that they are not free, and the cost associating with each commercial software can be astoundingly high, and not cut out to fit with the budget goals of the purpose of our project.

After careful considerations of all the available options, we came down to the conclusion that we will be using the Rembrandt software by EMBLA systems Software Company. Rembrandt is a Microsoft Windows based sleep monitoring and analysis system for both clinical and research applications. The software is capable of performing different kinds sleep study including OSAS, pediatric, insomnia and cardiovascular investigations. The software works on the physiological signals acquired via Monet and EMBLA systems hardware.

4.4.3 Method of Extracting the Data

Rembrandt software possesses the “Event Traces” functionality that can access all of the signals and information stored in the EDF file, and downloads it to the simple text file format. However, one downside of this feature is that there is no feature available for automated accessing and downloading the information from multiple EDF files.

Therefore, the download process has to go through the steps of opening the EDF file for each patient manually, and then extracting the “Event Traces” one by one. I accessed the computer server hosting the database via Microsoft’s Remote Desktop service, and manually extracted the desired data. There were 1046 patient EDF files in the database, and by the end of
the project, I have successfully extracted the data for each and every one of the EDF file in the database.

After the data download is done, Mr. Khasawneh already has the C program that will parse the output text file, and extract just the sleep stage sequence information from the downloaded files. He had been working on this since my download started, and he already has the working C program by the time I finished extracting the data.

4.4.4 Step-by-Step Guide of Data Extraction

This section shall document step-by-step the process of data extraction using the Rembrandt software.
1. Start up the Rembrandt software.

Figure 4.4.4.1: Step 1
2. Go to File→Open. You will see the following screen.

![Image of the screen](image.png)

**Figure 4.4.4.2: Step 2**
3. Click on Browse button in Step 2 to see the following screen. Choose the data server you would like to download from in the menu as can be seen below.

Figure 4.4.4.3: Step 3
4. Go to the folder where the data is stored on the server.

Figure 4.4.4.4: Step 4
5. Choose the file, and then click OK.

Figure 4.4.4.5: Step 5
6. The software will tell you that the data is in Read-Only format. Click OK.

Figure 4.4.4.6: Step 6
7. The file will be opened as it can be seen below. The opening of the file would take one to two minutes over the Remote Desktop.

Figure 4.4.4.7: Step 7
8. We can start extracting the data now. Go to Tools → Event Traces.

Figure 4.4.4.8: Step 8
9. You will see the Event Traces menu as it can be seen below.

Figure 4.4.4.9: Step 9
10. Select all of the Event Traces in the menu. And then click Export.

Figure 4.4.10: Step 10
11. Select the location on the hard-drive where you would like to place the extracted data.

You can change the file name here if you want to. After the location and the file name is set, click Save.

Figure 4.4.4.11: Step 11
12. After a few seconds, the extracted file will appear in text format as it can be seen below.

Check to make sure that file is extracted without defects.

Figure 4.4.4.12: Step 12
13. Step 12 concludes the process of extracting one file. You can go back to Step 2 for another file or close the Rembrandt application. Then Rembrandt will ask you if you would like to save changes to the recording. Click No.

Figure 4.4.4.13: Step 13
Chapter 5

Results

This chapter will present results of the project in terms of our goals, namely the societal aspects, technical aspect and the extraction of sleep data from the sleep data repository maintained by KDDRG group.

5.1 Societal Aspects of Sleep

As a result of the project, I learned the important impact of sleep on the daily life, and the detrimental consequences of the sleep disorders on the performance and the life of human beings.

It is a given fact that we, as human beings, know that sleep is an essential metabolic process that cannot be missed or made up. However, people are giving less and less attention to sleep due to the demands of the modern society.

In summary, analyzing the societal aspects of human sleep in the modern society has given me insights on the human sleep that I would not have been aware of otherwise.

5.2 Technical Aspects of Sleep

On the other hand, on the technical side of the project, I learned about the on-going research on sleep in the fields of Computer Science, Psychology and Medicine. In addition, I learned about how the sleep data is quantified via the physiological electrical signals, how the
process of sleep is partitioned into different stages by interpreting the features of those electrical signals, and how those rules are developed and updated over the time.

This portion of the project has made contributions to the on-going research at KDDRG group in terms of updating the knowledge of the sleep data representation.

### 5.3 Sleep Data Extraction

This is the portion of the project where the major contributions of the project comes along. For this part of the project, I researched on the information present in the PSG sleep study EDF files in joint efforts with Mr. Khasawneh. Then, we identified that the data desired for his algorithmic development research is the sleep stage sequence data.

Afterwards, I have looked into the software tools that will enable us to extract the sleep stage sequence data out of the EDF files within our constraints. In our joint effort again, Mr. Khasawneh and I identified that Rembrandt software is the best candidate available for the data extraction we are looking for.

Finally, using the Rembrandt software, I have successfully extracted the sleep stage sequencing data as per according to the requirements.
Chapter 6

Conclusion and Future Work

This chapter will present the concluding remarks for the project as a whole, and state the future work that could evolve from the results of this project.

6.1 Conclusion

As stated in the previous chapter, as of the end of the project, I have successfully researched and documented societal aspects of human sleep, physiological facts about the process of sleep, on-going sleep study research in the fields of Computer Science, Medicine and Psychology, and technical aspects of contemporary sleep research.

As outputs of the project, we have come to better understanding the societal and physiological facts about human sleep, and updating our knowledge on the technical aspects of contemporary sleep study research.

In addition, as part of the project, I have successfully downloaded, and extracted the sleep study data of 1046 patients as contained in KDDRG sleep data repository using Rembrandt sleep study software. This process has led the KDDRG group as well as me to better understand the sleep PSG study files as well as the Rembrandt software.

Overall, the project has met its goal of studying the societal and physiological aspects of human sleep, updating the concurrent technical aspects of human sleep research, and finally identified and extracted a significant number of sleep study data from the sleep repository we have access to.
6.2 Future Work

As future work, KDDRG group will continue working on the development of data mining approach for sleep disorder diagnosis via future graduate theses, Major Qualifying Projects, and Interactive Qualifying Projects utilizing the data from 1046 patients I have downloaded and extracted.

Possible immediate future works include utilizing the data extracted to test out novel data mining and algorithmic techniques to diagnose sleep disorders.
Bibliography


