Measuring College Students' Sleep, Stress, Mental Health and Wellbeing with Wearable Sensors and Mobile Phones

by

Akane Sano

B.Eng., Keio University (2003)

M.Eng., Keio University (2005)

Submitted to the Program in Media Arts and Sciences,

School of Architecture and Planning,

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

Signature redacted

Author _____

Program in Media Arts and Sciences

October 30, 2015

Signature redacted

Certified by _____

Rosalind W. Picard

Professor of Media Arts and Sciences

Director of the Affective Computing Group

Massachusetts Institute of Technology

Thesis supervisor

Signature redacted

Accepted by _____

Pattie Maes

Academic Head

Program in Media Arts and Sciences

Massachusetts Institute of Technology

Measuring College Students' Sleep, Stress, Mental Health and Wellbeing with Wearable Sensors and Mobile Phones

by

Akane Sano

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning,
On October 30, 2015, in partial fulfillment of the
Requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

This thesis carries out a series of studies and develops a methodology and tools to measure and analyze ambulatory physiological, behavioral and social data from wearable sensors and mobile phones with trait data such as personality, for learning about behaviors and traits that impact human health and wellbeing. This thesis also validates the methodology and tools on a selected subset of the questions that can be answered by the data collected.

First, I conducted a study to characterize wrist electrodermal activity (EDA) patterns with concurrent polysomnography and conventional palm EDA measurement. I developed a tool to analyze the EDA data quantitatively and found that wrist EDA peaks occur during Non REM2 and 3 sleep. Then, with multi-modal wearable sensor data, I conducted several studies showing how multi-modal wearable sensors can improve characterization of sleep/wake states over motion-sensing alone, and predict sleep-related memory consolidation. We found that wrist-EDA helps discriminate when there is improved sleep-related memory consolidation.

Next, with colleagues at MIT and Brigham and Women's hospital, I designed and carried out the first four semesters of the "SNAPSHOT study", which measured over 100,000 hours of multi-sensor and smartphone use data from 168 college students, recruited together with their social groups. Each student contributed intensive multi-modal ambulatory data (physiological, behavioral, environmental, and social) for 30 days. Each student also filled out standardized questionnaires on mental health, personality, stress, social interactions, sleep and GPA, and provided a measure of dim light melatonin, enabling circadian phase to be measured.

To investigate the value of the data, I examined a subset of the large set of questions that these new data enable us to answer: I examined the associations between sleep regularity and sleep duration on academic performance, physical/mental health, perceived stress and wellbeing-related measures using coarsened exact matching to control covariates. Our data showed that sleep irregularity was statistically significantly more associated with bad health, reported in the morning, and with worse mental health than sleep duration. I also identified features useful for recognition of monthly reported perceived stress (high vs low): daily activities, personality, sleep, physiology, social interactions, phone usage, and mobility.

Thesis Supervisor: Rosalind W. Picard
Title: Professor of Media Arts and Sciences, Program in Media Arts and Sciences, MIT

Measuring College Students' Sleep, Stress, Mental Health and Wellbeing with Wearable Sensors and Mobile Phones

By

Akane Sano

Signature redacted

Reader: _____

Rosalind W. Picard
Professor of Media Arts and Sciences
Director of the Affective Computing Group
Massachusetts Institute of Technology
Thesis supervisor

Signature redacted

Reader: _____

César A. Hidalgo
Associate Professor of Media Arts and Sciences
Director of the Macro Connections group
Massachusetts Institute of Technology, Media Lab

Signature redacted

Reader: _____

Charles A. Czeisler
Professor of Sleep Medicine, Harvard Medical School
Baldino Professor of Sleep Medicine and Director
Division of Sleep Medicine, Harvard Medical School

Acknowledgement

I am very happy that I spent the last 5 years at the Media Lab and have had freedom to explore several different research areas and obtained resources and funding to work on what I wanted to work on, long-term multi-modal ambulatory monitoring in "relatively" healthy populations for sleep and mental health fields. This 5 years passed very fast for me, giving me busy but happy and fruitful days. A lot of warm support and help from people around me has enabled me to finish my PhD as we confirmed the importance of positive social interactions in our SNAPSHOT study.

I would like to deeply thank my supervisor, Prof. Rosalind Picard for her warm support, cheer, patience and advice. I spent extremely happy for 5 years to work with you at the Affective Computing Group. Thank you so much for giving me an opportunity to join the group and letting me explore the field in sleep, mood, stress and long-term large-scale ambulatory multi-modal monitoring in relatively healthy populations.

I would like to also thank my committee members Prof. Charles Czeisler for providing me with insightful feedback and Prof. Cesar Hidalgo for giving me valuable advice on designing social network surveys and analyzing observational data.

The SNAPSHOT/College Sleep study could not have been done without help and support from a lot of people for designing the study, updating IRB protocols, running the actual studies and cleaning and analyzing the data. I greatly thank the SNAPSHOT/College Sleep project members: Brigham and Women's Hospital/Harvard Medical School team, Prof. Charles Czeisler, Prof. Elizabeth Klerman, Dr. Andrew Philips, Dr. Andrew McHill, Conor O'Brien, Justin Buie, Natalie Viyaran, Salim Qadri, Michael Shreeve, Omer Zaidi and Dr. Laura Barger, and MIT team members, Prof. Rosalind Picard, Prof. Cesar Hidalgo, Amy Yu, Sara Taylor, Yuna Hahn, Ian MacFalane, Marisa Rozzi Gia Min, Laura Breiman, Daniel Smilkov and Jade Philippom. I feel really grateful to have a chance to collaborate with the BWH team, write a NIH grant together which was fortunately accepted, and finally start running the actual studies and finding very interesting results. Especially, I would like to thank Beth for giving us precious and careful feedback on study design and papers. I am grateful to Conor, Justin and Salim for their tremendous support in running studies every semester together. I would like to thank Dr. Wei Wang for her advice on statistical analysis. I would like to express my gratitude to Sara Taylor for running this project together, helping and supporting me a lot, letting me focus on the analysis and on writing my thesis in the past half year and for suggesting ideas to make our project more effective and efficient and making them happen. I am very excited to continue working on this project with this great team!

I also thank Prof. Robert Stickgold and Hilary Wang for sharing PSG data with wrist EDA and giving me insightful advice on our wrist EDA journal paper.

I deeply thank members at Affective Computing Group for fun time every day and help and support: Yadid Ayzenberg, Vincent Chen, Karthik Dinakar, Micah Eckhardt, Szymon Fedor, Asma Ghandeharioun, Javier Hernandez, Natasha Jaques, Kristina Johnson, Grace Leslie, Daniel Lopez Martinez, Ehimwenma Nosakhare, Sara Taylor, Rob Morris, Daniel McDuff, Elliott Hedman, Joy Chen, Ehsan Hoque, Rana el Kaliouby, Rich Fletcher, Matthew Goodwin, Ming-Zher Poh, Sharon Tam, Jackie Lee, Seoungjae Yoo, Rashmi Karki, Kristina Bonikowski, Lillian Lai, and Alex Khitrik. I would like to thank especially, Javier Hernandez for being a great peer in the same group for the past 5 years and fighting in the thesis boot camp in the same office with the same schedule this spring and summer.

I also thank UROP students who worked with me, Anuhya Ghorakavi, Xinyu Liang, Alyssa H Smith, Tania W Yu, Shirley Chen, Barbara Duckworth, Sumit Gogia, George Ezenna, and Weihao Gao.

I also thank Mary Czerwinski, Paul Johns and Prof. Gloria Mark for providing me with a chance to work on the HealthSense project at Microsoft Research, which has allowed me to work on stress, productivity, diet, exercise and sleep on employee populations.

Thanks to MIT and Media Lab community for giving me a lot of stimulation, inspiration and fun time. Thank you, Linda, Keira, Mitch, Pattie and Joi for your support.

I would like to thank MIT Media Lab consortium, National Institute of Health, Samsung Electronics, Rotary Foundation and Japan Student Services Organization for financial support that made our projects possible and enabled me to stay in the US.

I deeply thank my friends in Boston, California, Seattle, Japan and the other side of the world for their friendship and support. Especially thank you, Blue, Wen, Aithne, Amy, Haritha, Selene, Jean, Naoya & Rie, Masayoshi, Takashi, Koki, Hayan, Mindy, Paul, Jonas, Ian, Kourosh, Vahid, and Yuki.

Lastly, I would like to give the deepest thank you to my Dad, Tetsuo Sano who encouraged me to come to the U.S. 5 years ago when I was thinking about whether or not to come and supported me from a long distance for this five year.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Motivation and Thesis Aims

Recently, many wearable devices have allowed us to monitor our daily lives: for example, a lot of wrist wearable devices have been commercialized and allowed us to collect activity and physiological data (e.g. pedometer or activity/sleep tracker using accelerators, heart rate from plethysmography, skin conductance, blood pressure, blood sugar level). Smart phones are capable of monitoring location, activity, social interaction over calls and SMS and environmental data such as ambient light exposure and humidity. How can we leverage 24/7 rich data from wearable devices? What if accumulated data could provide meaningful information that will make you aware of your health condition and what behaviors you should change to improve your health and wellbeing? Sleep, stress and mental health have been major health issues in modern society. What if we can figure out which factors influence your bad sleep or stress problem? What if we can predict your health condition based on your behaviors in the past 10 days? Could early warnings be given to change behaviors and predict conditions such as depression? People could be more aware of their health condition, think about how they could be healthier or solve their health problems and make actions to be healthier.

Much research has been done to compare healthy populations and patient populations or healthy populations under control tasks vs condition tasks (sitting down vs adding stress); however, even within healthy populations, we have variations: some people tend to feel more stressed or to be poor sleepers, and some people could feel down. In order to prevent being in a bad health condition or sliding from a good to

a bad health condition, we need to observe variations in the uncontrolled real world and understand which factors influence these variations, and examine if we can detect signs indicating a slide from a good health condition to a bad condition in the early stages. The cause of poor sleep habits and high stress as well as reactions to stressors and sleep habits can depend on many factors: internal factors are personality types and physiological factors and external factors are behaviors and environmental and social factors. We need to determine how we can collect and analyze these data and take advantage of them to keep people healthier.

The main goal of this thesis is to develop a reproducible methodology to capture ambulatory data on physiological and behavioral characteristics of human subjects using sensors installed in mobile phones and wearable devices, to collect data using this methodology, and to show that the data can be used to answer important health questions. To achieve this goal, I developed a suite of software tools that monitor and analyze the behavioral and physiological traits of individuals, and I deployed these tools in a cohort study of sleep among healthy college students, which we call the SNAPSHOT Study. The SNAPSHOT study seeks to measure **S**leep, **N**etworks, **A**ffect, **P**erformance, **S**tress, and **H**ealth using **O**bjective **T**echniques. The goal of developing this data collection methodology is to provide a means to study the relationship between the physiological, behavioral, and psychological characteristics of healthy individuals in an ambulatory setting, instead of in a controlled laboratory setting. I led the design of the SNAPSHOT study to collect data mainly to understand the associations between sleep, stress, mental health and wellbeing-related measures. Here, I validate the data collection methodology I developed by showing that the data collected in the study can be used to answer some of the important questions related to sleep, stress, and mental health. In particular, we show the influence of sleep duration and regularity on mental health and other health measures and physiological, behavioral and trait markers for stress. These results indicate that our methodology is able to generate the data needed to identify physiological, behavioral, and trait markers in an ambulatory setting and could be used to study a variety of conditions that are known to depend on social and environmental factors, such as sleep, stress and mental health.

In this thesis, first, I introduce the analysis of wrist electrodermal activity (EDA) during sleep. I describe a software tool I developed and show how it can be used to understand the properties of EDA patterns relative to sleep stages and relative to conventional palmar EDA. Then, I describe two studies conducted to investigate how wearable sensor features contribute to sleep/wake classification and to sleep-dependent memory consolidation, and findings we obtained with both studies. Lastly, I describe the SNAPSHOT Study, where we extended our measurement to a multi-modal, 24/7, long-term and large scale design to measure sleep, stress and mental health in MIT undergraduate students' daily life. With colleagues at MIT and Brigham and Women's Hospital, I have developed tools and designed and executed large-scale and long-term studies to collect intensive multi-modal and long-term data in daily lives using wearable sensors and mobile phones as well as some data that could be labels or biomarkers measured in the

laboratory. We started the SNAPSHOT study (previously we called College Sleep study) in fall 2013 and have collected 168 participants' data from MIT undergraduate students who are extremely busy and at high stress and sleep deprivation. We have recruited 50 MIT undergraduates each semester, making an unprecedented number of measures from each student for 30 days each: We have collected continuous physiology, behavior and social data 24/7 (120960 hours), including electrodermal activity, skin temperature, 3 axis acceleration and light exposure from wrist devices, and over 5000 day phone calls, SMS logs, location, application usage, and screen on/off logs from mobile phones. We have also collected questionnaire measurements twice a day and pre- and post-study to quantify daily diary items (exercise, academics and sleep schedule, caffeine, alcohol and drug intake, subjective daily measures about health, alertness, stress, mood and social interaction), and to provide standardized survey scores about sleep, stress, anxiety, and personality types. Participants also spend one night in the hospital for dim-light melatonin assessment of circadian phase.

Lastly, I describe the validation of the collected data by showing two kinds of analysis (1) the influence of sleep duration and sleep regularity on academic performance, stress level, physical/mental health and other health measures and (2) automated recognition of monthly perceived stress using multi-modal data.

This thesis describes how we developed tools, designed a series of studies and collected data using wearable sensors and mobile phones combined with laboratory measurement to find associations and predictors related to sleep, stress, and mental health in an ambulatory setting.

1.2 Thesis Contributions

This thesis provides contributions in:

- Designing and running multi-modal large-scale long-term studies (MIT SNAPSHOT Study) with wearable devices and surveys to measure sleep, stress and mental health data in daily life
- Developing software and methods to clean up and analyze ambulatory physiological and behavioral data (EDA, skin temperature and acceleration) data from wrists and mobile phone data (call, SMS, location, screen on/off timing logs)
- Characterizing EDA during sleep measured with dry wrist-worn electrodes and comparing it to palmar EDA, and characterizing its patterns relative to sleep stages
- Applying multi-modal wearable measurement to sleep/wake recognition and memory consolidation and finding related features in wearable measurement

- Providing a new and rich dataset to allow us to solve many research questions
- Validating the SNAPSHOT Study methodology by analyzing the data and comparing the results with the results previously reported
- Answering the following research questions with the multi-modal wearable data and tools:
    - What are characteristics of wrist EDA during sleep? (Chapter 3)
    - Can we recognize sleep dependent memory consolidation using multi-modal wrist sensor data? (Chapter 4)
    - Can multi-modal wrist sensor data classify sleep/wake epochs better than actigraphy? (Chapter 4)
    - What are characteristics in physiological and behavioral patterns from MIT undergraduate students in 30 days of daily-life data? (Chapter 6)
    - How do sleep duration and regularity influence academic performance, physical/mental health, stress level and other health measures? (Chapter 7)
    - How accurately can we recognize perceived stress and which features from which modality work better? (Chapter 7)

1.3 Thesis Outline

The outline of this thesis is the following:

Chapter 2 Background
This chapter describes background and related work about this thesis: sleep, stress, mental health, social interaction, and mobile and wearable sensing.

Chapter 3 Quantitative Analysis of Wrist Electrodermal Activity during Sleep
This chapter describes methods we developed to quantitatively characterize wrist sleep EDA measured on the wrist during sleep and the comparison of wrist sleep EDA with sleep stage and palmar EDA.

Chapter 4 Multi-modal Wearable Data Analysis
This chapter constructs new methods for human sleep dependent memory consolidation and sleep/wake states using multi-modal wearable sensor data.

Chapter 5 SNAPSHOT Study: Design and Measurement

This chapter describes the design and execution of the SNAPSHOT study. The precise measures we use are all presented here.

Chapter 6 SNAPSHOT Study: Data Cleaning, Pre-processing, Feature Extraction and Data Characteristics

This chapter describes methods of cleaning and pre-processing ambulatory data, extracting features and characteristics of our dataset.

Chapter 7 SNAPSHOT Study: Data Validation, Analysis and Results

This chapter describes specific questions we answer to validate the SNAPSHOT Study methodology, the analysis and the findings.

Chapter 8 Discussions, Limitations, and Contributions

This chapter discusses the main findings, summarizes the contributions, and lists several future direction.

Chapter 2
Background

In this chapter, we introduce background about sleep, circadian rhythm, stress, and mood and mental health: why they are important to study, and measurement and relationships among these that relate to the work of this dissertation. Then we also introduce methods and wearable and mobile devices to measure sleep, stress and mental health and previous studies to measure and recognize them.

2.1 Sleep and Circadian Rhythm

This chapter describes our main target in our study, sleep and summarizes the importance of sleep, measurement of sleep and circadian rhythm, the impact of sleep deprivation on learning and memory, and mood and emotion, and the influence of electric devices on sleep.

2.1.1 Sleep Behaviors and Health

Short sleep duration is highly prevalent in the United States, with approximately 30% of the adult population sleeping less than 7 hours per night [Marcelli, 2009]. One recent study reported that 5 healthy participants slept for 7.2 hours on average under 2 week prehistoric living conditions where the participants in the study had no access to electricity and any modern conveniences, such as mobile phones or newspapers [Piosczyk, 2014]. Numerous studies indicate that less than 7 hour sleep is insufficient for optimal health and cognitive function, recommending approximately 8 hours of sleep per night for the average young

healthy adult [Van Dongen, 2003]. Sleep is critical to a wide range of biological functions; inadequate sleep results in impaired cognitive performance [Van Dongen, 2003][Cohen, 2010], academic performance [Wolfson, 2003] and mood [Moturu, 2011], and adverse health outcomes including obesity [Gupta, 2002] [Weiss, 2010], diabetes [Zizi, 2010], and cardiovascular disease[Malhotra, 2009].

These studies were mostly done in laboratory settings or using questionnaires. Laboratory studies have advantages in collecting cleaner data in control settings; however, they have also disadvantages in not being able to measure natural sleep. Studies using questionnaires can capture data from a larger number of participants; however, the measurement might be subjective and might not be accurate. Some studies pointed out the difference between self-reported sleep patterns in sleep diaries and wrist actigraphy measurement [Martin, 2011]. Our study has captured both subjective and objective data in the real world to measure natural sleep. Also, most of these studies focused on a few days of data collection and a relatively small number of individuals; however, our work makes new advance not only in quantity of participants and nights, but also in richness of gathering and analyzing continuous behavioral, physiological, and social network data from a month of real-world college student life.

2.1.2 Circadian Rhythms and Sleep Homeostat

Sleep behaviors are affected by biological, social and environmental factors. Biologically, they are affected by two main factors: circadian rhythm and sleep homeostat. Circadian rhythm is an approximately 24-h periodic drive to sleep that can be synchronized by the daily light/dark cycle [Dijk, 2002]. The sleep homeostat is a biological drive to sleep that increases during wake and decreases during sleep [Daan, 1984]. The circadian and homeostatic drives act on a sleep/wake switch present in the hypothalamus [Saper, 2005].

The suprachiasmatic nucleus (SCN) in the hypothalamus in mammals is the master circadian pacemaker [Ralph, 1990]. The timing ("phase") of circadian rhythms is synchronized to local time by ocular light exposure. Inappropriately timed light exposure, such as light during the biological time for sleep, may adversely affect both the amount of sleep at night and alertness during the day.

2.1.3 Measurement of Circadian Rhythm and Sleep

To measure circadian phase, a constant routine protocol in a laboratory has been used in previous studies. Participants stay up with fixed posture on a bed under dim light (< 20 lux), and circadian phase is measured with core body temperature or melatonin from saliva or blood [Duffy, 2002] (Figure 2.1). Participants are kept in a temporally and environmentally isolated condition under the dim light during wake episodes so

as not to entrain the circadian rhythm. Time course of melatonin rhythm is shown in Figure 2.2. Melatonin secretion usually starts 6-7 hours before habitual bedtime with sleepiness and distal temperature. Melatonin secretion is suppressed by light. Light is a major entrainment in circadian rhythm. The entrainment happens depending on when light is applied. The relationship between the timing of light exposure and the light induced circadian phase shift is called a "phase response curve (PRC)" [Khalsa, 2003]. Figure 2.3 shows the PRC to the bright light stimulus using melatonin midpoints as the circadian phase marker and this curve explains how much phase advance/delay occurs when light is applied.

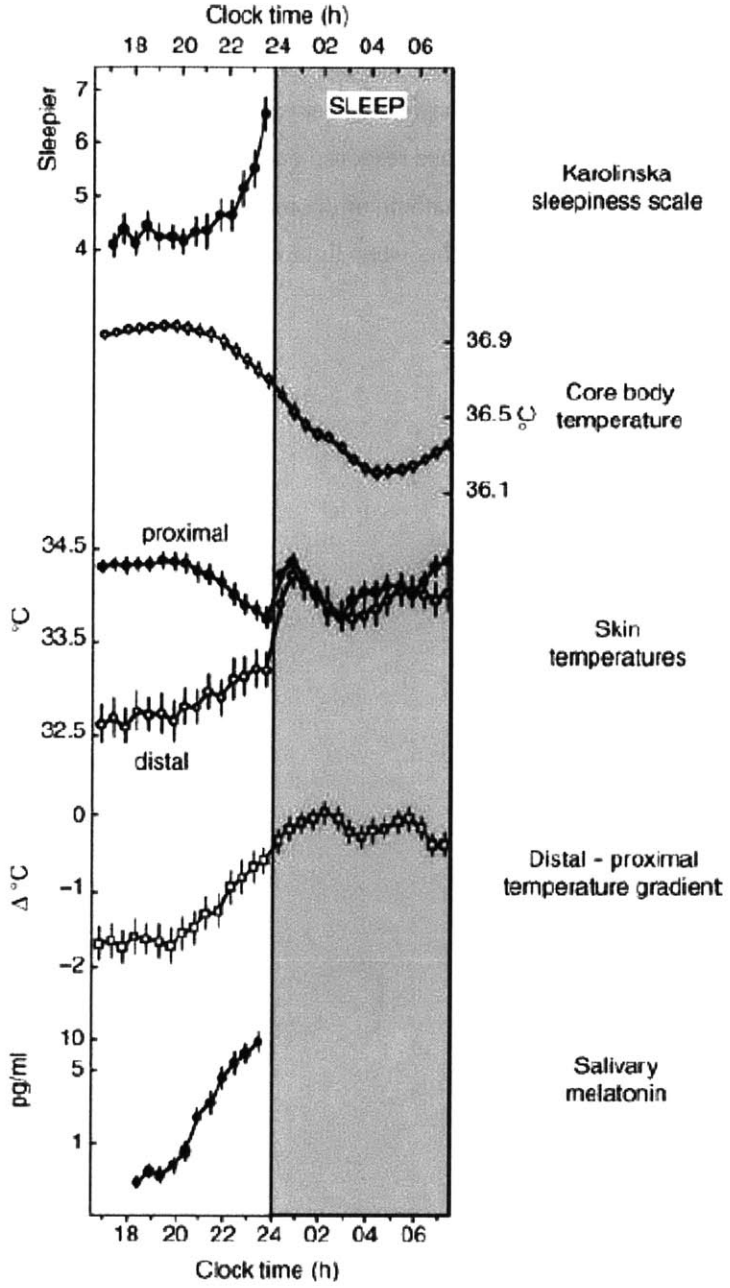


Figure 2.1 Time course of sleepiness scale, core body temperature, skin temperature, distal-proximal temperature gradient and salivary melatonin in constant routine [Cajochen, 2003].

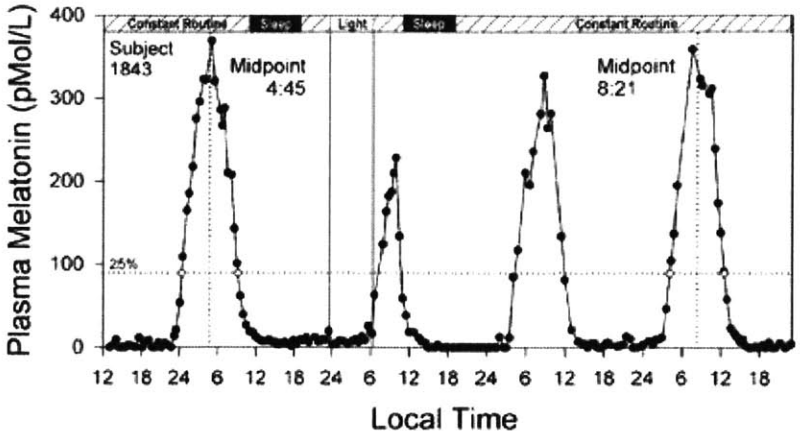


Figure 2.2 Time course of melatonin in constant routine and light exposure [Khalsa, 2003].

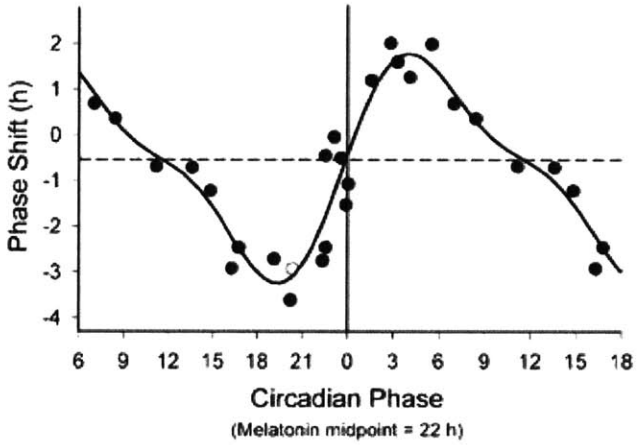


Figure 2.3 The PRC to the bright light stimulus using melatonin midpoints as the circadian phase marker [Khalsa, 2003].

To measure sleep clinically, polysomnography (PSG) has been used to monitor sleep and identify sleep disorders in sleep labs as a gold standard; however, it has disadvantages requiring the patient to stay one or more nights in the lab wearing uncomfortable sensors and wires. Actigraphy has been used to monitor long-term sleep wake cycles outside the laboratory. Cole et al. showed that sleep and wake are classified with an accuracy of 88% using wrist-worn actigraphy and regression analysis comparing the wrist data to PSG [Cole, 1992]. Some other researchers have applied machine learning or new algorithms to improve the accuracy [Sadeh, 1994][Pollak, 2001] or used other data (heart rate variability from electrocardiogram (ECG) or respiration) [Lewicke, 2008][Long, 2014].

In our study, we measured circadian rhythm through sampling melatonin in saliva and through gathering sleep/wake patterns with a wrist sensor and with self-reported electronic sleep diaries (more details will be in section 5.2 Data Collection).

2.1.4 Electric Device Use and Sleep Behaviors

One of the most important factors in influencing sleep behaviors among children, adolescents, and young adults is the growing use of electronic and social media [Shochat, 2012]. Previous survey based studies showed that social and electric media such as TV, computer games, and music before sleep can result in difficulties sleeping, making young adults sleep less and be more tired [Eggermont, 2006][Garrison, 2011] [Paavonen, 2006] [Shochat, 2010] [Sugawara, 2007], or simply spend less time in bed [van den Bulck, 2004]. In addition, excessive Internet and mobile phone use can cause anxiety and insomnia [Jenaro, 2007], and in the long term these effects can potentially lead to depression [Thomée, 2011].

Many studies have examined phase response curve and represented that electric light advances the circadian rhythm after circadian phase 0 hour and delays it before circadian phase 0 hour using a constant routine protocol [Khalsa, 2003]. Further, a laboratory study showed that exposure to a light-emitting screen with short-wave length light in the late evening suppressed melatonin secretion, increasing subjective alertness as well as objective alertness measured as increased eye movement and low frequency brain wave activity [Cajochen, 2011]. Another study showed that reading an electric book right before sleep prolonged sleep latency, increased alertness, suppressed a melatonin level and reduced alertness next morning in comparison to reading a paper book [Chang, 2014].

Our SNAPSHOT study is designed to collect data to investigate the relationship among electric device use (mobile phone/email monitoring), light exposure (light sensor data) and subjective and objective sleep and health measures and academic performance in extended daily life (section 5.2 Data Collection).

2.1.5 Impact of Acute Sleep Deprivation and Chronic Sleep Restriction

Several studies have examined the time course for the impact of acute sleep deprivation and chronic sleep restriction on sleepiness and on neurobehavioral performance.

One study [Van Dongen, 2003] examined how acute sleep deprivation and chronic sleep restriction influence sleepiness and neurobehavioral performance using dose-response experiments. Both chronic and acute sleep deprivation show increased sleepiness and neurobehavioral deficits, but impacts of the two types of sleep deprivation are different. Comparison of chronic sleep restriction to total sleep deprivation showed

that the latter resulted in extremely large neurobehavioral deficits in performance in psychomotor vigilance task (PVT) and sleep delta power responses relative to how much sleep was lost. Responses of chronic sleep restriction were cumulative in cognitive performance. Subjective sleepiness ratings showed an acute response to sleep restriction but did not significantly differentiate the 6 hour and 4 hour conditions. Even moderate sleep restriction (7 hour sleep) results in neurobehavioral deficit; however, the two types of sleep deprivation have also different recovery processes: acute sleep deprivation shows larger impacts than chronic sleep restriction and acute sleep loss takes less time to recover than chronic sleep loss [Belenky, 2003].

Learning and Memory

Many researchers have shown how sleep is related to learning and memory using sleep deprivation or sleep restriction. For example, performance in a visual discrimination task [Stickgold, 2000] significantly improves after sleep, but doesn't after sleep deprivation and even after the following 2 nights of recovery sleep. The study also showed that improvement in the task is proportional to the percentage of SWS in the first quarter of the night and the REM in the last quarter of the night, given sleep longer than 6 hours. Another study [Maquet, 2003] showed that visuo-motor skill performance was improved and the brain activity increased in the superior temporal sulcus after post-training sleep compared to post-training sleep deprivation.

Mood and Emotion

Some studies have shown that sleep loss could degrade mood and lead to depression. Medical interns who exposed to sleep deprivation showed higher score in depression compared to their baseline [Rosen, 2006]. Epidemiologic studies also showed the relationship between sleep deprivation and psychiatric disorders. People who suffer from insomnia have high percentages of mood disorders [Ford, 1989] and also long sleep onset and frequent night awakenings are major symptom of depression [Taylor, 2005]. On the other hand, total sleep deprivation improves depressed mood for short term till recovery sleep in 60% of patients diagnosed with affective disorders [Wirz-Justice, 1999]. This can be considered because sleep deprivation makes sleep phase on the next day advanced. In most depression patients, sleep-wake cycle is delayed to other circadian rhythm in their body and delayed to sleep-wake cycle in healthy people.

Moturu et al. investigated the association among sociability, sleep quality and mood with self-reported surveys and mobile phone proximity data from 54 participants for one month [Moturu, 2011]. Based on self-reported data from participants, they found that the good mood group had longer sleep duration (average ~7 hours) than the poor mood group (6.4 hours) and lower sociability was related to poorer mood. Another study showed the relationship between fatigue, mood and sleep need (difference between preferred sleep length and actual self-reported sleep length) in three different populations, school children, students and employees [Oginska, 2006]. These studies examined the relationship between self-reported sleep length and mood or sleep deprivation and mood; however, their sleep and mood measurement were all based on self-reported data only.

Jet lag and shift work can also degrade mood. In those cases, after an abrupt time-shift, desynchronization of the biological clock to a new time-zone or the external clock and sleep deprivation occur. One study showed that depression increased in abrupt phase delay [David, 1991]. In another study about bipolar patients [Frank, 2007], maniac episodes could be triggered by time-zone changes and by sleep deprivation. The study also showed that regularity of sleep, circadian rhythm and life style could regulate the symptoms of bipolar disorders.

Some studies have shown that taking melatonin after time shift led to faster recovery of mood degradation. One study [Petrie, 1993] showed that oral melatonin led to faster recovery in jet lag, self-reported mood, energy and alertness. Another study done for shift workers [Folkard, 1993] showed that melatonin given at bedtime increased alertness at wake time. Melatonin helps to adjust the biological rhythm to a new time zone or time shift and in improved mood. This also implies that desynchronized biological rhythm and sleep deprivation in jet lag and shift work could lead to degradation of mood, alertness and energy.

Next, we describe other targets of our study, stress, mood and mental health, together with their definitions, the importance of measuring them, and previous studies to measure or recognize them

2.2 Stress

Stress is one of the major problems in modern society. Sometimes people are aware of being under stress, for example, when they are occupied with deadlines of homework and projects; however, long-term conditions with high stress can be chronic and people may be less likely to notice whether they are under high stress, or they may be generally less sensitive to stressors. Stress detection technology could help

people better understand and relieve stress by increasing their awareness of heightened levels of stress that would otherwise go undetected.

Several technologies have been developed to measure or recognize stress level; some methods are based on physiological signals: blood pressure (BP) [Vrijkotte, 2000], heart rate (HR) [Vrijkotte, 2000], heart rate variability (HRV) [Dishman, 2000], skin conductance (SC) [Hernandez, 2011][Setz, 2010], cortisol [Dickerson, 2004][van Eck, 1996], pupil diameter [Mokhayeri, 2011]. Activity of the sympathetic and para-sympathetic nervous systems can be monitored through SC, BP, HR and HRV. Here, we describe the SC we used in our study. SC has been considered as a biomarker for stress [Boucsein, 1992], where eccrine sweat activity that is controlled by only sympathetic nervous activity is measured, For example, Healey et al. measured SC, HR, HRV, respiration and electromyogram to recognize stress levels in Boston drivers, finding that SC was the most accurate predictor of stress [Healey, 2005]. Hernandez et al. discriminated stressful and non-stressful calls at the call center environment using SC features. Setz et al. automatically classified SC responses from cognitive load and stress with accuracy higher than 80%. Kusserow et al. measured public speaker, an on-stage musician, an Olympic ski jumper, and people during everyday life, quantifying stress arousal using SC and other signals such as HR, HRV, and acceleration [Kusserow, 2013].

Other methods are based on surveys. For example, the Holmes and Rahe Stress Scale counts up events in the prior year that could lead to stress [Holmes, 1967]. Perceived stress has been used as a self-reported stress marker [Cohen, 1983]. Questions in the perceived stress scale (PSS) assess what degree in each situation a subject feels stressful.

In our study, we used the PSS-10 (surveys with 10 questions) to evaluate stress level in the past one month at the pre and post study, and a daily stressed-calm scale (0-100) (twice per day). We combined long-term physiological, social and environmental objective data with personality types and subjective data to understand not only spontaneous stress but also long-term stress reactions (section 5.2 Data Collection).

2.3 Mood and Mental Health

Mood is defined as "the appropriate designation for affective states that are about nothing specific or about everything-about the world in general" [Frijda, 2009]. Unlike emotions, which follow their eliciting stimuli closely or even instantaneously, a mood is usually temporally remote [Morris, 1992] from its cause (e.g., a person can wake up in a bad mood in the morning as a result of a confrontation the previous evening). Consequently, the cause of a mood may not always be easy to identify.

To measure mood, the Profile of Mood States (POMS) [McNair, 1989] has been used frequently in psychology. The questionnaire consists of 65 emotion adjectives and each of them is rated with a five point scale. POMS evaluates six different mood states: tension, depression, anger, vigor, fatigue, and confusion. The Positive and Negative Affect Schedule (PANAS) [Watson, 1988] has been also used often. PANAS consists of 20 adjectives and each of them is rated with a five point scale. In contrast to the POMS, the PANAS was developed not to assess distinct mood states but rather to assess positive affect and negative affect.

Positive mood is linked to good health including good performance, cognition, and memory [Chepenik, 2007] [Nadler, 2010]. These results are based on studies with either depressed people or with healthy people undergoing emotional induction or sleep deprivation in a laboratory. In the studies, a visual analogue scale or PANAS was used to assess mood. We conducted a study to examine sleep and behavior influences on mood for college students in their daily lives at home and work. New technology makes it possible to objectively measure sleep and other behaviors and examine their association with daily emotional ups and downs. Since the surveys described above have many items, we used daily Sad-Happy scale (0-100) to evaluate sadness and happiness twice per day in our study (section 5.2 Data Collection).

Mental Health includes emotionally, psychologically and socially positive conditions [MentalHealth.gov]. Questionnaires have been used to evaluate mental health. Here, we introduce three questionnaires, which are commonly used in previous studies. The Patient Health Questionnaire (PHQ-9) is one of the most commonly used surveys [Kroenke, 2001]. The PHQ-9 consists of nine diagnostic items of DSM-IV: anhedonia, depressed mood, trouble sleeping, feeling tired, change in appetite, guilt or worthlessness, trouble concentrating, feeling slowed down or restless, and suicidal thoughts. Each item is rated from 0=Not at all through 3=Nearly every day. The final score tells us severity of depression: no depression (0-4), and mild (5-9), moderate (10-14), moderately severe (15-19) and severe depression (20-). The Goldberg Anxiety and Depression Scale (GADS) is an 18-item self-report symptom inventory [Goldberg, 1988]. The GADS score is based on 'yes' or 'no' to nine depression and nine anxiety items about feeling in the past month. The Short-Form 12 (SF-12) is a generic health-related quality-of-life instrument for physical and mental health used in large population health studies [Ware, 1996]. Three items in the SF-12 ask about calmness, energy and feeling down related to depressive and anxiety disorders. In our study, we used SF-12 to evaluate both physical and mental health. We were not able to evaluate depression directly because of constraints it put on our hospital partners and the overnight stay expected there for our participants.

Next, we describe personality traits and social factors that influence our physical and mental health including sleep, stress, mood and mental health

2.4 Personality Types

Several studies have suggested that personality types affect physical and mental health [Janjhua, 2012][Martin, 1996]. The interactions between academic performance, sleep quality, self-reported stress, self-reported mental health and personality categories have been previously characterized using self-reported data. Specifically, academic performance has been correlated with personal traits (conscientiousness, openness and agreeableness) [Poropat, 2009, Noftle, 2007] and sleep parameters have been reported to be influenced by personality traits of neuroticism [Soehner, 2007], extraversion [Killgore, 2007], and agreeableness [Clark, 2007]. Vollrath summarized the relationship between stress and the Big Five Inventory Personality Test categories and identified that neuroticism was a predictor of stress [Vollrath, 2001]. Another study showed negative associations between perceived stress and extroversion, conscientiousness, agreeableness, and openness and positive association between perceived stress and neuroticism [Ebstrup, 2011]. In our study, we also used the Big Five Inventory Personality Test [John, 1999] to evaluate openness, extraversion, conscientiousness, agreeableness, and neuroticism with 46 questions (section 5.2 Data Collection).

Next, we describe what we measure using mobile phones and wearable sensors for our ambulatory studies and summarize previous studies to measure and recognize sleep, stress and mental health using wearable and mobile device data.

2.5 The Role of Social Networks in Mediating Behaviors

Several studies have shown that health-related behaviors can be mediated and spread within a social network. These include behaviors such as sleep [Mednick, 2010], obesity [Christakis, 2007], happiness [Fowler, 2008], depression [Rosenquist, 2011], and loneliness [Cacioppo, 2009]. Furthermore, the spread of these behaviors has been shown to extend beyond the dyad, spreading up to three degrees of separation. Social network analysis has shown that people who self-report being happy are more likely to be located at the center of their social network, where they connect to other happy people, furthering the importance of the social network [Fowler, 2008]. Another analysis showed that poor sleep behavior can spread in social networks from one person up to four degrees of separation. The effects were strongest between tightly

connected individuals, suggesting that the frequency and strength of interaction play important roles [Mednick, 2010].

These studies are based on surveys while our study addresses health-related behavior contagion in the real world by objectively measuring both social network structure from phone call, short message service: SMS and email usage and health-related behaviors. Previous studies using smart phones showed that individual's mood is associated with both individual's sleep and the spouse's mood and individual's sleep is also associated with both individual and the spouse's mood [Moturu, 2011].

2.6 Multi-modal Measurement using Mobile/Wearable Sensors

Recently, many wearable devices have allowed us to monitor our daily personal behaviors, extending measurements that had been done in laboratory or clinical settings before. In this section, we introduce the main ones used in our study.

Measurement from mobile phones

Mobile phones can measure data such as location, distance you travel, social interactions (phone call, SMS and email), application usage, and acceleration and light sensor data. Some smart phones are also equipped with temperature, pressure and humidity sensors. On Android, there are several frameworks to help us to log these data easily. In our study, we used funf framework to collect call, SMS, location, screen on/off timing and application usage (section 5.2 Data Collection).

Acceleration

Acceleration is one of the most commonly implemented measurement in wearable devices and it has been used to track steps and activity levels (e.g. Fitbit, Jawbone) and activity recognition (walking, sitting, climbing stairs etc) [Aggarwal, 2011]. It has also been used to track sleep. In sleep studies, actigraphy uses a wrist-worn accelerometer to identify sleep vs wake patterns [Ancoli-Israel, 2003]. For regular consumers, commercialized devices and smart phone applications track sleep duration, quality and sleep patterns based on tracking movement during sleep; however, it is not clear how accurately these devices track sleep

patterns as each device has different algorithms which have not been disclosed compared to polysomnography, a gold standard used to diagnose sleep disorders with multi-modal physiological signals (electroencephalogram, electrocardiogram, respiration, and electromyogram etc). In our studies, we used a 3-axis accelerometer and an actiwatch (sections 3.2.1 Measurement, 4.1.2 Methods, 4.2.2 Methods and 5.2 Data Collection).

Electrodermal activity (EDA)

Electrodermal activity (EDA) provides a measure of activity in the sympathetic nervous system, one of the main branches of the autonomic nervous system. It has been widely used in psychophysiology studies including for pain [Ledowski, 2007, Storm, 2008], schizophrenia [Ohman, 1981, Schell, 2005], emotion [Kreibig, 2010], epilepsy [Poh, 2012], depression [Ward, 1983] and stress [Healey, 2005]. In laboratory settings, the measurement has been carried out on palms or fingers; however, wearable devices allow us to measure it on the wrist or ankle in ambulatory settings. In this thesis, first we compared wrist sleep EDA with palm sleep EDA and measured EDA on the wrist (sections 3.2.1 Measurement, 4.1.2 Methods, 4.2.2 Methods and 5.2 Data Collection) and we extended the EDA measurement and analysis to day and sleep time to quantify sympathetic activities.

Skin Temperature

Several wearable devices enable us to measure skin temperature on the surface of our body (wrist sensors, and button type sensors). In previous studies, skin temperature has been used to investigate insomnia [Lack, 2008], circadian rhythm [Martinez-Nicolas, 2013] and circadian phase [Kolodyazhniy, 2012]. Furthermore, low rhythmicity in skin temperature in depression patients have been reported [Barbini, 1998]. In this thesis, we measured wrist skin temperature (sections 3.2.1 Measurement, 4.2.2 Methods and 5.2 Data Collection).

Ambient Light

Light is one of the most important factors that regulates sleep and circadian rhythm (See sections 2.1.2 and 2.1.3). Wearable light sensors measure light intensity of photopic and RGB light and they have been used in studies to investigate how light exposure during day and night influence wellbeing and sleep [Sander, 2015] [Harb, 2015]. In this thesis, we measured photopic light intensity on the wrist (section 5.2 Data Collection).

2.7 Understanding Sleep, Stress and Mental Health from Multi-modal Mobile/Wearable Data

This section describes previous research with wearable sensors and/or mobile phone data to measure, recognize or find markers related to sleep, stress and mental health in ambulatory settings. This section helps readers to understand their study design, measurement and data analysis methods.

The "Student Life" study, which monitored 48 college students across a 10 week term at Dartmouth College using Android phones, investigated the relationship between wellbeing measures such as stress, depression, flourishing and loneliness and academic performance and objective mobile phone sensors and usage [Wang, 2014]. Their correlation analysis showed associations between higher conversation frequency (day and night) and longer conversation duration (day), and lower PSS, and longer sleep duration and lower PSS.

To understand sleep, several studies have been done using multi-modal data. A bedside standalone system was developed to capture movement on beds and bedroom environmental data (sound, temperature, and humidity) to give feedback about sleep-related behaviors to users [Kay, 2012]. Another study used mobile phone sensor and usage data (sound, accelerometer, light sensor, proximity sensor, running app, battery, and screen status) from a month of data (N=27) to build models to recognize sleep-wake states and overall and daily sleep quality (good or poor sleepers based on the PSQI score, and daily subjective sleep quality, good or poor quality) [Min, 2014], resulting in 93% accuracy for sleep-wake recognition and 84% and 81%, respectively, in accuracy of overall and daily sleep quality classification.

For recognizing stress levels, several studies have used mobile phones and wearable sensors. Bauer et al. found behavioral modification (the number of places they visited, social interactions, calls, and SMS) in the comparison of stress related situations (during 2 week stressful events and 2 weeks after them, N=7). Bogomolov et al. used multi-modal 7-month data from 120 participants including phone usage, weather and personality type information to classify high/low daily stress levels [Bogomolov, 2014]. They showed 73% accuracy with the combination of 32 features (5 features from Big Five test, 4 features related to weather (e.g. temperature, humidity), 12 call and SMS features (e.g. # of calls and SMS, entropy and reaction time to SMS) and 11 proximity features (e.g. # of Bluetooth IDs and entropy)) and personalized models. We also investigated 5 day high vs low stress recognition (N=18) and 30 day high vs low stress recognition (N=66) using wearable sensor and mobile phone data and showed 75% and 90% respectively [Sano, 2013b] [Sano, 2015]. Muaremi et al also did stress recognition using sleep parameters, and

physiological responses (heart rate, heart rate variability and EDA) (N=10, 19 days), and phone usage and sleep heart rate variability (N=35, 4 months) and showed 61% 3 class stress level classification accuracy with combination of phone usage and HRV features and 73% with sleep duration and upper body posture, or HRV features [Muaremi, 2013, Muaremi, 2014].

For mood and mental health recognition, Moodscope used mobile phones to predict two axes of daily emotion, activeness and pleasure, using multi linear regression models (N=32, 2 months) [LiKamWa, 2013]. They captured phone usage such as email, SMS, call, application usage and mobility and showed 93% accuracy with personalized models. Grunerbl et al. recognized mental states and state changes in bipolar disorder patients using android phone usage (4 different modalities: call, sound, acceleration and location) and showed 76% recognition accuracy with the feature fusion (N=10, 12 weeks) [Grünerbl, 2015]. Our earlier paper developed daily happiness-sad prediction using physiology, mobility and phone data from the SNAPSHOT Study and showed 70% accuracy with features such as skin conductance, acceleration, pre-sleep activity, social interaction, exercise, screen-on duration and time indoors (N=68, 30 days) [Jaques, 2015].

Chapter 3
Quantitative Analysis of Wrist Electrodermal Activity during Sleep

In this chapter, we present quantitative characterization of electrodermal activity (EDA) patterns on the wrists of healthy adults during sleep using dry electrodes. We compare the new results on the wrist to prior findings on palmar or finger EDA by characterizing data measured from 80 nights of sleep consisting of 9 nights of wrist and palm EDA from 9 healthy adults sleeping at home, 56 nights of wrist and palm EDA from one healthy adult sleeping at home, and 15 nights of wrist EDA from 15 healthy adults in a sleep laboratory, with the latter compared to concurrent polysomnography. While high frequency patterns of EDA called "storms" were identified by eye in the 1960's, we systematically compare thresholds for automatically detecting EDA peaks and establish criteria for EDA storms. We found that more than 80% of EDA peaks occurred in non-REM sleep, specifically during slow-wave sleep (SWS) and non-REM stage 2 sleep (NREM2). Also, EDA amplitude is higher in SWS than in other sleep stages. Longer EDA storms were more likely in the first two quarters of sleep and during SWS and NREM2. We also found from the home studies (65 nights) that EDA levels were higher and the skin conductance peaks were larger and more frequent when measured on the wrist than when measured on the palm. These EDA high frequency peaks and high amplitude were sometimes associated with higher skin temperature, but not always. More work is needed looking at neurological and other EDA elicitors in order to elucidate their complete behavior.

3.1 Introduction

EDA is widely used in psychophysiology and provides a measure of activity in the sympathetic nervous system, one of the main branches of the autonomic nervous system. Studies on EDA (also known as galvanic skin response, GSR) during sleep have shown that elevated levels of EDA, with high frequency "storm" patterns are more common during deep, slow wave sleep (SWS) [Koumans, 1968], while the frequency of EDA peaks is lower in the first cycle of the night [Freixa i Baqué, 1983b] (Table 3.1). Classically, EDA has been measured as skin conductance level or skin conductance responses and involves attaching wired and gelled electrodes to the skin, usually on the fingers or palm [Boucsein, 1992; Fowles, 1981]. However, several studies have shown valid measurement of EDA on other locations including the forearm (Table 3.2). Studies using dry electrodes on the forearm have demonstrated reliable long-term measures of EDA [Poh, 2010] and have also led to the discovery of correlations between EDA and significant neurological events measured from EEG [Poh, 2012].

In this study, we used a wireless non-invasive EDA sensor worn as a wristband on the distal forearm, which made it easy for subjects to be monitored in the same manner in the sleep lab and at home. We collected and analyzed 80 nights of EDA data more than ever previously reported in a single study.
Our study makes three main contributions: First, we compare wrist EDA (convenient for continuous long-term measurement) to palmar EDA (inconvenient). When we began this work, there was concern that the wrist measures would primarily reflect thermal sweating. Our work finds significant EDA patterns in sleep from the forearm while simultaneously measuring skin temperature at the same position.

Second, we characterize EDA in natural sleep, proposing an automated method to extract features from the EDA, and using these features to create a taxonomy of EDA patterns during sleep. For 15 nights where we have concurrent synchronized polysomnography (PSG), we also characterize the EDA-PSG relationships and compare the new measures with results published in the 1960-70's. PSG is currently the gold standard to evaluate and diagnose sleep patterns; however, the use of PSG requires scalp EEG electrodes and other sensors that tend to be uncomfortable and expensive, time-consuming to apply, and arguably interfere with the sleep they are measuring. Actigraphy is a much less invasive method often used to estimate daytime and sleep activity with a wrist-worn device; however, it does not measure neural activity such as stages of sleep. In this study, we measure both EDA and actigraphy to develop a quantitative characterization of EDA in natural sleep.

Third, we also compare EDA responses with skin temperature. It has long been recognized that thermoregulatory processes are suppressed during REM, while they persist during NREM [Adam, 1986]. In a study of five healthy men, the largest sweating, averaged across multiple sites on the body, was recorded during SWS while the lowest was recorded during REM, although sweating was not completely

blocked during REM [Sagot, 1987]. But this occurred in the absence of significant changes in skin temperature across sleep stages. We provide the first characterization of the interaction between wrist/palm EDA, skin temperature, and sleep stages.

Table 3.1 Summary of previous sleep EDA studies

	Description	Location
Asahina et al., 1964, N=20,	GSR high activity in stage 4	galvanic skin response (measurement location unknown)
Broughton et al., 1965, N=unknown	Responses are frequent in stage 4, and rare in REM sleep	electrodermal response on palm and dorsal forearm
Lester et al., 1967, N=53	More GSR peaks in stage 4	Galvanic skin response on finger
Koumans et al., 1968, N = unknown	Electrodermal fluctuations increase during SWS and decrease during REM	skin potential and response on palm and dorsal surface of forearm
Hori et al. 1970, N=15	Skin potential response max: SWS, low: REM	skin potential activity on the palmar surface of finger and dorsal surface of hand
McDonald et al., 1976, N=46	Storming in stage 3-4	skin potential and resistance, unknown location
Freixa i Baqué et al., 1983a, N=8	Spontaneous skin potential responses increase during 2-4 sleep cycles	electrodermal activity on palm and dorsal surface of hand
Ware et al., 1984, N=12	Storming occurs during NREM sleep	skin resistance response on hands
Burch, 1985, N=unknown,	GSR storms during sleep stage 4	skin response (location unknown);
Liguori et al., 2000, N=53	Spontaneous sympathetic skin responses was highest in stage 4 and lowest in REM sleep	Sympathetic skin response on hand
Kobayashi et al., 2003, N=8	The GSR peaks and sweat rate were significantly less frequent during REM sleep than during NREM sleep.	Galvanic skin response on the dorsal side of hand;

Table 3.2 Summary of previous EDA studies

	Location
Johnson et al, 1966, N = 29	Finger, GSR and SCR, sleep lab
Johns et al, 1969, N=31	Finger, GSR, sleep lab
Liguori et al., 2000, N=53	Hand, sympathetic skin response, sleep lab
Shiihara et al., 2000, N=5	Finger, Skin conductance, Palm, Skin potential, sleep lab
Kobayashi et al., 2003, N=8	Hand, galvanic skin response, sleep lab
Poh et al., 2010, N=26	Finger and inner wrist, Electrodemal Activity, Physical, cognitive and emotional tasks
Poh et al. 2012, N=80	Wrist, electrodermal acitivity, epilepsy patient admitted to the long-term video-EEG monitoring unit
van Dooren et al., 2012, N=17	16 positions (fingers, distal wrist, central wrist, vertical wrist, chest, foot (instep), calf, forehead, neck, shoulders, back, buttock, abdomen, armpit, upper arm, and thighbone), skin conductance, watch emotional film clips

3.2 Methods

3.2.1 Measurement

Our studies examined EDA during sleep by monitoring skin conductance on the outer or inner wrist (dorsal or ventral forearm) or on the palmar surface, using the Affectiva Q^{TM} sensor with 1cm diameter Ag-AgCl dry electrodes. The sensor logged EDA, actigraphy (3-axis accelerometer) and skin surface temperature at 32 Hz. The Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects (COUHES) approved both studies.

EDA at home from wrist and palm from healthy adult (65 nights)

Nine healthy adults (two females) wore the Q sensors on the right palm and wrist for one night each. A tenth person (healthy adult female) wore the Q sensors for 56 nights. Participants put the sensor on before going to bed, and took it off after waking.

EDA with concurrent PSG (15 nights)

Fifteen healthy university students (ages 18-22, 10 males) participated in a night of measurements in a sleep laboratory, wearing the Q sensor on the wrist. Sleep was simultaneously monitored with standard PSG and scored by standard criteria [Rechtschaffen, 1968].

3.2.2 Definition

We define the following terms:

EDA peak: Local EDA maximum that exceeds a defined threshold (see analysis below for details).
EDA-peak epoch: A 30 second section of EDA having at least one EDA peak
EDA storm: Consecutive EDA peak epochs. Thus, an EDA storm has a minimum duration of one minute, and has at least two peaks during that minute.
Burch storm: "A minimum of 5 galvanic skin response (GSR) peaks per minute for 10 consecutive minutes of sleep" [Burch N, 1965; Lester, 1967]
EDA event: A section of EDA data having one or more EDA peaks or storms (*e.g.*, an EDA isolated peak, EDA peak epoch, EDA storm or Burch storm)

3.3 Analysis

In this work, we automate the processing of EDA data in order to remove noise and to extract features that are robust and meaningful for characterizing sleep, and in order to provide objective measures that can be used across nights, across participants, and across studies. In PSG, it is standard practice to label sleep stages in 30-second epochs; thus, we adopt the length of 30-second segments for our comparison analyses. The EDA data were processed in four steps.

1. Detection of sleep from actigraphy: Standard zero-crossing detection and Cole's function were applied to the accelerometer data to discriminate between sleep and wake [Cole, 1992]. Only EDA data that corresponded to the times scored as sleep were further processed. Thus, EDA data that might be associated during the night with getting out of bed and moving around were not included in the analyses below.

2. Pre-processing of EDA: All EDA data that corresponded to segments of sleep were subsequently low-pass filtered (cutoff frequency 0.4 Hz, 32nd order FIR filter).

3. EDA peaks: After EDA data were low-pass filtered, we computed the first derivative and determined where it exceeds a threshold. Part of our effort asked, "What is the optimal threshold that has meaning for sleep data?" We conduct in this study tests varying the threshold over these values: 0.005, 0.01, 0.02, 0.03,

0.04 and 0.05 μS/s and describe below how dependent the results are on the particular value. In subsequent analyses comparing wrist and palm EDA, we used a threshold of 0.01 μS/s. We define EDA "peaks" as those whose rise phase exceeds the threshold. Peaks must be separated by at least one second or they will be counted as a single peak. Thus, this method can detect up to 30 peaks per epoch, although in sleep the most we have seen is 13 peaks in one epoch.

4. EDA storms: Our definition above is that an EDA storm must consist of at least two adjacent peak epochs. Thus, the slowest possible storm would have 2 peaks per minute. Often during sleep we see regions with much faster bursts of 5-8 peaks per minute (ppm), and once we saw 26 ppm. During our analysis, we compared EDA storms to previous ones in the literature. Thus, for the analyses below, we compared definitions requiring 1, 2, 3, and 4 EDA peaks per epoch, before clustering the adjacent epochs into "storms."

The EDA peak detector we developed is fully automated and has been quantitatively and qualitatively validated for accuracy. Figure 3.1 shows 10 seconds and 1 minute of EDA raw data and its derivative. Peaks shown here (black dots) are automatically detected when the derivative exceeds the threshold of 0.005 μS (red line). An asterisk marks the location of the rising edge of the peak. All peaks during sleep that meet the criteria are detected except when 2 peaks occur less than 1 second apart. When 2 peaks are less than 1 second apart then it marks only the first of the two peaks. The third peak in the bottom of Figure 3.1 (x and arrow) is not detected as two peaks occur within a second.

Figure 3.2 displays one night of filtered EDA data, the number of EDA peaks for each 30-second epoch, along with a 4-min segment of the filtered EDA data and the detection of EDA peaks for the 4-min segment using the most sensitive threshold of 0.005 μS/s.

Our analysis, below, has three main parts:

(1) Compare EDA amplitude (skin conductance level) and the number of peaks for wrist and palm recordings.

(2) Compare wrist EDA amplitude and the number of peaks in sleep stages and during the four quarters of the night (ANOVA and post hoc t-test); also characterize storm durations.

(3) Compare EDA and skin-surface temperature at the EDA electrodes (correlation analysis)

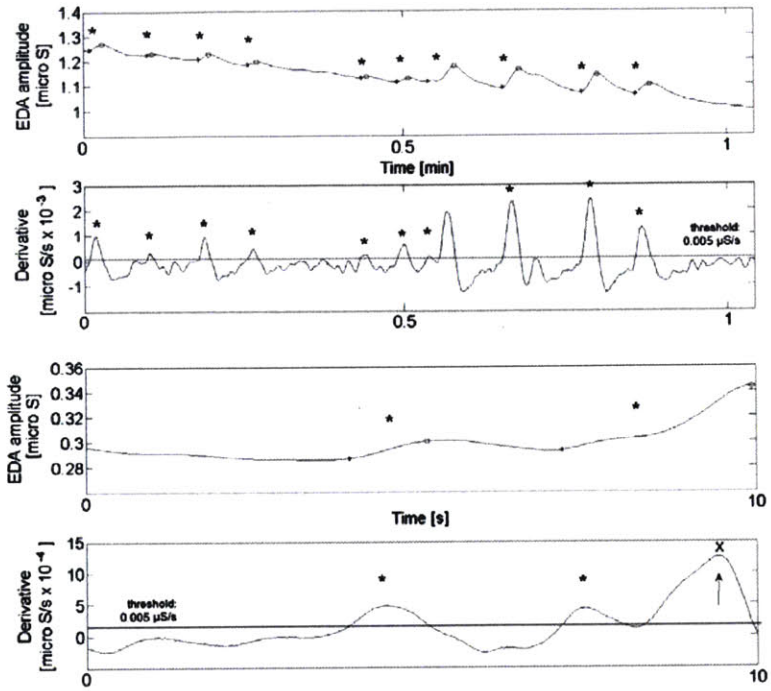


Figure 3.1 EDA peak detection (EDA amplitude and derivatives)

The black asterisks show detected peaks and x shows a peak detected within 1 second after the previous one and counted as one peak.

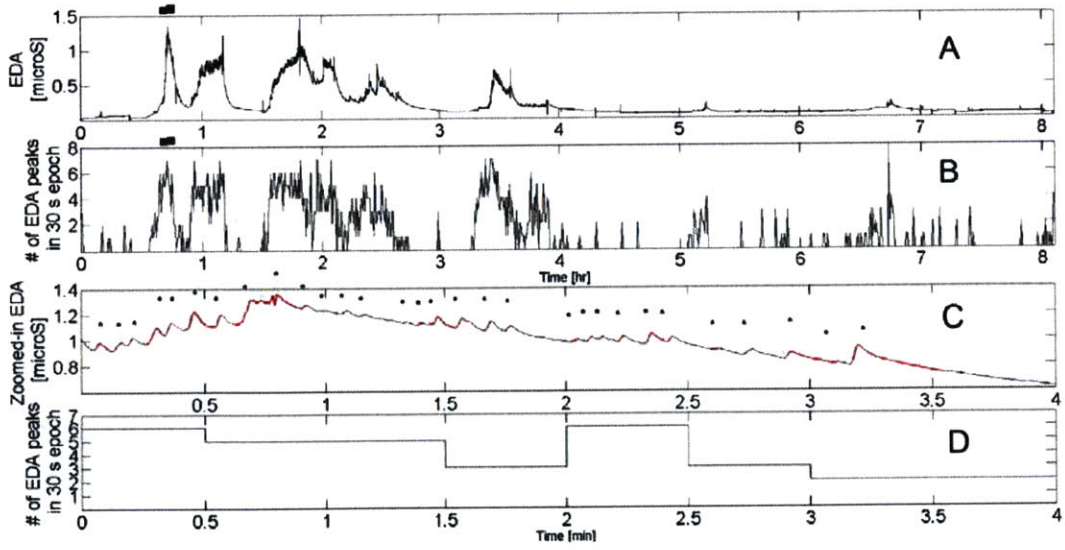


Figure 3.2 A:filtered EDA data for one night in a healthy adult. B: detected EDA peaks in 30-s epochs. C: zoom of region marked with a bar on A. D: # of EDA peaks in each 30-s epoch.

3.4 Results

3.4.1 Wrist vs. Palmar EDA

Most prior studies of EDA during sleep have looked at palmar skin conductance as a measure of EDA, e.g. Doberenz et al collected one night of palmar data from each of 48 subjects [Doberenz, 2011]. We found that EDA measured on the wrist usually gives a larger signal than that measured on the palm, although otherwise the two signals are usually reasonably correlated during sleep (e.g., Figure 3.3). To quantify this, we analyzed the difference between the wrist and palm EDA data (after filtering as above) from 9 healthy adults using 0.03 μS as tolerance (epsilon). Across participants, the palmar skin conductance measured during sleep was at least 0.03 μS lower than the inner wrist skin conductance during 74% of samples. Despite this difference, the palm and the inner wrist EDA show the same number of EDA peaks for 71% of 30-sec sleep epochs, with more EDA peaks on the wrist seen during 21% of sleep epochs and less on the wrist during the other 8%.

We also analyzed the difference between the wrist and palm EDA data for 56 nights (longitudinal case study) because, increasingly, long-term measurement is important in understanding intra-individual differences as well as in treatment and intervention studies, and we wish to compare a set of individual results to the group results. As shown in Figure 3.4, on 48 of the 56 nights (86%), the average skin conductance level measured from the inside of the wrist was higher than the palmar level during sleep (both measured on the right side of the body). On the remaining 8 nights, the palmar skin conductance had larger amplitude than the wrist. When analyzed by hour of sleep, the wrist EDA was higher than the palmar EDA 71% of the time (255 hours of sleep), while 23% of the time (84 hrs of sleep) the palmar EDA exceeded the wrist EDA, and 5% of the time (18 hrs of sleep), the difference between wrist and palmar EDA was less than 0.03 μS.

Our software used the 0.01 threshold as mentioned above and detected EDA peaks during sleep both for palm and wrist on all 56 nights. As seen in Figure 3.5, on 42 of the 56 nights, more EDA peaks were detected on the inner wrist. Of 357 hours of sleep, the wrist and palmar EDA-peak counts per epoch were equal 83% of the time (296 hours of sleep); 12% of the time the wrist EDA showed more peaks (42 hrs of sleep), and 5% of the time the palmer EDA had more peaks (19 hrs of sleep). Thus, overall the wrist appears to be a more sensitive location for capturing EDA events during sleep. Moreover, these results were consistent both across individuals and long-term within an individual.

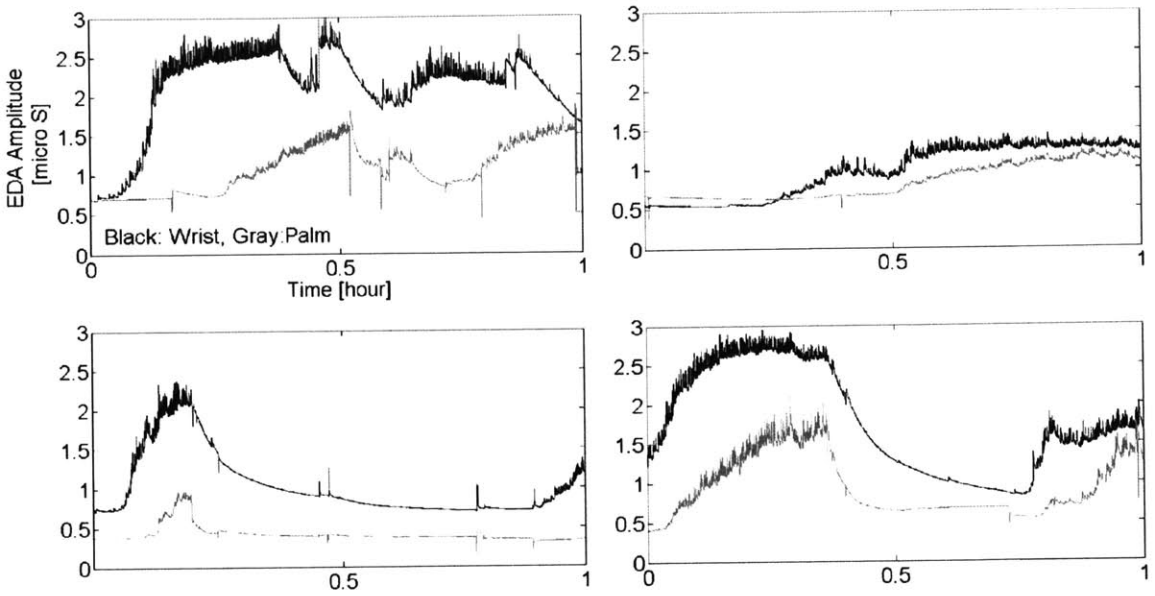


Figure 3.3 Examples of wrist and palm EDA during sleep

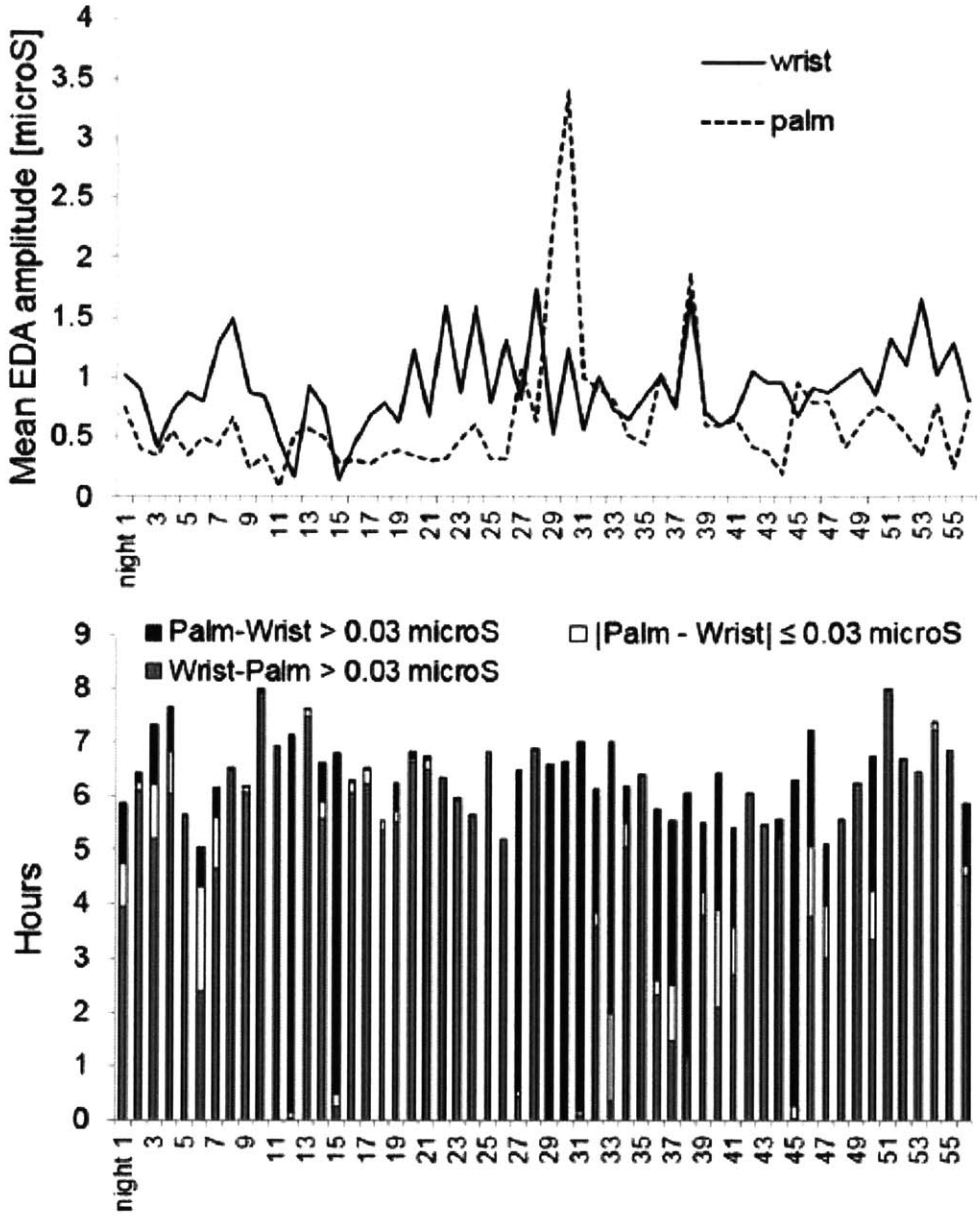


Figure 3.4 EDA amplitude comparison between the palm and the wrist

(56 nights, 357h, one participant long-term)

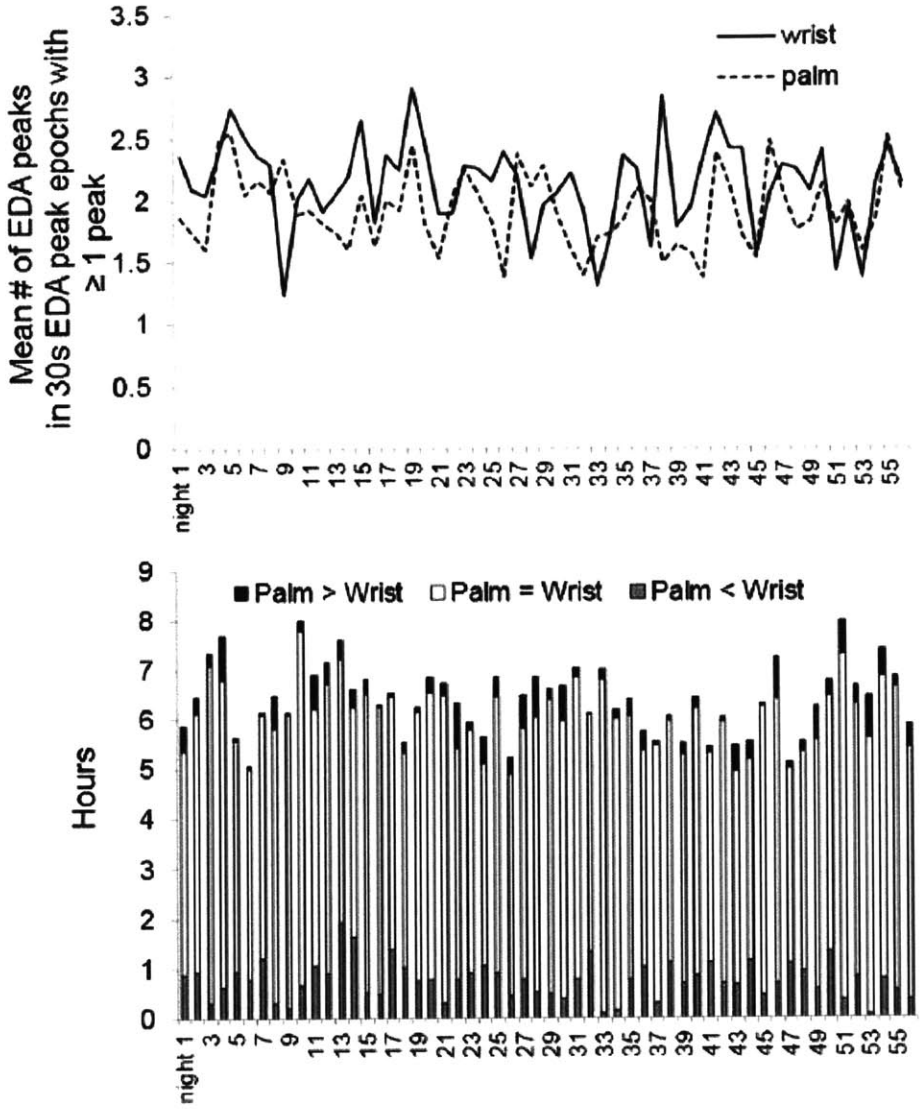


Figure 3.5 # of EDA peak comparison between the palm and the wrist (56 nights, 287 h)

Y-axis of the top figure: mean # of EDA peaks per 30s epoch containing more than 1 peak, not for all epochs

3.4.2 Characteristics of EDA

We wish to characterize EDA peaks and their relation to sleep stages. First, we examine the sensitivity of the peak-detection parameters for our automated algorithm. We computed the distribution of the number of EDA peaks per 30s epoch for thresholds from .005 to .05 μS/sec (n = 15 in the laboratory). Over the fifteen

nights, more than 60% of the 30-s epochs did not show any peaks, regardless of peak threshold. As expected, a lower threshold for EDA peaks showed more peaks.

We then analyzed how the peaks that occurred are distributed across the sleep stages Most of the night was spent in NREM2, and indeed we see most of the peaks (55 ± 4%) occurred in NREM2. The next highest are 25±4% in SWS, 12±1% in REM and 4±0% in NREM1. This relative ordering of NREM2 > SWS > REM > NREM1 holds regardless of the threshold that we used for detecting peaks. Thus, this finding is robust over a large range of parameter values. However, the relative number found in each stage varied: the ratio of EDA peaks in REM compared to SWS varies systematically from 39% at the highest threshold to 77% at the lowest.

Figure 3.6 shows that SWS has the highest percentage of epochs with EDA peaks during sleep. The percentage of sleep epochs containing EDA peaks varied significantly across sleep stages (repeated measures ANOVA, F=12.70, df=3, p< 4.82E-06). Overall, EDA peaks were more than 1.5 times more frequent in SWS than in NREM2 and more than 3 times more frequent in SWS than in REM (post hoc t-test, p=0.05). While the exact percentages of peaks decrease as the threshold gets higher, the main findings relating EDA to sleep stages are consistent for thresholds from .005 to .05 μS. Thus, the EDA peaks measured on the wrist with dry electrodes show robust properties related to sleep stages. Figure 3.7 shows the distribution of EDA peak epochs over the night. Most of the EDA peak epochs occurred in the first half of the night.

Next, we analyze the basic properties of EDA amplitude, peaks and storms. Median EDA-amplitude (averaged median across participants) was 0.44, 0.26, 0.18, and 0.26 in SWS, NREM2, NREM1 and REM. The median EDA amplitude in SWS was significantly higher than in the other sleep stages. (ANOVA and post hoc t-test, $p < 0.05$). (We computed the median because the distribution of EDA amplitude is far from Gaussian.) Thus, the wrist EDA median amplitude varies with sleep stages. We also compared the EDA amplitude between epochs with EDA peaks and those without EDA peaks. In twelve out of 15 participants, median EDA amplitude was higher in epochs with EDA peaks. The EDA-peak frequency (peaks per epoch) was also significantly higher in SWS than in NREM2, NREM1 and REM (ANOVA and post hoc t-test, p=0.05).

We also validated the robustness of the new automated criteria for detecting EDA storms: the number of EDA peaks required per epoch. We again found that the relative distribution of storms is robust across the criteria: About 85% of storms lasted under 5 minutes regardless of the various amplitude gain thresholds for EDA peaks (0.005 – 0.05 μS) and regardless of the various peaks-per-epoch thresholds for EDA storms (1-4 peaks/epoch).

Burch was the original scientist identifying EDA storms, which he and his colleague identified visually after measuring GSR on the left middle finger with Ag-AgCl electrodes and a sodium-chloride

paste [Lester, 1967]. We wanted to compare today's objective sensor data and automated algorithm to their original hand-counted values. Among all wrist EDA events in our data, only 11% of EDA events met Burch's criteria (≥5 EDA peaks/min, and duration ≥ 10 minutes). Of these Burch storms, 95% occurred during NREM2 and SWS, compared to 89% of isolated EDA peaks and non-Burch storms. Similarly, 77% of Burch storms occurred during the first half of the night, compared to only 43% of the other peaks and storms. Thus, we have qualitative similarities between our automated and objective measures and Burch's hand-count observations in EDA peaks and storm occurrences in NREM2 and SWS, but difference in their distribution across the night.

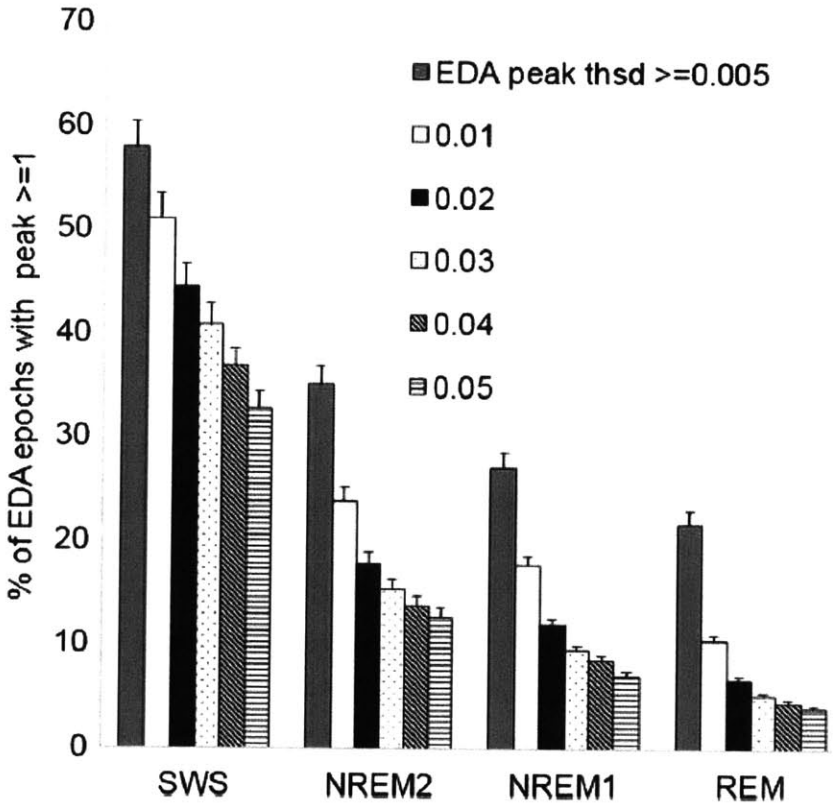


Figure 3.6 Mean percentages of sleep stage epochs containing EDA peaks (N=15, error bars: s.e.m.)

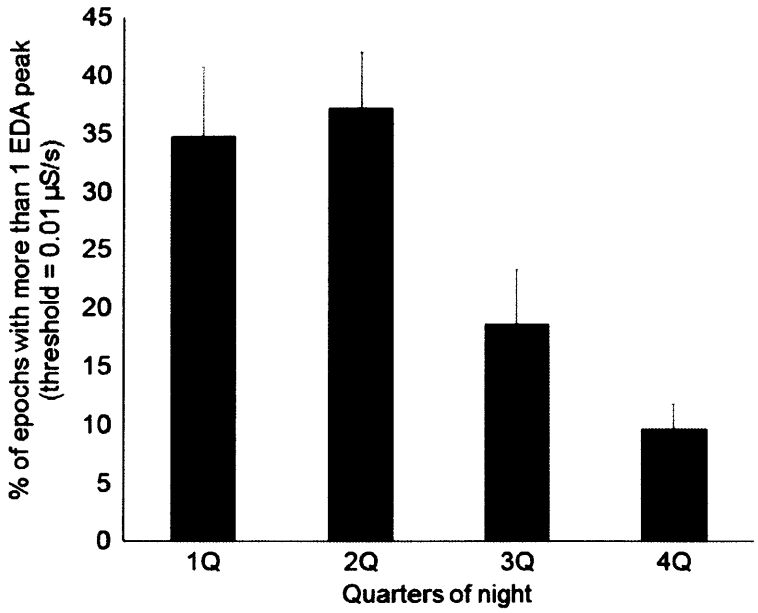


Figure 3.7 Percentages of epochs with more than 1 EDA peak (threshold = 0.01 μS/s)

3.4.3 EDA vs Skin Temperature

The purpose of the analysis here is to determine whether skin surface temperature is the cause of the EDA changes we see during sleep. Note that skin surface temperature is not the same as core body temperature; core body temperature drop is usually preceded by wrist temperature increase [Sarabia, 2008]. We have also found that skin temperature tends to climb for most of our participants during sleep, which is consistent with the previous finding [Martinez-Nicolas, 2013]. We do not have measures of ambient temperature or of whether or not the person's wrist was under a blanket, which is likely to make the skin warmer; nonetheless, it is still interesting and meaningful to examine correlations between the skin surface temperature and the EDA, both measured at the position of the same pair of electrodes because it still tells us if temperature change influences EDA activation. We first examine the correlation between skin temperature and EDA overall as well as during each sleep stage. Out of 15 participants, 12 participants showed significant positive correlations between 30s epoch averaged skin conductance level (SCL) and 30s epoch averaged skin temperature level. Also, 9 of the 15 participants showed significant positive correlation between the number of EDA peaks and skin temperature per epoch. However, 13 out of 15 participants also showed higher wrist temperature in SWS than in REM generally, making causal links unclear. While EDA amplitude and peaks do have a statistical relationship with skin temperature in our 30-sec data, the

correlation breaks down at a finer time scale. Examples can be found (*e.g.*, Figure 3.8), where EDA and skin temperature are completely dissociated. Thus, increases in EDA amplitude and peaks are not simply the immediate consequence of changes in skin surface temperature.

Both the wrist and the palm contain eccrine sweat glands, which have a primary function of thermoregulation, and which are denser on the palm than on the wrist [Dawson, 2007]. We examined if wrist or palm differed in how their EDA responded to temperature during sleep, comparing wrist and palm temperature when there were and were not EDA peaks. On the wrist, 6 out of 9 participants showed higher temperature during epochs without peaks than with peaks; thus, the EDA peaks were not simply associated with warmer skin on the wrist. In contrast, on the palm, 7 out of 9 showed higher temperature during epochs with EDA peaks than without (wrist *vs.* palm, $c^2 = 3.6, p = 0.058$). Thus, there may be a slight tendency for higher temperature on the palms to lead to more peaks on the palms (binomial, $p = .089$). All 9 participants showed higher mean temperature on the wrist than on the palm during EDA peak epochs. Also, 7 out of 9 showed higher mean wrist temperature than palm during non-storm epochs. When the wrist temperature was higher than the palm temperature, then the wrist EDA was almost always higher than the palm EDA (95% of these epochs).

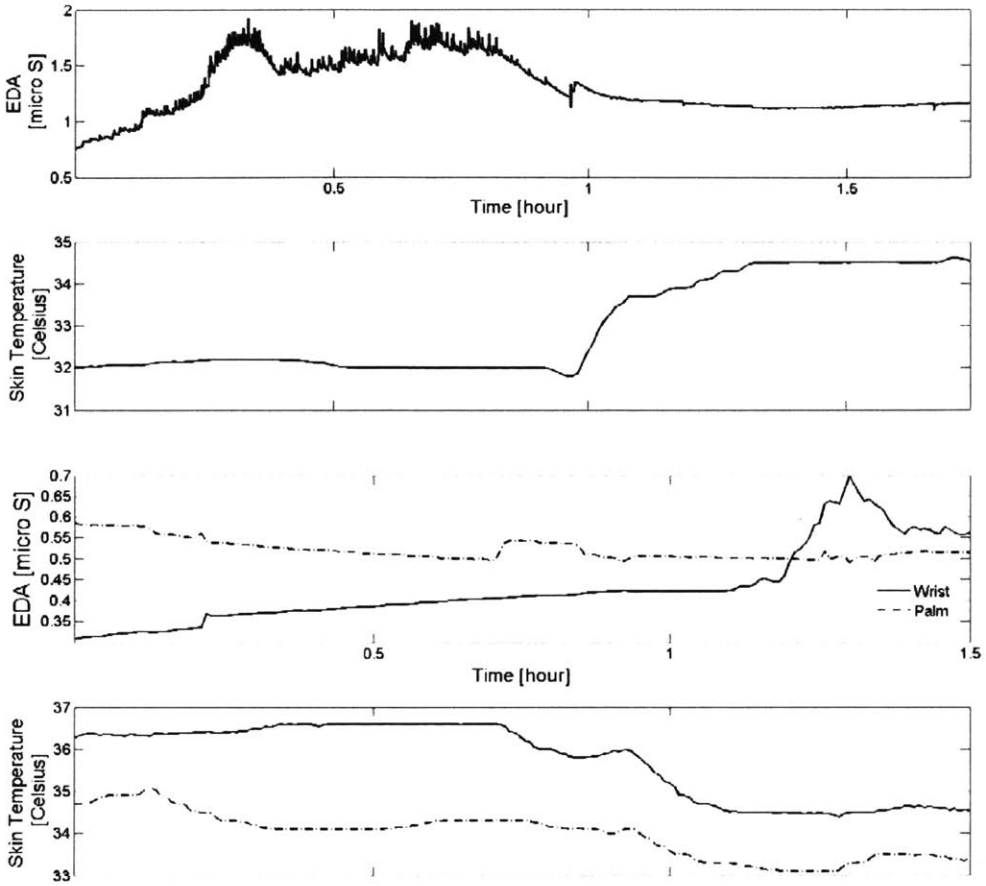


Figure 3.8 Examples from sleep showing that changes in EDA are not always caused by changes in temperature (upper two: skin temperature on the wrist was flat when EDA showed storms and there are no storms when temperature climbs, lower two: wrist EDA goes up when wrist temperature has no change)

3.5 Discussion

This EDA study, with 80 nights of data, examined and characterized basic EDA properties during sleep. Our study includes the first longitudinal characterization (56 nights) as well as 15 nights with synchronized PSG and nine additional nights of healthy adults at home. Consistent with previous studies [McDonald, 1976, Burch 1965], our data showed that the mean EDA amplitude in SWS is significantly larger than in

other sleep stages. We also observed a decreased number of peaks in EDA during REM sleep, which was consistent with these prior studies (Asahina, 1964, Broughton, 1965, Hori, 1970, McDonald, 1976, Ware, 1984, Burch 1965, Ligori, 2000). These common findings are noteworthy because ours is a significant sleep study to use a convenient-to-wear dry-electrode EDA skin conductance sensor on the wrist, while most prior studies measured the EDA on the palmar surface or fingers with wired gelled electrodes. We also developed a fully automated EDA sleep peak detection algorithm providing objective measures across a range of thresholds, and showed that the findings were robust across these thresholds. We will further discuss comparisons of forearm vs. palmar EDA below, but these significant findings serve to validate both the occurrence of EDA peaks and the sleep-stage dependence of the EDA peaks for this alternate convenient location of wearing a sensor.

In our study, EDA peaks were not distributed uniformly over the night, but were more likely to be located in the first half of the night. This can be because more SWS occurs in the earlier half of the night. However, some nights showed no EDA peaks in the first SWS cycle. It is important to note that EDA peaks and storms did not happen in all cycles of SWS and NREM2; thus the EDA peaks provide different information than that normally obtained from PSG. In fact on some nights, some participants have no EDA storms, while on other nights they may have many. Meanwhile, when EDA storming does happen, it is most likely to appear during SWS and NREM2.

We found that the largest number of peaks per epoch occurred in SWS and NREM2. Freixa i Baqué et al. (1983b), Johnson et al. (1966) and Hori et al. (1970) also found more peaks in SWS, and McDonald showed a decrease in the EDA storm rate in NREM 1 and NREM 2 sleep (1976), all of which are consistent with our results. Liguori et al. (2000) showed that the frequency of spontaneous sympathetic skin conductance peaks in stage 4 (SWS) was 5-9 per minute. This result is slightly different but consistent with our tendency (the most frequent in SWS, 2- 26 per minute). One earlier finding that did not match ours is that of Freixa i Baqué et al. (1983b) who found that spontaneous EDA activity showed a smaller number of EDA peaks per minute (i.e., 60%) during the first sleep cycle (the different EEG stages from sleep onset appearance of alpha rhythm (NREM1) until the end of the first REM) compared to the subsequent three sleep cycles (defined as different EEG stages between the ends of two REM periods) (Freixa i Baqué, 1983a, N=8). In our data, the first and second quarters of the night showed a larger number of EDA peaks per 30-second epoch than the latter two. Hori et al. also visually found that EDA peak frequency was less frequent in the latter half of sleep, especially after the third full REM cycle (Hori, 1970, N=15), consistent with our findings based on objective wrist EDA data.

Most of the EDA storms in our data lasted under 5 minutes. Of all EDA peaks detected over the 80 nights, only 11% were in EDA storms that met Burch's storm criteria. Nevertheless, we found more storms per night than the 2-3 storm nightly average reported by Freixa i Baqué [Freixa i Baqué, 1983a]; this may

in part be due to the stronger EDA signal obtained when measured on the wrist. In addition, our result showed that longer storms, with a larger number of EDA peaks, are more likely to occur in the earlier part of the night and in SWS and NREM2.

We measured EDA on the forearm using a wristband, while previous studies examined EDA mostly on the palmar surface or fingers. Our results showed that the EDA amplitude and storm patterns during sleep are usually more pronounced on the forearm than on the palm, and thus peaks are more likely to be detected when measured with a wristband. These observations are the opposite from activities during daytime awake tasks [Van Dooren, 2012] where peaks tend to be more pronounced on the palm. The stronger signal we observed on the wrist during sleep may explain why we found more EDA peaks than earlier studies, not only during SWS but also during NREM2 as well. This sensitivity on the wrist was found even using dry electrodes, which avoids the problem of a gel breaking down over long-term wear and interfering with signal level over time.

Our findings of a higher mean skin temperature during SWS may appear to contradict those of Sagot et al. (1987) who showed no statistical relationship between skin temperature and sleep stages; however, they averaged skin temperature from 10 different points on the body, including distal and proximal skin temperature, while our findings were specific to the wrists.

Warmer wrists help explain the higher SCL and larger number of peaks found on the wrists overall. That said, we cannot say that the higher SCL and peaking are always associated with skin surface temperature changes: There are instances, such as Figure 3.8, where SCL on the palm is higher than on the wrist, while the skin temperature is higher on the wrist than on the palm. An overall correlation is present, but the relatively rapid changes we see in EDA do not appear to be caused only by changes in skin surface temperature.

When we began these studies, we were initially perplexed by this discrepancy: During sleep, we would expect low emotional arousal and low EDA responsivity; however, we found higher EDA responsivity, even after removal of sleep-motion artifacts, and even at times when skin surface temperature was dropping. Since that surprise, we have learned about key neurological findings showing, for example, that the amygdala and hippocampus, when directly stimulated with depth electrodes, elicit large skin conductance responses [Mangina, 1996]. The amygdala and hippocampus regions of the brain are known for being involved in memory and emotion. In fact, in recent work we have found that automatically computed features of the skin conductance over a night's sleep are more accurate predictors of improvement in a learning task (learned before sleep, tested after sleep) than are classic features measured from EEG or from PSG (Sano, 2013a, Chapter 4 in this thesis). It is thus possible that neurological memory-related processes may also be contributing to the patterns of EDA responsivity we measure during sleep.

This study has several limitations. Several factors can influence an individual's EDA. For example, thermal regulation influences sweating and we did not measure core body temperature or environmental temperature, nor did we videotape to track the position of participants' wrists. Only the temperature on the skin location of the EDA electrodes was measured. Core body temperature is usually higher earlier in sleep (when there is usually more SWS) and tapers down over the course of sleep. Sleep stages such as SWS and NREM2 have been associated with higher core body temperature on average than REM [Sagot, 1987]. Core body temperature behaves in ways different from distal skin surface temperature [Kräuchi, 2002]; thus, thermoregulation remains a potential driver of some of our findings, even when there is no strong correlation between temperature at the electrode location and the skin conductance measured at the same position. Another mystery is that some nights had no EDA responses, despite that we might still expect that core body temperature dropped over the night. One possible explanation for the women in the study is that they have reduced sweating during the luteal phase (latter half) of their menstrual cycle, and this could cause a reduction in measured EDA storm peaks [Mackinnon, 1954]. Future sleep studies should examine the timing of the measurements made relative to female participants' menstrual cycles. Our longitudinal study of one subject, who was female, showed quite a bit of variation from night to night in the EDA patterns. Future work is needed to characterize inter- and intra- individual differences in long-term EDA features.

3.6 Conclusion

This work presents the systematic taxonomy of autonomic activity patterns measured in healthy adults based on forearm skin conductance and actigraphy during sleep. Our analyses focused on the automated detection of EDA peaks and on regions of continuous peaks called "storms," and their comparison with concurrent PSG as well as with skin surface temperature.

Most of the EDA data in this study were measured from the wrist and on most nights the results showed greater activity at this location than at the traditional palmar location in terms of both amplitude and the number of peaks; thus, the wrist is a viable location to get long-term data about EDA patterns during sleep.

About 80% of wrist EDA peaks are observed in SWS and NREM2 sleep, and mostly in the first half of the night. This property is robust over different thresholds to detect EDA peaks. Only 11% of all EDA peak epochs were contained in Burch's EDA storms (classically defined as more than 5 peaks per minute and durations longer than 10 minutes), and these occurred mostly in the first half of the night. EDA amplitude was also on average higher during EDA-peak epochs.

We analyzed the relationship between EDA and skin temperature, where we found a higher frequency of EDA peaks and a higher average skin conductance level in SWS, measured on the wrist, tending to co-occur with higher temperature on the wrist, although not always in association with higher temperature. While we know that thermoregulation influences EDA, the temperature on the surface of the skin does not fully account for the EDA patterns measured at that location.

Overall, our work has characterized strong patterns in EDA that can be measured at home or in the lab, using automated methods that are robust to different parameter settings. Our findings characterize consistent EDA patterns related to sleep stages derived from gold standard PSG. Future work is needed to elucidate the many neurological, environmental, and thermoregulatory influences contributing to the rise of these EDA patterns.

Chapter 4
Multi-modal wearable data analysis

In the last chapter, we characterized wrist EDA responses during sleep. This chapter introduces two studies we conducted using wearable sensors including sleep EDA and describe how they contribute (1) to understand sleep-related memory consolidation and (2) to classify sleep/wake epochs.

4.1 Recognition of Sleep Dependent Memory Consolidation with Multi-modal Sensor Data

This study presents the possibility of recognizing sleep dependent memory consolidation using multi-modal sensor data. We collected visual discrimination task (VDT) performance before and after sleep at laboratory, hospital and home for N=24 participants while recording EEG (electroencepharogram), EDA (electrodermal activity) and ACC (accelerometer) or actigraphy data during sleep. We extracted features and applied machine learning techniques (discriminant analysis, support vector machine and k-nearest neighbor) from the sleep data to classify whether the participants showed improvement in the memory task. Our results showed 60-70% accuracy in a binary classification of task performance using EDA or EDA+ACC features, which provided an improvement over the more traditional use of sleep stages (the percentages of SWS in the 1st quarter and REM in the 4th quarter of the night) to predict VDT improvement.

4.1.1 Introduction

Past studies have shown that sleep can enhance memory consolidation. Some studies have shown the relation to REM sleep [Siegel, 2001] [Karni, 1994]. Stickgold et al. showed that consistent and significant performance improvement on a Visual Discrimination Task (VDT) became proportional to the amount of sleep in excess of six hours, and subjects with an average of eight hours then exhibited a correlation in performance to the % of sleep stages: % of SWS (Slow Wave Sleep) in the first quarter of the night, and % of REM (Rapid Eye Movements) in the last quarter [Stickgold, 2000]. To our knowledge, no prior studies have attempted to classify whether sleep-dependent memory consolidation occurred by using automated analysis of sensor data during sleep. This study examines whether EEG (electroencephalogram), EDA (electrodermal activity) and Actigraphy data during sleep can predict task performance improvement on the VDT.

4.1.2 Methods

A. Measurement

Twenty-four healthy university students (ages 18-22, 16 males) participated in 3 nights of measurements, one night in a "homey" sleep laboratory, one night in a hospital GCRC, and one night at home to measure physiological changes related to task performance on a VDT, measured before and after sleep. The physiological measures consisted of EDA (a measure of sympathetic nervous system activity), skin temperature and actigraphy (all three measured from the wrist at 32Hz) each night and EEG (C3, C4, O1 and O2 under international 10-20 method, 100Hz) for the nights in the sleep lab and hospital GCRC. The Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects (COUHES) approved both studies.

Each night (PM), participants trained on a different version of the VDT, slept, and were tested the next morning (AM). Sleep in the sleep lab and GCRC was also monitored with standard PSG (polysomnography), consisting of 30-second epochs of sleep stages labeled by experts (Wake, REM, Non-REM 1, 2 and SWS).

We evaluated task improvement by overnight change (PM-AM) in VDT performance (a lower score is better performance). We obtained standard PSG sleep staging as well as subjective sleep quality evaluations on a scale of 1 to 4. Unfortunately, out of all data, only 15 nights of data from 15 participants (10 males) in the hospital had accurately time-synched EDA data with concurrent PSG, so while we have EDA and actigraphy for all nights, we only have it synchronized with sleep stage information for a subset of the nights.

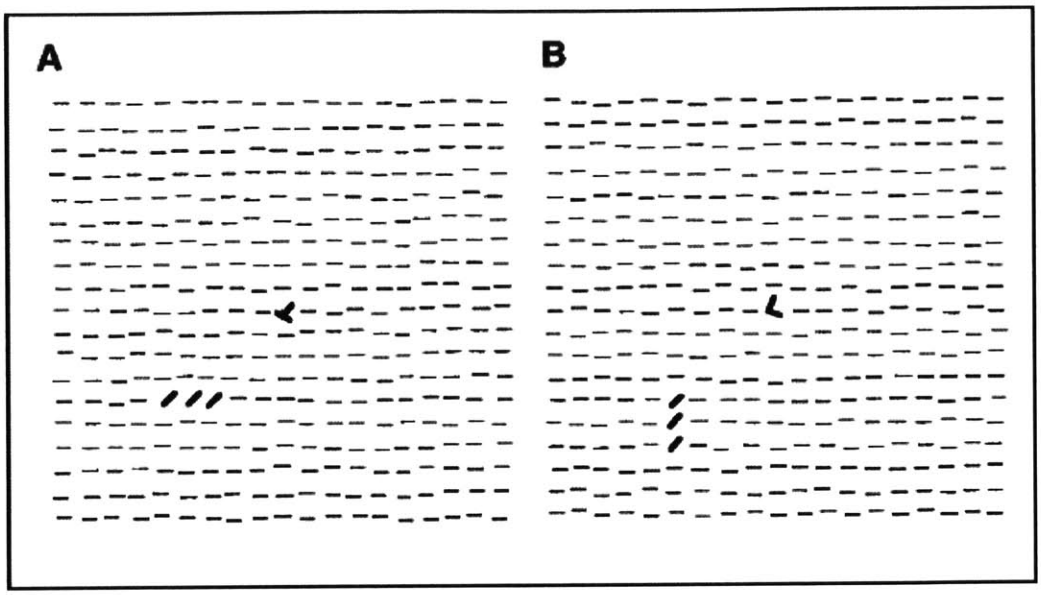


Figure 4.1. Sample target screens for the visual discrimination task.

The VDT was a task that has been used routinely in sleep and memory studies by Stickgold et al. (2000). Each target screen contained a rotated "T" (A) and "L" (B) at the fixation point and a horizontal (A) or vertical (B) array of three diagonal bars in the lower-left quadrant of the visual field (Reprinted with permission from [Stickgold, 2000]).

Figure 4.1 shows one trial of the VDT method, consisting of 5 screens that appear in typically less than 1 second. The first screen is a fixation screen (black with a white centered crosshair) that remains until the participant hits a key. This is followed by a 16-ms target screen, a 0-400-ms blank "interstimulus interval" (ISI), and then a 16-ms mask screen. The ISI varies over the course of the task, starting at 400 ms and is progressively shortened to 0ms over the 25 blocks of 50 trials. Each participant was asked to determine two features of the target screen: whether the capital letter in the center of the screen was "T" or "L" and whether an array of three diagonal bars in one quadrant of the screen is horizontal or vertical. By interpolating the ISI at which 80% accuracy is achieved on the horizontal versus vertical decision, a 'threshold' ISI (in ms) was extracted from each session of the VDT. A lower threshold is a better performance. Overnight improvement or deterioration on the VDT was then calculated as a subtraction of the AM VDT threshold from the PM threshold. For example, 30 ms indicates the person performed 30 ms better in the AM than in the PM, while a negative value signals deterioration from PM to AM. One VDT session consisted of 25 blocks with 50 trials in each.

B. Feature Extraction

Figure 4.2 shows a sample representation of one night's data from one participant when the PSG was fully synchronized.

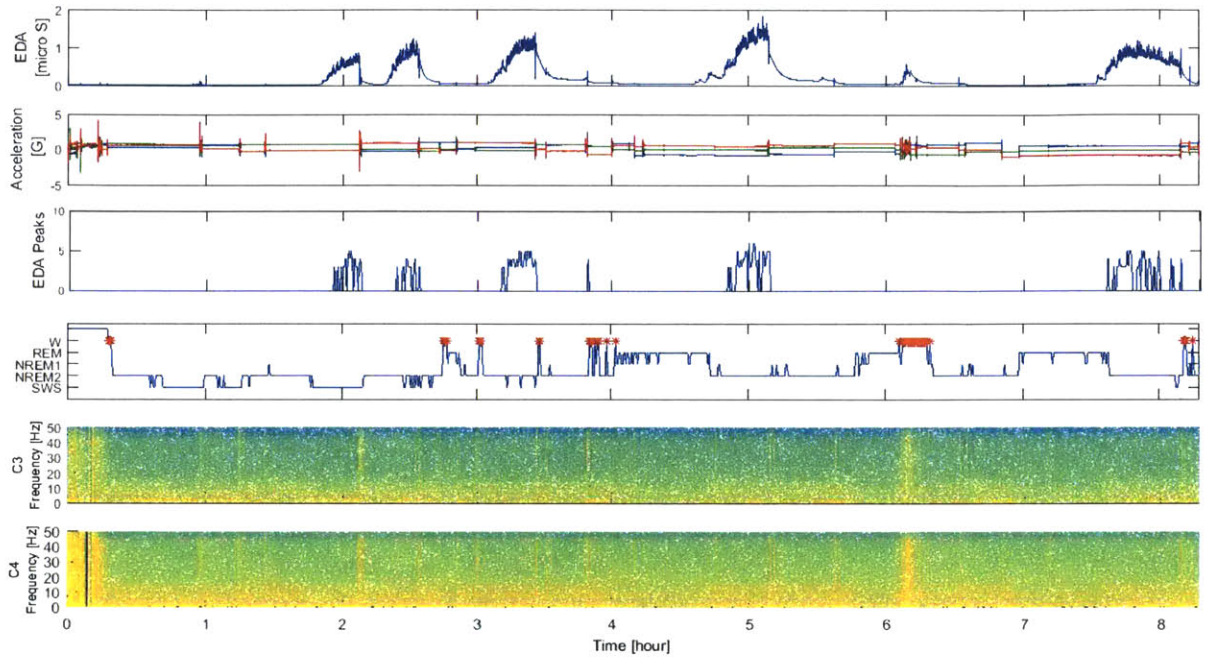


Figure 4.2 Raw EDA, Actigraphy (derived from 3-axis accelerometer data, red marks means wakefulness), detected regions with EDA peaks, manually scored sleep stages from EEG(red marks mean REM sleep), and two EEG channels for one night measured from a healthy adult.

We computed the following features for building a machine learning classifier.

a) EEG

We calculated power spectrum density of the frequency band (delta, theta, alpha, beta) of the quarters of the night for electrode locations C3 and C4. We also computed the features using the average amplitude of C3 and C4 over the whole night and per epoch.

b) EDA

The EDA was processed first by low-pass filtering (cutoff frequency 0.4 Hz, 32nd order FIR filter) before computing the features. We normalized the amplitude of the EDA by dividing all values by the maximum amplitude over the night, then obtained the first derivative of the filtered EDA, then determined where the slope exceeds a value of 0.5 micro Siemens per second. We detected EDA "peaks" based on those that exceeded this 0.5μS/s threshold and counted the number of peaks per each 30-second epoch. We also

computed the mean, standard deviation, median of the normalized EDA amplitude (normalized by the maximum EDA amplitude over the night) before sleep and during sleep. For EDA peaks we computed the total number for the night, the mean, standard deviation, and median of the number of EDA peaks per 30 s epoch over the night, the averaged number of peaks, the % of epochs with EDA peaks for each sleep stage and the mean # of EDA peaks per epoch. Previously, we have shown that EDA peaks are related to SWS or NREM 2 sleep [Sano, 2014a].

EDA data that corresponded to non-sleep epochs (as determined by actigraphy, see below) were removed from the analysis before computing features related to sleep.

c) Actigraphy or Accelerometer (ACC)

Sleep and non-sleep epochs were determined using standard zero-crossing detection and Cole's function applied to the accelerometer data [Cole, 1992]. From this motion information we further computed sleep latency, sleep duration, the % of wake in each quarter of the night, and the mean and standard deviation of the motion level.

d) Sleep stages

The sleep stages were scored by standard criteria [Rechtschaffen, 1968]. The features we used included the % of each sleep stage over the night, the sleep efficiency derived from the EEG (the percentage except wake and others during sleep), the time to first deep sleep, and the percentage of each sleep stage for each quarter of the night.

C. Classification

Figure 4.3 shows the distribution of performance improvement across the 24 participants x 3 nights = 72 nights.

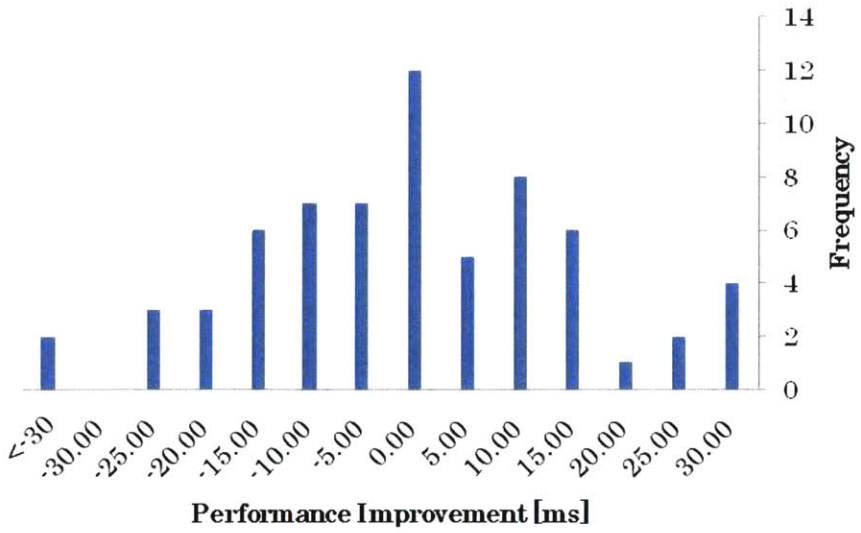


Figure 4.3 VDT performance improvement (N=24, 3 nights)

We grouped the participants into the following three groups.

1) the highest and lowest 33% of VDT improvement

2) the highest and lowest 20% of VDT improvement

3) the highest and lowest 20 % of VDT improvement only in hospital and laboratory nights (because we have PSG only for those nights).

For each of the three groups of data, we compared 6 methods:

A) Support vector machine with linear kernal
B) Support vector machine with Gaussian (radial basis function, RBF) kernel
C) Principal component analysis (PCA) and linear discriminant analysis
D) PCA and support vector machine with linear kernal
E) PCA and support vector machine with Gaussian (RBF) kernel
F) PCA and k nearest neighbors (k=1-5)

Each method was run with 4 variations of features (a-d) for the nights of hospital, laboratory and home and run with 7 variations of features (a-g) for only the nights of hospital and laboratory,

a) All features

b) EDA features only

c) ACC features only

d) EDA+ACC features

e) EEG features only

f) Sleep stage features only

g) EEG+EDA+ACC features

and compared classification accuracy with the 10-fold cross validation (trained the model with 90% of the data, tested with the remaining 10% and repeated this procedure for 10 times)

4.1.3 Results

As a baseline, because of prior published findings on how SWS and REM interact with VDT improvement, we examined the classification result using the percentage of SWS in the 1st quarter and percentage of REM in the 4th quarter of the night (Figure 4.4). The accuracy of predicting VDT performance is mostly around 60% or below it.

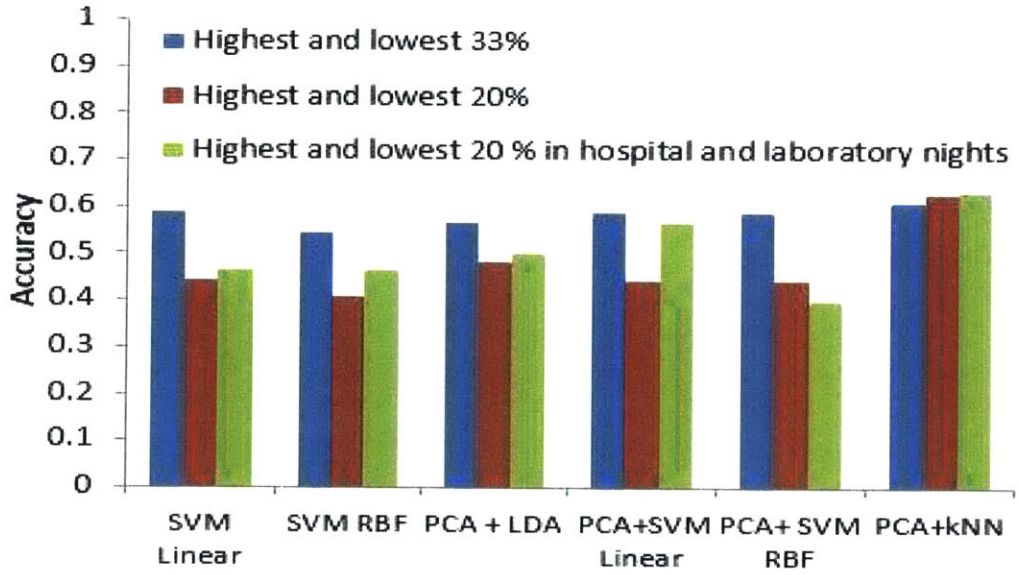


Figure 4.4.Accuracy of classification of VDT performance using the percentage of SWS in the 1st quarter and REM in the 4th quarter of the night

Figure 4.5 shows the comparison of classification accuracy using physiological features and classification methods for the highest and lowest 33% of VDT performance improvement. The features from EDA alone showed the highest accuracy, around 60-70% beating every one of the other three feature set combinations, while being tested on all six machine learning systems (these results were quantitatively proven with 10-fold cross validation even though we did not use statistical analysis).

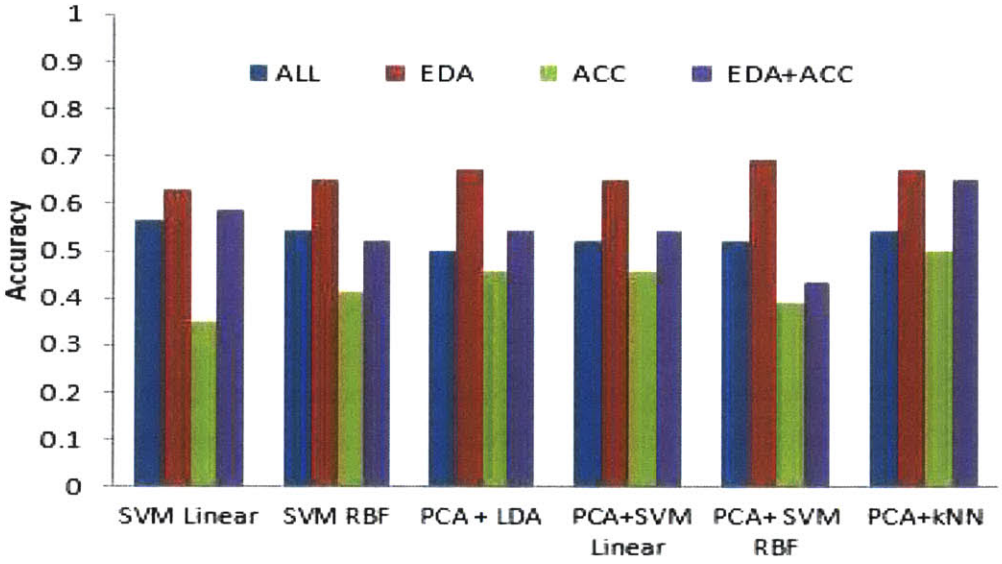


Figure 4.5.Accuracy of classification using the highest and lowest 33% of VDT performance improvement

Figure 4.6 shows the comparison of classification accuracy with features and classification methods for the highest and lowest 20% of VDT performance improvement. The features from EDA again showed the highest accuracy, 74%, this time either by appearing solo as the top performer or in three cases appearing in combination with accelerometer data.

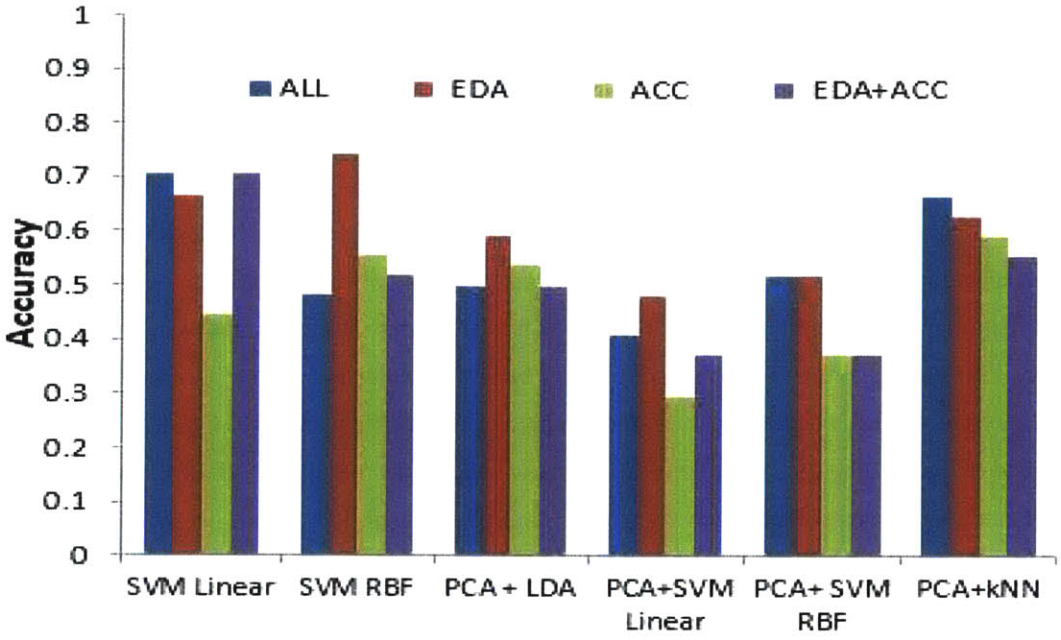


Figure 4.6 Accuracy of classification using the highest and lowest 20% of VDT performance improvement

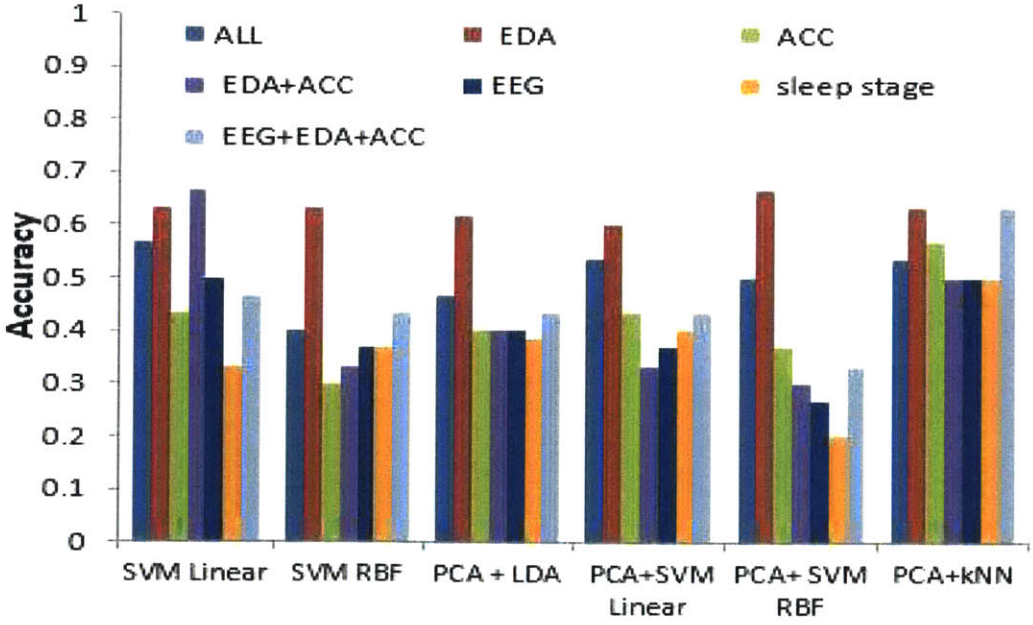


Figure 4.7 Accuracy of classification using the highest and lowest 20 % of VDT performance improvement only in hospital and laboratory nights

Figure 4.7 shows the comparison of classification accuracy with features and classification methods for the highest and lowest 20% of VDT performance improvement for the nights where we also had PSG at the hospital and laboratory. The features from EDA solo or EDA + ACC showed the highest accuracy, 67%, followed by EEG + EDA + ACC. Thus, EDA was again a part of all the top performing features.

In all the comparisons here, overall, either solo EDA features or EDA + ACC features improved the classification accuracy compared to use of sleep stages and to use of only EEG. In this paper, we applied PCA to reduce feature dimensions, but as a next step, we will apply feature selection methods to investigate relevant features in EDA and compare which features work better to predict the memory consolidation task. Why would EDA changes play a significant role in prediction of performance on a VDT task? This cannot be fully explained by the finding that EDA activity can show some signs of SWS or NREM2 sleep. While this question remains unanswered, this phenomenon is intriguing to look into more, in part because it's known that EDA during wake is increased with greater engagement and arousal, which then is believed to help predict memory. Now we also see that for sleep, the use of electrodermal physiology is showing patterns that invite greater exploration.

4.2 Comparison of Sleep-Wake Classification using Electroencephalogram and Wrist-worn Multi-modal Sensor Data

This study presents the comparison of sleep-wake classification using electroencephalogram (EEG) and multi-modal data from a wrist wearable sensor. We collected physiological data while participants were in bed: EEG, skin conductance (SC), skin temperature (ST), and acceleration (ACC) data, from 15 college students, computed the features and compared the intra-/inter-subject classification results. As results, EEG features showed 83% accuracy while features from a wrist wearable sensor showed 74%. The combination of ACC and ST played a more important role in sleep/wake classification by 30-s epochs than other subsets of the wearable sensor data.

4.2.1 Introduction

Sleep/wake identification has been used both in clinical fields and personal health/wellness fields. Clinically, polysomnography (PSG) has been used to monitor sleep and identify sleep disorders in sleep labs as a gold standard; however, it has disadvantages requiring the patient to stay one or more nights in the lab wearing uncomfortable sensors and wires. Actigraphy has been used to monitor long-term sleep wake cycles [Blackwell, 2011]. Cole et al. showed that sleep and wake are classified with an accuracy of 88% using wrist-worn actigraphy and regression analysis comparing the wrist data to sleep stages from PSG in bed with 1 min epoch data [Cole, 1992]. Some other researchers have applied machine learning or new algorithms to improve the accuracy [Sadeh, 1994][Tilmanne, 2009][Pollak, 2001] or used other data (heart rate variability from electrocardiogram (ECG)) [Elsenbruch, 1999][Lewicke, 2008]. Recently, many wearable devices have been on the market and most of them have multiple sensors (accelerometer, photoplethysmogram, etc). Due to advances in device technology, more wearable devices will come to the market with multi-modal sensors.

In this study, we compared the sleep/wake classification using physiological data taken while participants were in bed, using polysomnography (PSG) as a gold standard and skin conductance (SC), skin temperature (ST) and acceleration (ACC) data from a wristband sensor. We investigated which features from which modality play the most important roles in the sleep/wake classification.

4.2.2 Methods

A. Data Collection

Fifteen college students participated in sleep measurement in a hospital sleep laboratory. They wore electrodes for electrooculogram (EOG), for EEG on C3 and C4 (International 10-20 system), and for electromyogram (EMG) on their chin and a wrist sensor (Q Sensor by Affectiva) to measure SC, ST, and ACC on their dominant hand. The EEG, ECG and EMG were sampled at 200 Hz and the wristband data were sampled at 8 Hz. Sleep stages were scored for each 30-s epoch sleep data based on standard PSG criteria measuring EEG, EOG and EMG [Rechtschaffen, 1968]. Figure 4.8 shows a sample representation of one night's data from one participant. The experimental procedure was pre-approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology.

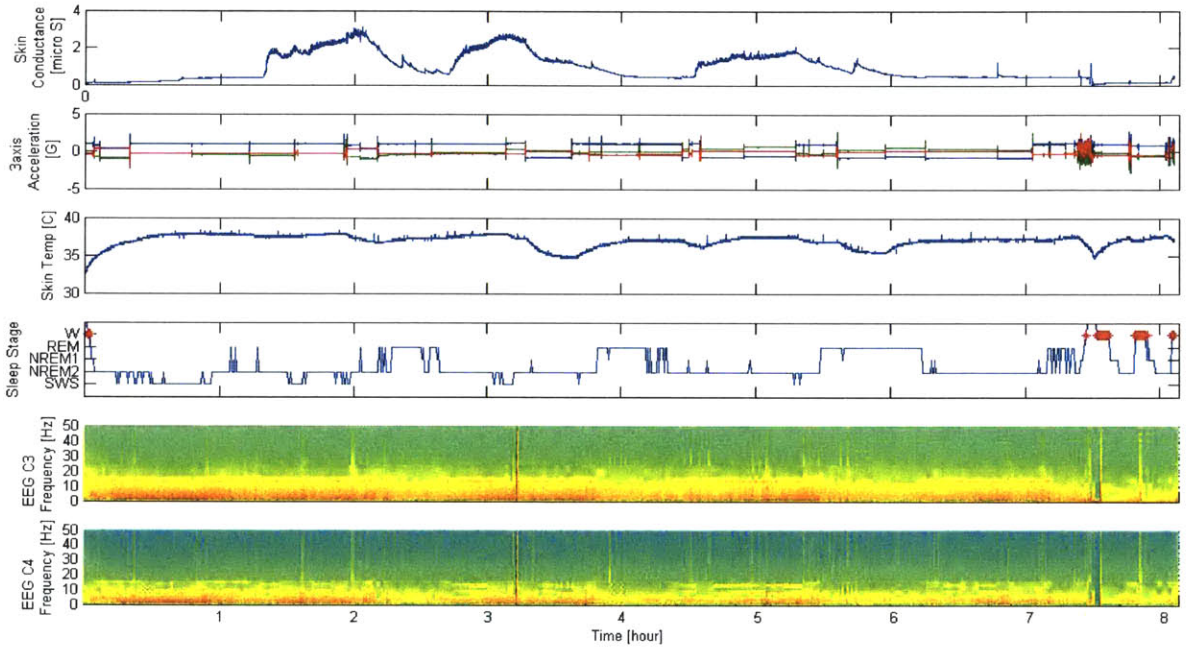


Figure 4.8 Raw skin conductance, 3-axis accelerometer data, skin temperature, manually scored sleep stages from PSG (red marks mean wakefulness), and EEG spectrogram (channels C3 and C4) for one night for a healthy college student

B. Feature Extraction

We computed the following features for building machine learning classifiers.

a) Electroencephalogram (EEG) (16 features were computed per epoch)

We calculated the z-score of the power spectrum density for the frequency bands (delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz) and beta (12-30 Hz) over the night for averaged EEG at electrode locations C3

and C4. We then computed the average, standard deviation, maximum and minimum of the z-scored power spectrum density of averaged EEG at electrodes C3 and C4 per epoch.

b) Skin Conductance (SC) (7 features per epoch)

First, for de-noising, we low-pass filtered the SC data (cutoff frequency 0.4 Hz, 32nd order FIR filter) before computing the features. We normalized the amplitude of the SC in a range between minimum and maximum amplitude over the night, then obtained the first derivative of the filtered SC. We detected SC "peaks" based on those that exceeded 0.02 μS/s threshold and counted the number of peaks per each 30-second epoch. Our previous study has shown that SC peaks are much more likely to occur during SWS or Non-REM (rapid eye movement) sleep [Sano, 2011]. We also computed the mean, standard deviation, median, maximum and minimum of the normalized SC amplitude (normalized by the maximum and the minimum SC amplitude over the night) and gradient from linear least square fitting for each 30-s epoch. For SC peaks, we computed the total number for each 30-s epoch and the standard deviation of the number of SC peaks per 30 s epoch over the night.

c) Acceleration data (ACC) (7 features for an epoch)

We applied a 2-3 Hz band pass filter to the accelerometer data and then counted the number of times of the three axis amplitude root mean square (RMS) values crossed 0.01 (the number of zero-crossings) for each 30-s epoch. We then applied Cole's "D" function to score "wake" or "sleep" for each 30-s epoch [Cole, 1992]. We also computed the mean, standard deviation, maximum and minimum of the root mean square of three axis acceleration data for each 30-s epoch.

d) Skin Temperature (ST) (5 features for an epoch)

We normalized the temperature data using the maximum and minimum values over the night and computed the average, standard deviation, maximum, minimum and gradient from linear least square fitting for each 30-s epoch.

B. Classification

We grouped SWS, non-REM2, non-REM1 and REM into sleep. As we have more sleep epochs than wake epochs given all the data was from lying in bed, we extracted sleep epochs randomly to equalize the number of sleep and wake samples.

We defined the following 6 methods (A-F), 2 datasets (1-2) and 15 feature combinations (a-o) for systematic evaluation.

Classifiers:

A) Support vector machine with linear kernel (SVM linear)

B) Support vector machine with Gaussian (radial basis function) kernel (SVM RBF)

C) K-nearest neighbor (kNN, k=1-4)

D) Feature selection (exhaustive feature selection by maximizing the J3 measure associated with the scatter matrices (maximize separability of the features from 2 classes) [Theodoridis, 2010] to find the best 2-6 features for EEG or data from the wrist sensor) and SVM linear

E) Feature selection (exhaustive) and SVM RBF

F) Feature selection (exhaustive) and kNN (k=1-4)

Data-sets

1) Intra-subject classification

Within each participant, we identified 11-98 epochs from each of wake and sleep. We trained the models using 90% of the data, tested with the remaining 10% of the data and repeated this procedure 10 times, each time leaving out a different 10% of the data (10-fold cross validation). Then, we reported the mean accuracy.

2) Inter-subject classification

We divided the data (total # of epochs = 661) into 10 sets, performed training with the one set from all except one subject data, tested with one set of the remaining one subject data and repeated this procedure 10 times.

Feature sets

a)EEG, b)EEG+ACC, c)EEG+SC, d)EEG+ST, e) EEG+ACC+ST, f) EEG+SC+ACC, g)EEG+SC+ST, h)EEG+ACC+SC+ST, i)SC, j)ACC, k)ST, l)ACC+SC, m)SC+ST, n)ACC+ST, o)ACC+SC+ST

We compared the classification results for the following combinations.

1) Intra-subject + Feature sets a - o + Classifiers A - C

2) Inter-subject + Feature sets a - o + Classifiers A - C

3) Inter-subject + Feature sets a - o + Classifiers D – F

4.2.3 Results

A. Intra-subject classification

Figure 4.9 shows the classification comparison for intra-subject data-sets. EEG showed 91% and EEG + other features boosted the accuracy to 95% (EEG+SC+ACC) and 96% (EEG+ACC+SC+ST). The features from only the wristband sensor showed 86% (ACC+SC+ST) and 84% (SC+ST, ACC+ST). Of the single wristband features, ST was the best (79%) followed by SC (75%), with ACC the lowest (67%).

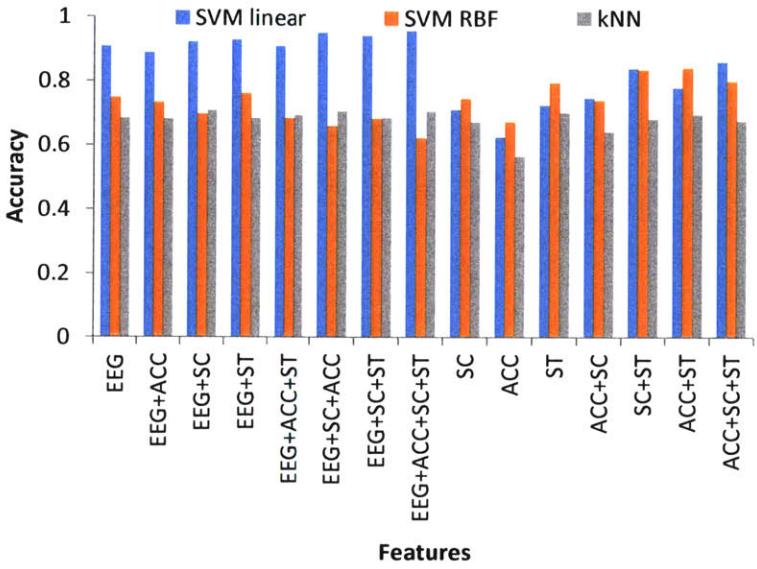


Figure 4.9 Accuracy of intra-subject classification.

B. Inter-subject classification

Figure 4.10 shows the classification comparison for inter-subject data sets. All of the features EEG+ACC+SC+ST and EEG+SC+ACC showed the best classification rates (85%). Of the wrist features, ACC+SC+ST and ACC+ST showed 74%. Of the single wristband features, ACC was the best (68%), followed by ST (67%) with SC the lowest (51%). Thus, we found ACC worked better when looking across subjects than within subjects. Figure 4.11 shows the Receiver Operating Characteristic (ROC) curves of EEG features vs wrist features. Of the wrist features, ST was dominant.

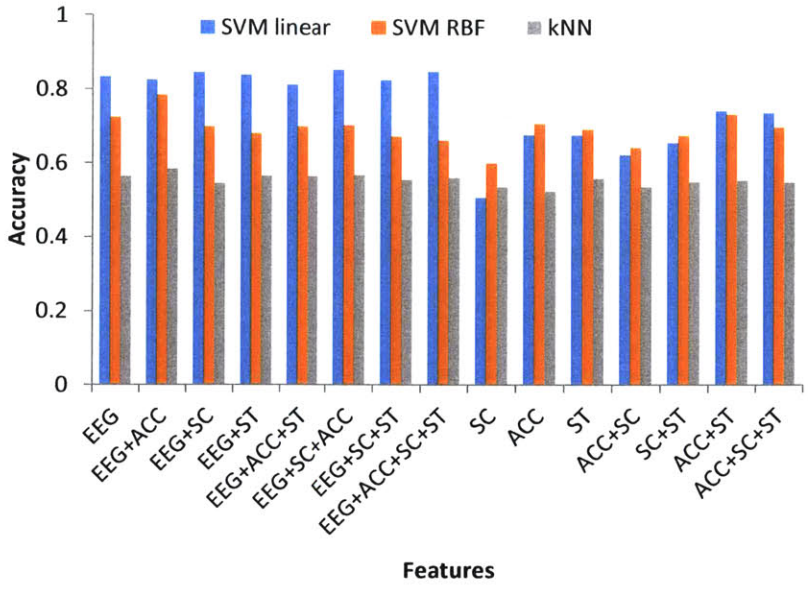


Figure 4.10 Accuracy of inter-subject classification

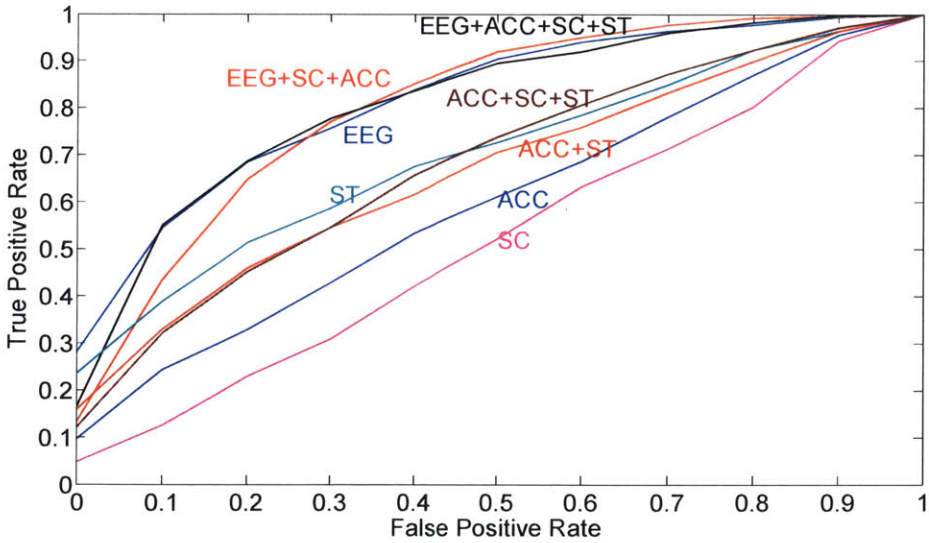


Figure 4.11 Inter-subject ROC curves for sleep-wake classification with EEG+ACC+SC+ST, EEG+SC+ACC, ACC+SC+ST and ACC+ST

C. Inter-subject classification with selected features

We applied exhaustive search over sets of sizes 2-6 features for each modality and compared the classification rate. Table 4.1 shows the best feature subsets for each modality for each classifier with the

highest classification rates. Some of the features were chosen multiple times by different classifiers. The best EEG features and the best wrist features showed 83% (EEG features: mean (Z-score, theta), mean (Z-score, alpha), SD(z-score, alpha), SD(z-score, beta), min(z-score, alpha)) and 73% (ACC: Coefficient from the Cole's function and ST: mean, min and gradient (normalized ST)) respectively.

4.2.4 Discussion

In all of the classifications, as expected since EEG is part of the sleep/wake ground truth in PSG, the EEG features showed the best accuracy - over 85%. The ACC features performed much lower, only 67-68%, which is lower than Cole reported. This may be because we used 30s epochs and Cole et al used one minute epochs. Furthermore, we did not apply Webster's rescoring methods [Webster, 1982] which serve as a filter that allows rescoring of sleep scores as wake scores when short periods scored as sleep are bounded by longer waking intervals on either side. Jean-Louis et al used different actigraphy than the one Cole et al used and showed that the rescoring rules can decrease the sensitivity of sleep detection but increase the specificity and that they improved the accuracy in detecting sleep and wakefulness in 24-h recordings, but not for in-bed recordings [Jean-Louis, 2001]. In addition, we did not use previous epochs to estimate current state and we also used a different source of ACC. As next steps, we can customize and apply rescoring methods for our dataset or test temporal machine learning models or features. SC was very person-dependent, performing 75% when it had some training data from the subject it was testing on, but only 51% when trained and tested on different people. Some SC features related to sleep that have been reported as robust such as decrease in palm skin potential level after sleep onset [Hori, 1982] were not included in our data. Also, SC storms in non-REM sleep [Sano, 2014a] are not represented in but 51% of the randomly selected sleep data in our tests, while SC tends to be active in wake before and after sleep; thus, we are not surprised that use of SC in only a 30-sec epoch is not a strong indicator of sleep vs. wake when judged by the score of only epochs. Skin temperature showed an important role in epoch sleep/wake classification. Kräuchi et al. indicated distal skin temperature increase at sleep onset and dramatic decrease at wake up [Kräuchi, 2004]. Although we maintained the balance between the number of wake epochs and sleep epochs for training and testing data, the quality of the dataset could improve if we collected more of the day's data over wake epochs, and used this larger set of data. Once again, in this chapter, we aimed to classify the sleep/wake states using physiological data taken while participants were in bed. Sleep-wake detection from 24/7 wearable sensor data is another problem to solve, which our research group is also working on separately from this thesis.

4.2.5 Conclusion

We compared sleep-wake classification accuracy with EEG from the scalp, and ACC, SC and ST from a wrist sensor, taken while participants were in bed. We applied three types of machine learning, paired with feature selection methods, in order to identify features that best discriminated sleep and wake. We found EEG features showed 83% accuracy while features from a wrist wearable sensor (the combination of ACC and ST showed 74% in inter-subject classification. This result implies that the combination of ACC and ST is better than ACC solo which is currently used in commercialized actiwatches.

Table 4.1 Summary of the best features (Features with higher grand total are more robust regardless of classifiers).

Modality	Features computed for every 30sec epoch	SVM linear	SVM RBF	kNN	Grand Total	
EEG	$\mu(z\text{-}\delta)$	1			1/3	0.33
	$\mu(z\text{-}\theta)$	1	1	1	3/3	1
	$\mu(z\text{-}\alpha)$	1	1	1	3/3	1
	$\mu(z\text{-}\beta)$	1			1/3	0.33
	$\sigma(z\text{-}\delta)$	1		1	2/3	0.66
	$\sigma(z\text{-}\theta)$	1		1	2/3	0.66
	$\sigma(z\text{-}\alpha)$	1	1	1	3/3	1
	$\sigma(z\text{-}\beta)$	1	1	1	3/3	1
	max(z-δ)	1			0/3	0
	max (z-θ)				0/3	0
	max (z-α)	1			1/3	0.33
	max (z-β)	1			1/3	0.33
	min(z-δ)	1			1/3	0.33
	min (z-θ)				0/3	0
	min (z-α)	1	1	1	3/3	1
	min (z-β)	1			1/3	0.33
ACC	# of ZC	1			1/3	0.33
	coefficient from the Cole's function	1	1	1	3/3	1
	W/S score from Cole's function	1	1		2/3	0.66
	μ (RMS of 3axis ACC)				0/3	0
	σ (RMS of 3axis ACC)				0/3	0
	max (RMS of 3axis ACC)		1		1/3	0.33
	min (RMS of 3axis ACC)		1		1/3	0.33
SC	# of SC peaks		1		1/3	0.33
	μ (normalized SC)		1		1/3	0.33
	M (normalized SC)		1		1/3	0.33
	σ (normalized SC)		1		1/3	0.33
	max (normalized SC)		1		1/3	0.33
	min(normalized SC)		1		1/3	0.33
	Gradient	1		1	2/3	0.66
ST	μ (normalized ST)	1	1	1	3/3	1
	σ (normalized ST)				0/3	0
	max (normalized ST)				0/3	0
	min(normalized ST)	1	1	1	3/3	1
	Gradient	1	1	1	3/3	1

Chapter 5
SNAPSHOT Study: Study Design and Measurement

Now I have completed describing the introduction, related work and some of our previous projects related to the new kinds of methods and measures developed in this thesis. In this chapter, I describe the design, execution and high level research questions of the SNAPSHOT Study, a large new study that seeks to measure: **S**leep, **N**etworks, **A**ffect, **P**erformance, **S**tress, and **H**ealth using **O**bjective **T**echniques.

This study was initially designed to investigate how interactions in a social network influence sleep behaviors. We also added measurement of other multi-modal factors to investigate how daily behaviors influence sleep, stress, mood, and other wellbeing-related factors.

Each semester, we study N=50 MIT undergraduate students who are socially connected (calling or texting at least once a week, with a minimum # of people in a group >=5) for about 30 days (Figure 5.1). So far we have collected data from N= 168 over 3 semesters and a pilot study. We collected over 100,000 hours of continuous ambulatory measurements together with surveys and measurements in the laboratory to obtain subjective and objective labels about sleep habit, stress, academic performance, social interactions, mood, sickness and circadian phase (dim light melatonin onset) (Table 5.1). Details of recruitment (section 5.1) and measures (section 5.2) will be described in the following sections.

Table 5.1 Our measurement (more details will be described in section 5.1.2)

Objective measurement	Wearable Sensor (continuous)	Skin conductance Skin temperature 3-axis acceleration Light Exposure
	Mobile Phone (continuous)	Call, SMS logs (only timestamps and phone numbers), Screen on/off timing Location Application usage
	Email (continuous)	Email logs (only timestamps, to, cc)
	Lab measurement (once during 30 days)	Melatonin Cognitive performance Stress task responses
Subjective measures (Surveys)	Pre-study (once)	8 surveys about demographics, morningness-eveningness, sleep habit, personality, stress, mental and physical health and social interactions
	Daily survey (every morning and evening)	16 morning questions, 18 evening questions about academic, exercise, extracurricular activity, sleep, caffeinated/alcoholic drink and drug intake, social interactions, wellbeing-related measures (alertness, happiness, energy, healthiness and calmness)
	Post-study (once)	5 surveys about stress, anxiety, mental and physical health and social interactions

Figure 5.1 Study timeline

The Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects (COUHES) approved this study and all participants gave informed consent.

5.1 Recruitment

We intentionally recruited a group of MIT undergraduate students who were socially connected. Our definition of "socially connected" was "making a call or SMS at least once a week" and we recruited a group of at least 5 people who knew each other and interacted socially. We posted our study advertisement to undergraduate students' mailing lists and the potential participants filled out the screening questionnaire. Our exclusion criteria were the following

- Non-Android phone users
- People who have a problem wearing wrist sensors (e.g. irritated skin on wrist, etc).
- Pregnant women
- People who have traveled more than one time zone away one week prior to the study or have plans to travel more than one time zone away during the study
- People under 18 years and over 60 years old

In our study, we targeted only Android phone users because other smart phones (e.g. iPhone) did not allow us to monitor phone usage in detail. Specifically, our participants consented to let us know whom they communicate with over calls and SMS, which we needed to create social network models for each participant.

After we screened the potential participants, we invited the screened participants to the SNAPSHOT information and consent sessions. For each session, we invited about 15 participants and explained about the study and tasks participants will do during the study period. After participants signed a consent document on joining the study, they filled out pre-study questionnaires, started wearing devices, and installed an Android application on their phone. Details of measurement will be explained in the next section.

5.2 Data Collection

Since we started the SNAPSHOT study in fall 2013, we have recruited 169 participants (age: 18-25, 19.6 ± 1.5, male: 107). 7 participants dropped out before completing the study, leaving us with 162 who completed the full 30-day SNAPSHOT. (See Appendix B and C for more details about demographics about participants)

Pre-study Questionnaires

Prior to the study, participants completed the morningness-eveningness questionnaire [Horne, 1976], the Pittsburgh sleep quality index test [Buysse, 1989] to determine their habitual sleep patterns, the Myers Brigg Personality test and the Big Five Inventory Personality Test [John, 1999] to understand their personality factors, the Perceived Stress Scale (PSS) to understand their stress level, SF-12 to understand their physical and mental health condition, and a set of social network surveys to help us map their social networks. We also collected age, sex, academic major and living situation (both living group name and whether single or multiple occupancy room).

Ambulatory Monitoring

Throughout the 30 days, participants wore two wrist band sensors: Q-sensor (Affectiva, USA) to measure EDA, skin temperature, 3 axis acceleration on their dominant wrist and Motion Logger (AMI, USA) on their non-dominant wrist to measure acceleration and ambient light data, taking them off only to shower, swim, or when privacy was desired. In addition, an Android phone application that we modified based on funf [Aharony, 2011] monitored location, receiver, sender and timing of calls and text messages, screen on/off timings, and use of mobile phone applications on the mobile phone. Participants also kept daily morning and evening diaries about sleep and wake times, nap, exercise, academic and extracurricular

activity times, social interactions, caffeine, alcohol, and drug intake, overall health condition, sleep, mood, and stress each day upon awakening and at bedtime (See Appendix). Participants also signed a certificate of confidentiality when they consented on joining the study.

Acceleration was measured to estimate activity levels and sleep/wake patterns. Skin conductance (SC) was measured because it represents autonomic arousal during the day and provides a stress index; its responses during sleep are highly likely to occur in either non-REM Stage 2 sleep or Slow Wave Sleep, and help to characterize sleep better than using only acceleration data from actigraphy in some cases [Sano, 2014a]. Skin temperature also helps to understand sleep/wake patterns [Sano, 2014b], while acceleration helps show activity and sleep patterns. We hypothesize that physiology combined with daily behavior data can be used to predict aspects of sleep behaviors, academic performance, and self-reported stress and mental health better than any of these measures alone.

Phone and email usage was measured for two main reasons: First, lighting from the interaction with mobile phones or emailing late at night could disturb the biological circadian clock and increase alertness, both of which can influence sleep patterns [Cajochen, 2011, Chang, 2014]. Second, phone and email usage and location data give clues to sociability. The timing of calls, SMS, emails and "screen on" provide an estimate of how often participants interact with their phone during the day and the night, while the number of calls, SMS and emails and the number of people they interact with helps quantify their social interaction.

Melatonin Assessment of Circadian Phase

Circadian phase is a potent influence on human sleep timing and content. The dim light melatonin onset (25% of the fitted nighttime peak), is a highly robust marker of circadian phase [Lewy, 2007]. Once in the 30 day study period, participants spent a night at the Brigham & Women's Hospital. Saliva samples were collected every 60 minutes from late afternoon to the next morning in dim light conditions, beginning 8 hours before normal bedtime to calculate circadian phase for each participant.

Post-study Questionnaires and Other Measures

In addition to the above measures, academic performance (overall GPA) in the previous and at the end of the study semester were collected by self-report from each participant. Email usage during the experiment (to, from, cc and timestamps) was collected through the MIT website *Immersion* (https://immersion.media.mit.edu/) at the end of the study. In addition, based on their phone call, SMS and email usage during the experiment, participants were asked to identify these most frequent contacts and whether they had positive/neutral/negative interactions with these contacts and which category they belong

to (family/social/work/others) as a whole over the month. At the end of the study, participants filled out the PSS, SF-12 and social network surveys again and also the State-Trait Anxiety Index [Spielberger, 1983].

Chapter 6
Data Cleaning, Pre-processing, Feature Extraction and Data Characteristics

In this chapter, we describe data cleaning and pre-processing methods and characteristics of our data.

6.1 Data Cleaning and Pre-processing

6.1.1 Sleep/wake Scoring and Sleep Regularity

Sleep/wake onsets were determined by the Brigham & Women's Hospital (BWH) team using a combination of wrist actigraphy and sleep diaries and their written standard operation procedure. In addition to bed time, wake time and regularity, BWH's team defined and computed sleep regularity as a value of 0 - 1 based on the likelihood of sleep/wake state being the same time-points 24 hours apart (equation 6.1). Sleep regularity is another index to evaluate sleep wake patterns which is not evaluated with conventional sleep surveys such as PSQI. Some studies have pointed out the importance of this sleep regularity measure [Clerx, 2014, Clerx, 2015, Bei, 2015] in addition to sleep duration.

$$Sleep\ regularity\ index = \frac{1+\frac{1}{T-\tau}\int_0^{T-\tau} s(t)s(t+\tau)\,dt}{2} \qquad \text{(eq. 6.1)}$$

,where s(t)=1 during wake and s(t)= -1 during sleep.

Suppose data are collected for y=[0, T]. Choose τ =24.

Figure 6.1 shows the examples of the most regular and irregular participant in the study.

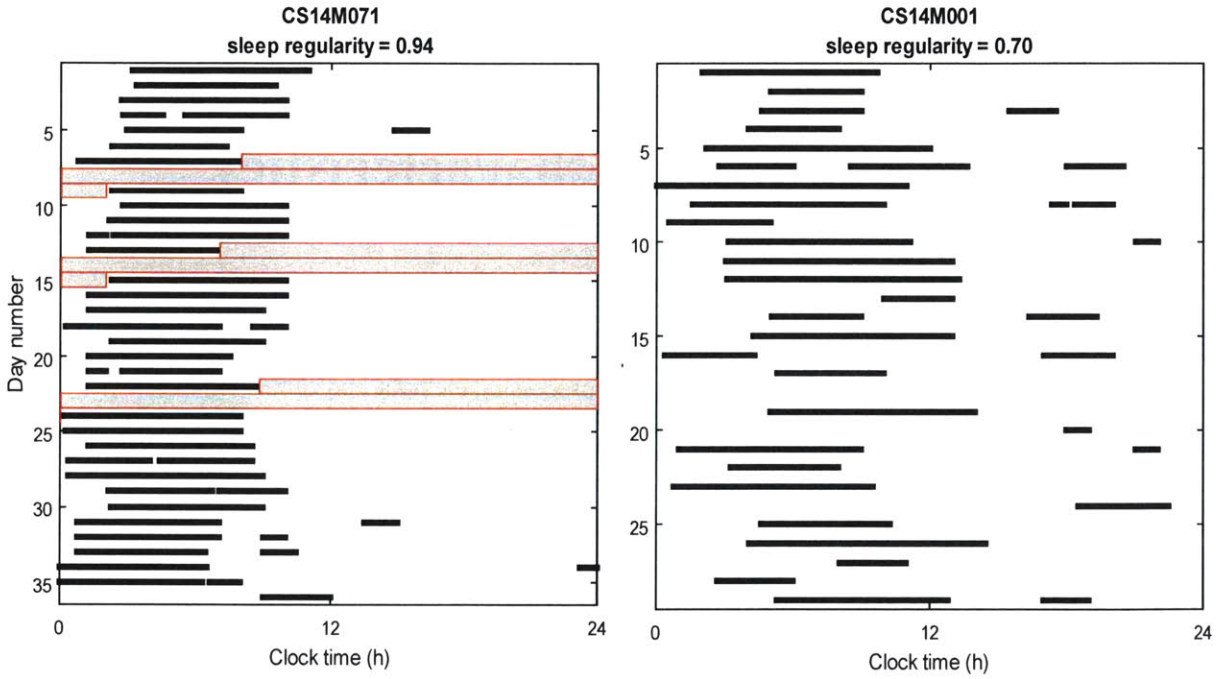


Figure 6.1 Raster plots of the most regular and the most irregular sleepers in our study

(black lines: sleep episodes, red lines shows missing data)

6.1.2 Separation of Ambulatory Electrodermal Activity in Day and Sleep Activities based on Activity Magnitude and Sleep-Wake Scoring

With ambulatory measurement systems, researchers have started measuring 24/7 EDA data in daily life and identifying EDA features related to stress and sleep [Muaremi, 2014, Sano, 2015]. EDA responses can be elicited from multiple processes, including thermoregulation, motor, and affective processes. In order to understand long-term ambulatory EDA, we separate daytime and sleep EDA. Within the daytime, we also separate exercise vs non-exercise-related EDA activity. A lot of research has been conducted to recognize

activities [Chernbumroong, 2011]. In this work, we aim to obtain rough estimate of three activity levels (sitting, walking and running) in a simple way and leverage it to compute EDA features under different activity levels. We used wrist acceleration data (ACC) and identified activity magnitude thresholds to separate daytime activity into sitting, walking and running levels. We compare ambulatory EDA amplitude and peaks under sitting, walking, running level activities and 1st-4th quarters of sleep.

Methods

In order to train a classifier for sitting, walking and running, we collected non-dominant outer wrist 3-axis ACC data from N=68 participants using the Q-sensor (Affective, USA) while participants went through the following experimental procedure 1) sit still and watch a relaxing video for 5 minutes 2) sit and fill out surveys 3) sit and perform a "counting backwards by 7's" and stroop tasks for 5 minutes 4) walking 5) running. We computed activity magnitude (AM) using the equation 6.2 and drew histograms of mean activity magnitude under 5 different tasks. Then, we applied a maximum likelihood decision rule to identify thresholds to separate sitting (1-3), walking and running distributions. We applied these thresholds and sleep-wake scorings (from activity data and sleep-wake diaries) to ambulatory ACC and EDA data from N=20 people collected over ~30 days per person, for a total of 600 days. We compared EDA amplitudes and number of peaks per 30s epoch for sitting, walking and running and for the first through fourth quarters (1Q-4Q) of each night's sleep (see the detailed peak detection method in [Sano, 2014a]).

$$AM = \sum_{t=0}^{N} AM_t + \sqrt{(Raw_{xt} - Rm_x)^2 + (Raw_{yt} - Rm_y)^2 + (Raw_{zt} - Rm_z)^2} \qquad \text{(eq. 6.2)}$$

where AM = Activity Magnitude, Raw = Raw accelerometer sample

Rm = running mean in a previous 5 seconds window, N = number of raw data samples received in one second

Results

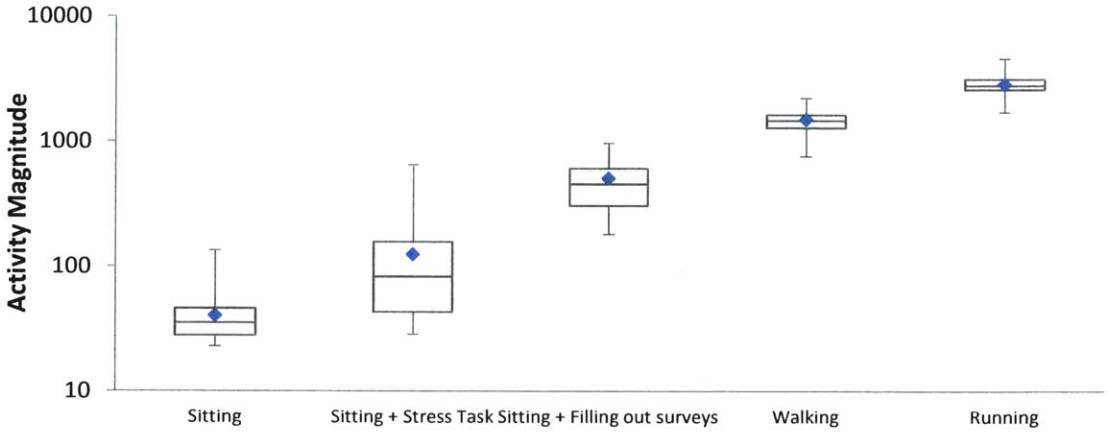


Figure 6.2 Boxplots of activity magnitude

Figure 6.2 shows boxplots of the activity magnitude from 5 activities. We obtained values 966 and 2059 for sitting-walking and walking-running thresholds. We compared the mean of EDA amplitude and # of peaks (Figure 6.3). EDA amplitude and peaks increased as activity level got higher. For sleep, we found the highest amplitude in 2Q sleep and the largest number of peaks in 1Q sleep (ANOVA, post-hoc).

Figure 6.3 Comparison of ambulatory EDA amplitude

EDA artifact detection

We collected 24/7 EDA data except during times when they took a shower or were at risk of breaking their sensor. We applied the EDA artifact detection algorithm developed by our research group at MIT [Taylor and Jaques, 2015] to 5 second epochs of EDA data and obtained a label of clean/artifact/unknown output

for each 5s epoch. Using these 5s labels, we defined 30s epochs as clean or artifact: if we had artifact epochs for more than 15 seconds, that 30s epoch was defined as artifact.

6.2 Feature Extraction

Table 6.1 shows 183 features extracted from the collected data. We computed both monthly-averaged features and daily features from 30 days of data. In this thesis, we computed only features we can interpret relatively easily for physiology in contrast with more non intuitive features in our previous paper [Sano, 2015].

EDA was processed with the same method we used in chapter 3. For ACC data in wakefulness, we separated the data into sit, walk and run episodes based on thresholds we computed in the previous section and computed EDA features for sit, walk, and run episodes in order to separate EDA responses into psychological and activity-related ones. Then, we extracted mean, median and SD of the whole day's ACC, EDA and TEMP patterns. In previous work, we tried very detailed features from histograms and power spectrum density data of each signal [Sano, 2015]; however, in this thesis, we focused on simple features that can be interpretable.

Table 6.1 shows the features we computed:

Table 6.1 Features

Surveys	
Sleep (11 features)	Pittsburg Sleep Quality Index score, Morningness Eveningness Score, sleep time, wake time, sleep latency, sleep regularity, how they wake up (alarm or spontaneously), # of awakenings, duration of awakenings, # of naps, duration of naps
Stress (2 features)	Perceived Stress Scale (PSS) (pre-study and post-study)
Anxiety (2 features)	State and Trait Anxiety Score
Personality Traits (5 features)	Big Five Test (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism)
Physical and Mental Health (4 features)	Physical and mental health composite scores (PCS and MCS) from SF-12 (pre-study and post-study)
Academic Performance (2 features)	Grade point average (GPA) (previous and current semesters)
Wellbeing-related measures (10 features)	Alertness, happiness, sluggishness, healthiness and calmness when wake up and before sleep (0-100 scales)
Social Interactions (14 features)	Social interactions before sleep (with person in person or through electronic devices), frequency of memorable positive and negative and very negative social interactions, and social interactions in the past one month (% of interactions through face to face, email, SMS and phone, % people with positive, neutral and negative interactions, % of family members, friends, work-related colleagues each participant interacted frequently in the past one month)
Activities (5 features)	Total hours of academic(including classes, e-classes, sections, seminars, labs, study groups), study(studying alone), exercise (including sports, gym, cycling, etc), and extracurricular activities, last intake time of caffeinated drinks

Table 6.1 Features (cont)

Monitoring using phones or wearable sensors	
Email (10 features)	Total # of sent emails, mean and SD of # of daily received/sent emails, # of people emails were addressed to, mean and SD of timestamps of received and sent emails
Phone (CALL) (6 features)	Mean, median, SD of timestamp of each call, duration for each call, total # of people called, incoming call %
Phone (SMS) (8 features)	Mean, median, SD of timestamp of each SMS message, total # of SMS messages, total # of people SMS messaged, Incoming SMS %, 0-3am outgoing SMS%, 3-6am outgoing SMS%
Phone (Screen on/off) (10 features)	Time of each screen on/off, total # of screen on/off, total duration % of screen on between 0-3am, 3-6am, 6-9am, 9am-12pm, 12-3pm, 3-6pm, 6-9pm and 9pm-0am
Phone (MOB: mobility) (3 features)	Total distance per day and median and standard deviation of the distance per day
Wearable sensor (ACC) (31 features)	Mean % of sit, walk and run activities per day, mean, median and SD of RMS values for day time, sit, walk, run and entire and 1-4Q sleep and mean objective sleep quality from actigraphy
Wearable sensor (EDA) (27 features)	Mean, median and SD of amplitude for day time, sit, walk, run and entire and 1-4Q sleep
Wearable sensor (ST) (27 features)	Mean, median and SD of amplitude for day time, sit, walk, run and entire and 1-4Q sleep
Wearable sensor (light) (8 features)	Light exposure mean and SD for 0-6am, 6am-12pm, 12pm-6pm, 6pm-0am

6.3 Characteristics of Our Population

In order to understand the characteristics of our population, we describe statistics of our measurement from our population. Table 6.2 shows statistics of general parameters (N=168). Big five personality test results showed that our participants are less open, more conscientious and less neurotic than the average (50).

Table 6.2 Characteristics of populations (N=168)

Parameters	Mean	Median	SD
Age	19.63	19.00	1.53
BMI	23.78	22.54	4.67
Openness	42.48	41.00	27.20
Conscientiousness	53.27	58.00	28.53
Extraversion	49.64	48.00	30.03
Agreeableness	50.06	50.00	28.41
Neuroticism	35.17	27.00	27.73
GPA previous semester	4.39	4.50	0.62
GPA for the semester	4.30	4.40	0.56

6.3.1 Characteristics of Sleep Related Behaviors, Stress, and Health

Table 6.3 shows characteristics of sleep related behaviors, stress and health (N=168). The averaged bed time was 02:24 AM and wake time was 09:42 AM. On average, they slept for 7.0 hours; however, if we look at the distribution of their daily sleep duration (Figure 6.4), 23.5% of the nights were less than 6 hours, 3.4% was the average percentage of days each participant had no sleep, and the maximum days a participant did not sleep was 16.7% = 5 days per month. Their averaged PSQI score was 4.55: 49 % of the participants are considered to be a poor sleeper (PSQI >=5). The averaged Owl & lark score was 42.8 and it was still within neutral type but towards evening type (extreme morning type 70-86, moderately morning type 59-69, neutral type 42-58, moderately evening type 31-41, extreme evening type 16-30). Thus, 43% of the population was evening type. Within 15 minutes before their bed time on 35% of their study days, they interacted with people through electronic media and on 58% of their study days, they interacted with people in person.

Their average PSS was 14.8 in the pre study survey and 17.3 in the post study survey. Thus, 45 % of the population was more stressed than the average for their age at the beginning of the study and it raised up to 60% at the end of the study (14.2 is the average PSS over this age group). The averaged PCS was

slightly higher than 50; that means our population was slightly physically healthier than average; however the averaged MCS was lower than 50, which means their mental health scores were lower than average. Detailed answers in SF-12 related to MCS are illustrated in Table B.7. Fifty-seven % of our participants answered they felt downhearted and blue during the past 4 weeks either "some of the time", "most of the time" or "all of the time" at the end of the study. Forty-five % of our participants said either "some of the time", "most of the time", or "all of the time" to the question "During the past 4 weeks, how much of the time have you had a problem that you accomplished less than you would like with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?". In addition, the averaged MCS difference was smaller than the averaged PCS. That means more participants experienced reduced mental health scores at the end of the study.

We showed the distributions of daily subjective measures in Figure 6.5-6.9. Most of the distributions were skewed and not Gaussian. Mean alertness was lower in evenings than in mornings. The distributions of alertness, happiness and energy were different in mornings and evenings (Kolmogorov-Smirnov test, p= 8.27e-54, 0.0090, 5.54e-12). Both morning and evening healthiness and calm distributions were from the same distributions.

Table 6.3 Characteristics of sleep behaviors, stress and health (N=168)

Category	Parameters	Mean	Median	SD
sleep	Regularity	0.85	0.87	0.11
	PSQI Score	4.55	4.00	2.20
	Owl & lark	42.76	43.00	9.05
	DLMO	23.11	23.08	1.74
	Sleep quality (from actiwatch)	94.86	95.64	2.70
	Probability of pre_sleep_activity interacting with people through electronic media (e.g. emails, calls, SMS, skype, chat, online games)	0.35	0.27	0.27
	Probability of pre_sleep_activity interacting with people in person	0.58	0.61	0.26
	Spontaneously awoke	0.40	0.38	0.20
	Awoken by an alarm	0.52	0.53	0.21
	Awoken by disturbance	0.08	0.07	0.09
	Bed time [hour]	26.4	26.3	1.9
	Sleep latency [mins]	15.0	10	28.2
	Wake try time [hour]	9.7	9.5	1.8
	Sleep duration	7.0	7.2	2.3
	# of awakening per night	0.72	0	1.2
	Awakening duration [min]	6.0	0	35.1
	Nap duration	22.5	0	59.8
Stress	PSS Score_pre	14.77	14.00	7.35
	PSS Score_post	17.32	17.00	7.12
Physical Health	PCS_pre	57.62	58.26	4.27
	PCS_post	57.52	58.20	4.91
	delta PCS	-0.01	0.00	4.31
Mental Health	MCS_pre	44.33	46.47	8.20
	MCS_post	40.33	42.05	9.14
	delta MCS	-4.25	-2.79	7.90
Subjective daily measures	Alertness_Morning	51.8	52	26.5
	Happiness_Morning	62.3	63	22.2
	Energy_Morning	52.1	51	24.6
	Health_Morning	65.7	70	24.9
	Calmness_Morning	54.6	55	25.2
	Alertness_Evening	42.4	38	26.9
	Happiness_Evening	62.0	64	23.7
	Energy_Evening	48.0	48	24.2
	Health_Evening	65.7	70.5	24.9
	Calmness_Evening	54.5	55	26.1

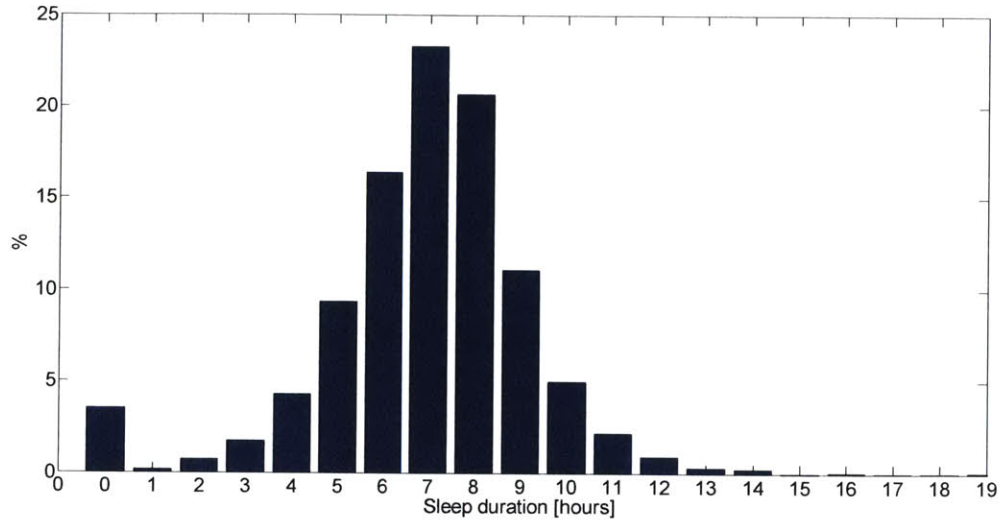


Figure 6.4 Distribution of sleep duration (4823 nights, N=165)

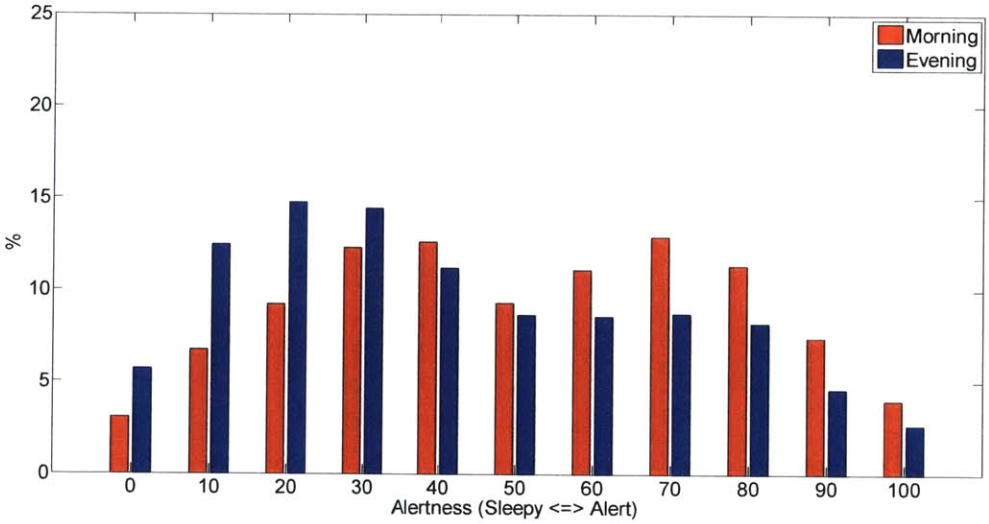


Figure 6.5 Subjective alertness distributions (4808 days, N=165)

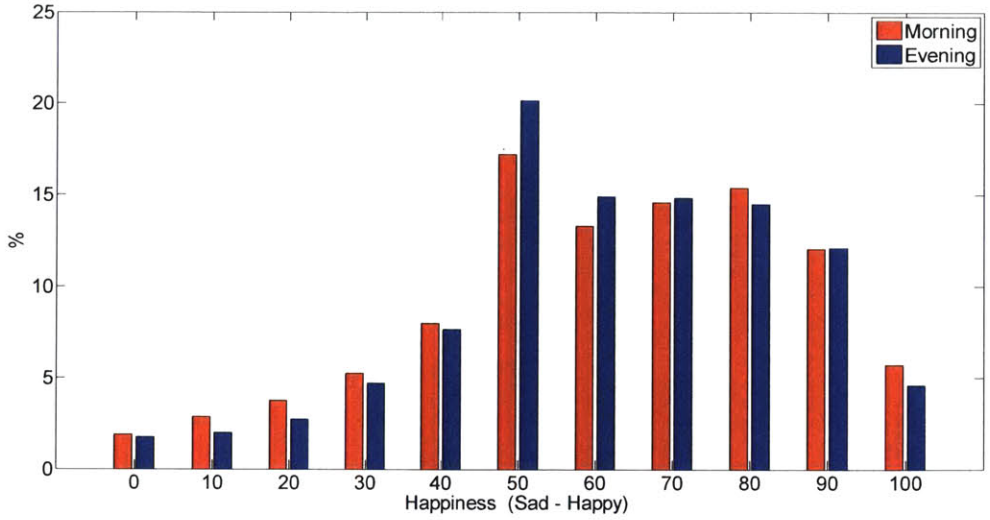


Figure 6.6 Subjective happiness distributions (4808 days, N=165)

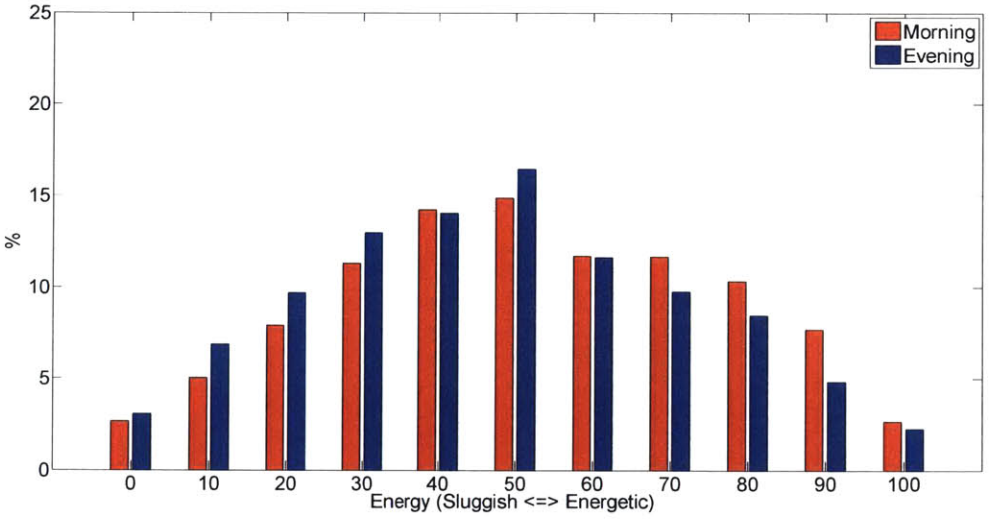


Figure 6.7 Subjective energy distributions (4799 days, N=165)

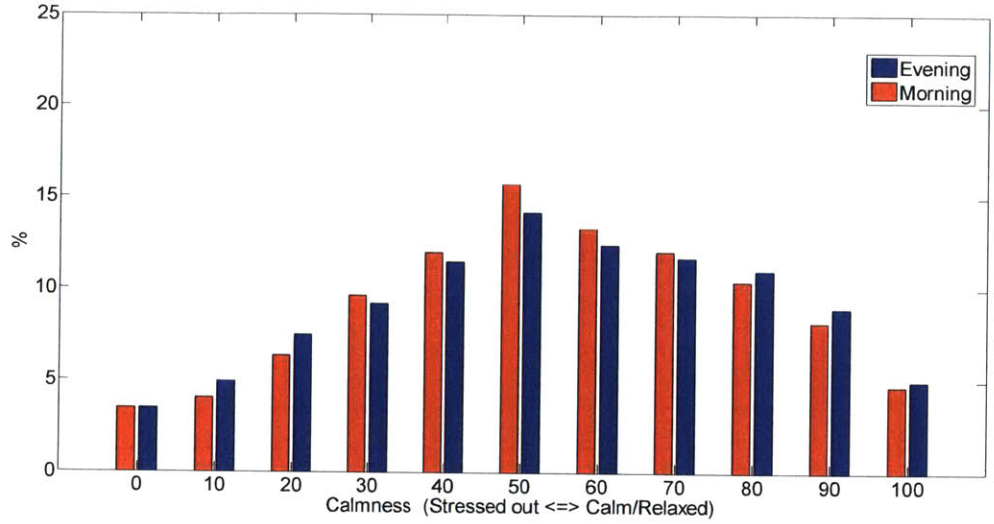


Figure 6.8 Subjective calmness distributions (4799 days, N=165)

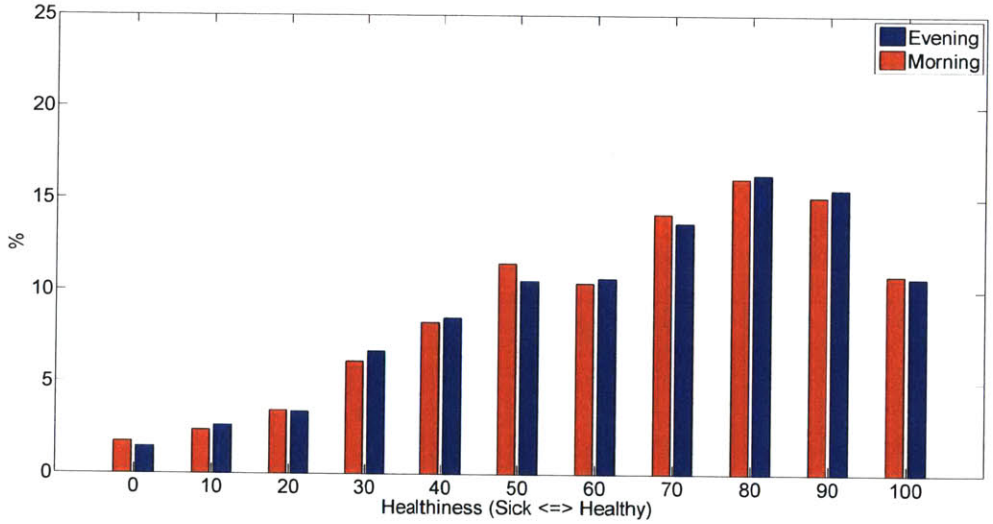


Figure 6.9 Subjective healthiness distributions (4801 days, N=165)

6.3.2 Characteristics of Social Interactions

Statistics about social interaction are shown in Table 6.4. The percentage of contacts reported as "social" in the frequent contacts was more than 50% and the % of interactions reported as "family" was the smallest. The majority of their interactions were positive (average 63%) and face-to-face interaction was the most common followed by email, SMS and phone. During the 30 day study period, they had positive interactions on 25% of the days and negative interactions on 11% of the days.

Table 6.4 Characteristics of social interactions (N=164)

Category	Parameters	Mean	Median	SD
Social interaction	Family contacts /# of total frequent contacts	0.06	0.06	0.05
	Social contacts /# of total frequent contacts	0.52	0.54	0.18
	Work contacts /# of total frequent contacts	0.14	0.10	0.12
	Positive contacts /# of total frequent contacts	0.63	0.64	0.19
	Neutral contacts /# of total frequent contacts	0.14	0.11	0.13
	Negative contacts /# of total frequent contacts	0.02	0.00	0.03
	Phone contacts /# of total frequent contacts	0.38	0.38	0.23
	Email contacts /# of total frequent contacts	0.56	0.57	0.21
	SMS contacts /# of total frequent contacts	0.47	0.52	0.21
	Face-to-face contacts /# of total frequent contacts	0.62	0.65	0.21
	Days with a memorable positive interaction / total # of days in study	0.25	0.19	0.24
	Days with a somewhat negative interaction/ total # of days in study	0.08	0.03	0.12
	Days with very negative interactions/ total # of days in study	0.03	0.00	0.05
	Days with neither positive nor negative interaction / total # of days in study	0.63	0.70	0.25

6.3.3 Characteristics of Mobile Phone Usage

Table 6.5 shows the statistics about phone calls, SMS and screen-on time on their mobile phones. Participants turned on their phone screen 113 times on average for 126 minutes per day on average. Figure 6.10 shows the percentages of call, SMS and screen on every 3 hours. Screen-on, call and SMS percentages for night time (9pm-3am) were 25.2 %, 26.7% and 21.3%.

Table 6.5 Characteristics of phone usage (N=152, 3065 days)

Category	Parameters	Mean	Median	SD
Phone call	Total number of phone calls per day	4.6	3	6.1
	Total number of people each participant interacted with per day by calls	2.4	1	3.6
	Total duration of phone calls per day [mins]	14.1	3.2	42.0
SMS	Total number of SMS per day	40.7	26	45.5
	Total number of people each participant interacted with per day by SMS	4.9	4	3.8
Screen on	Total number of times screen turned on	113.3	93	87.8
	Total duration of screen on [mins]	125.8	110.2	87.8

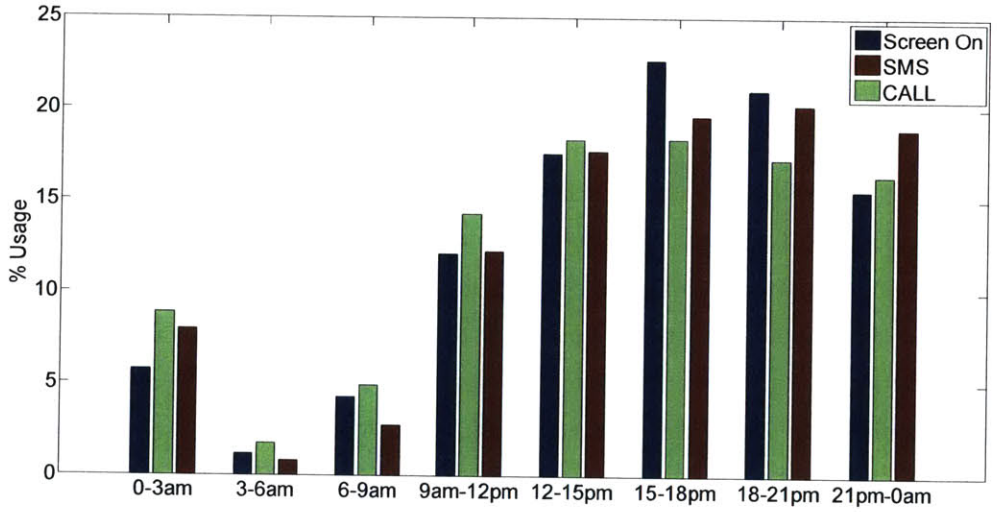


Figure 6.10 Phone usage distributions

6.3.4 Characteristics of Daily Activities

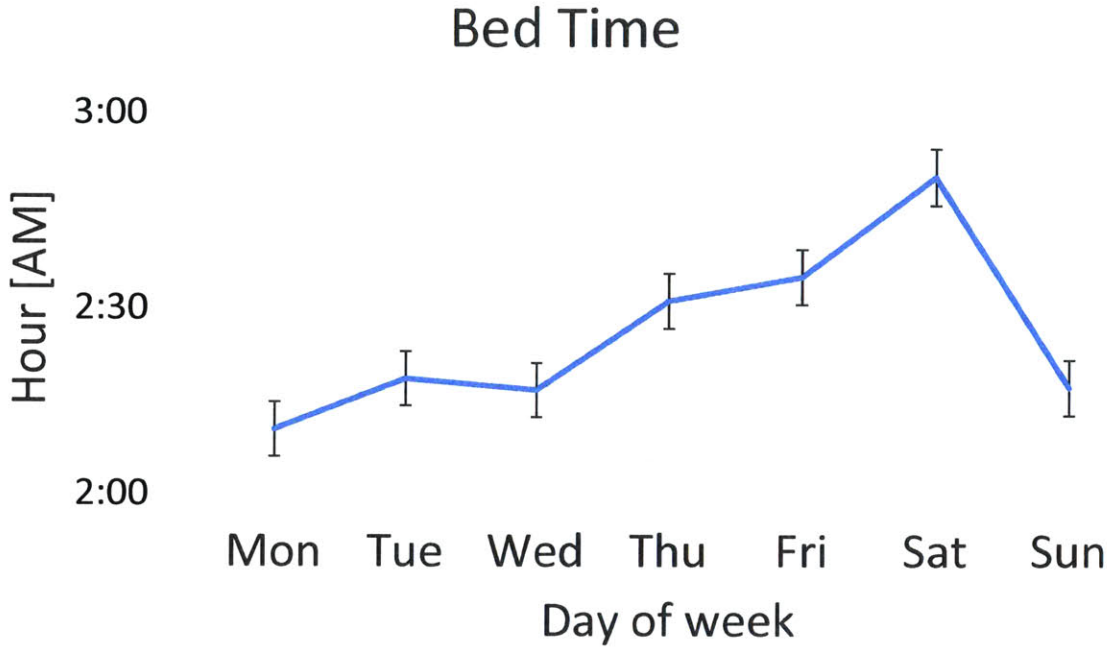
Table 6.6 shows the characteristics of daily activities. Participants reported spending 5 hours on average per day for study including classes and studying by themselves. We also measured total amount of caffeinated and alcoholic drinks and timing of drugs, but the analysis of these is saved for future work.

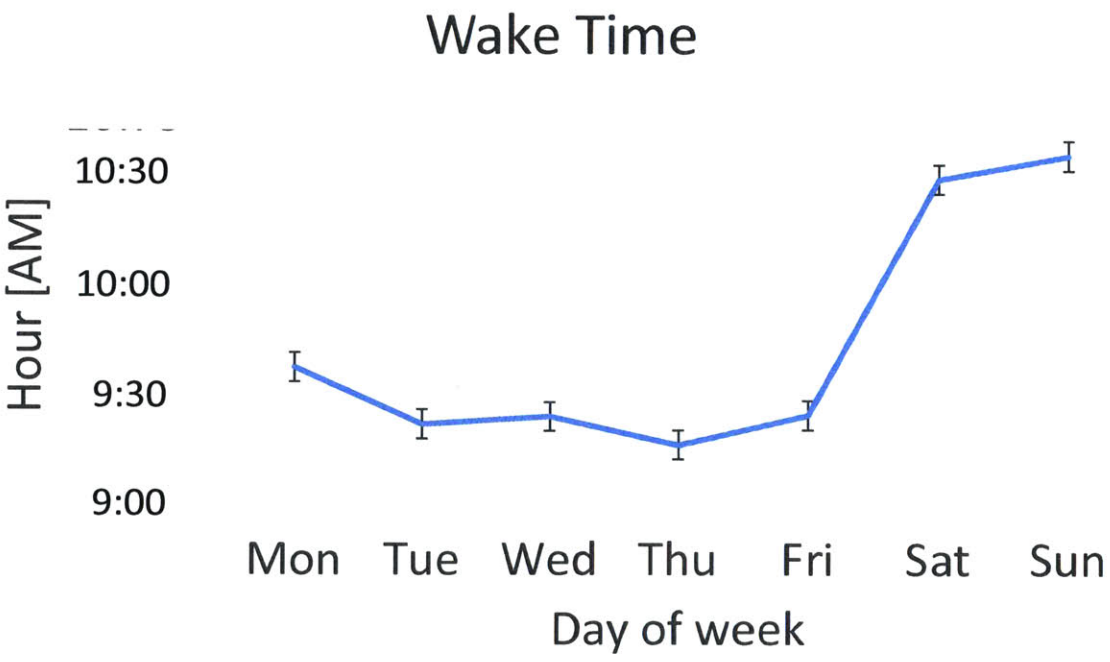
Table 6.6 Characteristics of activities

Category	Parameters	Mean	Median	SD
Activities	Total minutes of academic activities (including classes, e-classes, sections, seminars, labs, study groups)	132.9	120	136.8
	Total minutes of study activities (studying alone)	178.7	180	166.9
	Total minutes of exercise (including sports, gym, cycling, etc)	26.4	0	70.9
	Total minutes of extracurricular activities	59.7	0	607.9
	Time of day of last caffeinated drink	14.3 (2:18pm)	16 (4pm)	6.9

6.3.5 Participant Behaviors on Days of Week

Next, we look at behavioral characteristics on days of the week (Figure 6.11-13). Bed time became later from Monday to Saturday. Saturday bedtime (Saturday night) was significantly later than the ones from Sunday to Thursday ($p=6.79e-11$). Saturday and Sunday wake time were significantly later than the rest of the days ($p=5.57e-83$). Like wake time, sleep duration from Friday to Saturday and from Saturday to Sunday was significantly longer than the rest of the days ($p=1.34e-12$). Sleep duration from Thursday to Friday was shortest of all and significantly shorter than duration on Sunday, Monday, Friday and Saturday nights. On weekends, they did have much shorter academic activities such as classes. On Friday and Saturday, their study time was significantly shorter than on other days ($p=1.34e-12$).


Bed Time
3:00
2:30
2:00
Hour [AM]
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Day of week


Wake Time
10:30
10:00
9:30
9:00
Hour [AM]
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Day of week

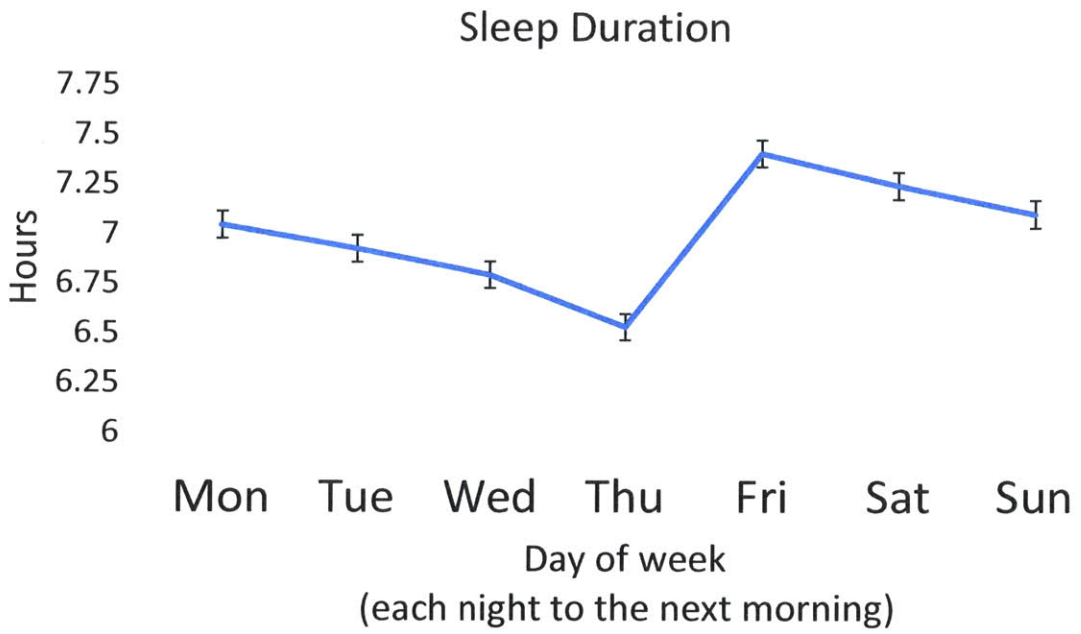


Figure 6.11 Bed time, wake time and sleep duration on days of week (N=165, 4736 days)

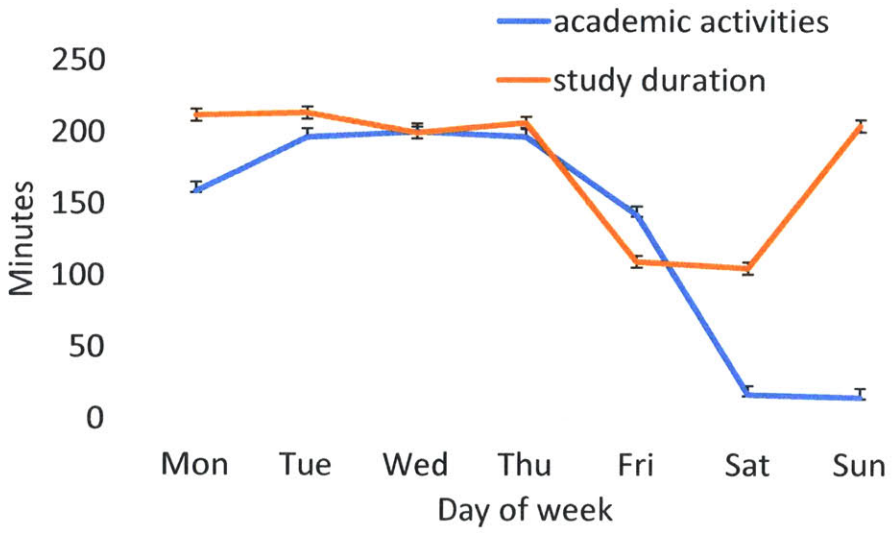


Figure 6.12 Academic activities and study duration on days of week (N=165, 5025 days)

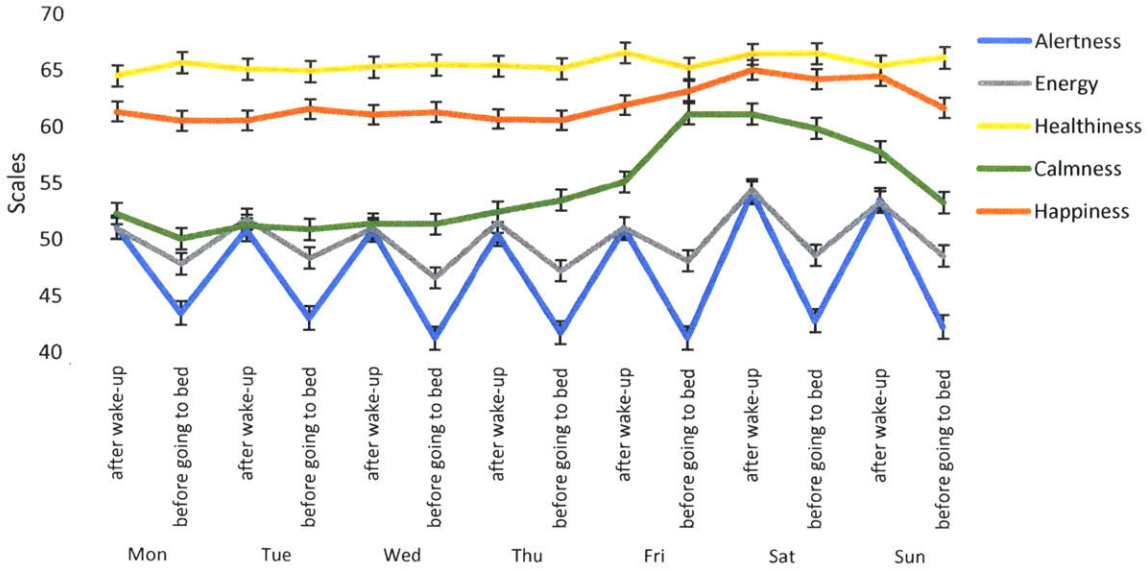


Figure 6.13 Subjective wellbeing-related measures on days of week (N=165)

Figure 6.13 shows subjective measures on each day of week. Here, we show subjective measures right after they woke up (morning) and right before they went to bed (evening). Alertness and energy in the mornings were higher than in the evenings. Calmness was the scale which showed the largest difference between the maximum and minimum, followed by happiness. Happiness on Saturday morning was the highest and significantly higher than on Monday evening, Tuesday morning and Thursday morning and evening. Calmness on Friday night, Saturday morning and evening was statistically significantly higher than on the rest of the days (ANOVA, post-hoc test, $p < 0.05$).

6.3.6 Sensor Data

We have collected 4318 days of Q-sensor data from N=166 (total 103632 hours). We computed how much data were within a normal range (Table 6.7). Also, 80% (± 23%) of the collected EDA data were classified as clean data using the algorithm we described in the last chapter. Thus, among our collected EDA data, on average, 64% (80% of the collected data within a normal range x 80% of the data was clean) can be used for further analysis.

Table 6.7 Statistics about our sensor data

Category	Normal Range	Percentage of the data within normal range	
		Mean	SD
EDA	0.01-30 microS	80	30
Skin temperature	20-40 Celsius degrees	99.7	2.6

Next, we computed mean and SD of EDA amplitude and number of peaks per 30-s epoch over sitting, walking and running during daytime and 1-4 quarters of sleep (Figure 6.14). The EDA amplitude was higher in day time than during sleep on average. EDA amplitude increased as activity magnitude became higher. During sleep, the 2nd quarter of sleep showed the highest amplitude. The number of EDA peaks was highest during running, followed by walking, the 1st and 2nd quarters of sleep and when sitting.

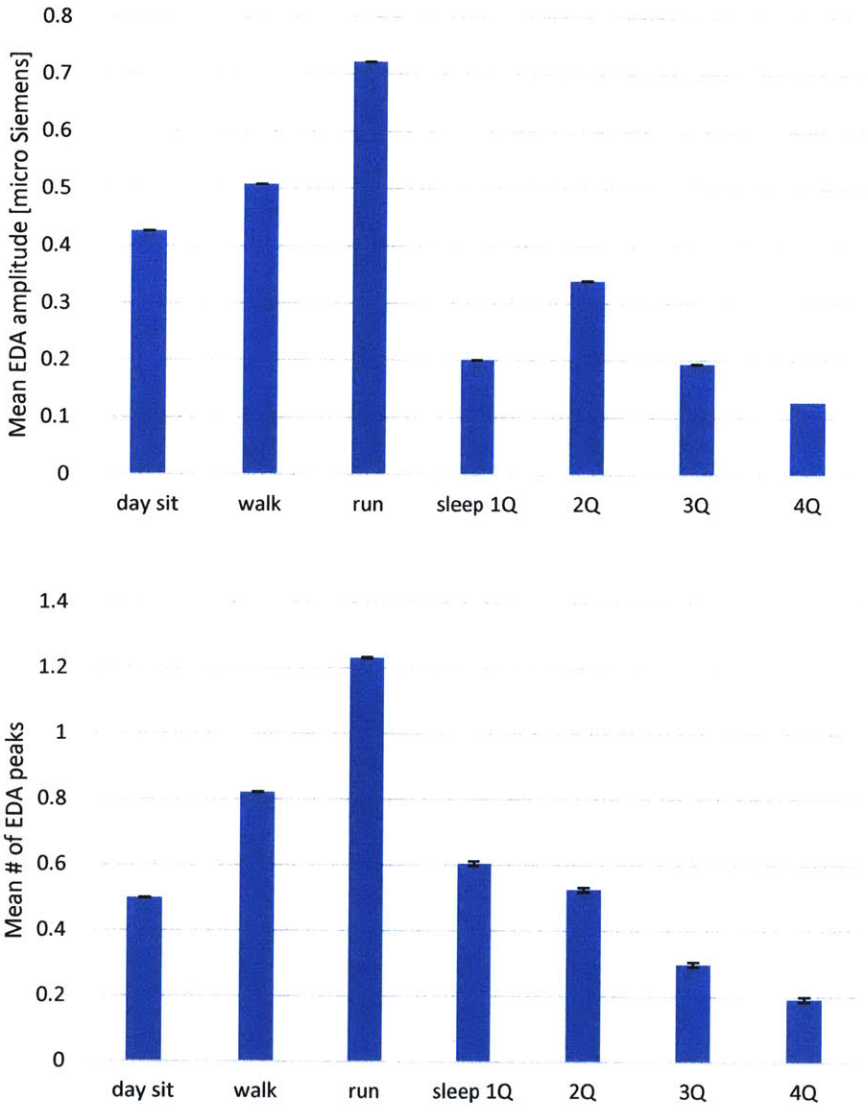


Figure 6.14 Comparison of EDA amplitude and number of peaks over different conditions

Chapter 7
SNAPSHOT Study: Data Validation, Analysis and Results

In this chapter, we give examples of how the SNAPSHOT data is valid and useful for answering important questions related to sleep, stress, mood and health. While the data and methodology enable hundreds of questions to be addressed, we focus on only a few examples here.

As I described in the introduction chapter, the SNAPSHOT study is designed to collect rich daily ambulatory data to understand interactions between sleep, and stress, mental health and other wellbeing-related measures. In this validation, we focus on the following two things (1) the influence of sleep duration and regularity, two different measures on sleep behaviors on stress, mental health and subjective wellbeing-related measures (2) stress recognition using multi-modal data.

7.1 Influence of Sleep Regularity and Sleep Duration on Academic Performance and Health

7.1.1 Methods

We analyzed the effect of sleep regularity and sleep duration on academic performance, physical/mental health score (PCS, MCS), Perceived Stress Scale (PSS) and subjective wellbeing-related scores (alertness,

happiness, sluggishness, healthiness and calmness). For the analysis, we used coarsened exact matching [Iacus, 2009] to control several covariates in two groups and then, we applied t-test or Man-Whitney U test (for parameters with non-Gaussian distributions) to compare how these two groups are different.

In the following analysis, we tested several different coarsening methods to evaluate the consistency of results.

(1) Sleep regularity

First, we analyzed the influence of sleep regularity on GPA for the semester, PCS, MCS, and PSS at the end of the study, and the daily average subjective wellbeing-related measures. Sleep regularity was computed from 30-day sleep diary (sleep and nap logs). We matched our samples with gender and with monthly averaged sleep duration. For sleep regularity, we compared the following two criteria ((a) top and bottom 40%, (b) >= 0.85 (mean), < 0.85). For coarsened matching of sleep duration, we used following criteria, making sure these sets had the same representation in the two sleep regularity groups: Sleep duration of (a) >= 6 hours, < 6 hours excluding naps; (b) >= 7hours, < 7 hours excluding naps; and (c) >= 7 hours, < 7 hours including naps. We also controlled for gender, and stress level (PSS) at the beginning of the study. Therefore, we compared the following:

(a) Sleep regularity (top 40% vs bottom 40 %) matched with gender and sleep duration (sans naps, >=7 hours vs < 7 hours)

(b) Sleep regularity (top 40% vs bottom 40 %) matched with gender and sleep duration (sans naps, >=6 hours vs < 6 hours)

(c) Sleep regularity (>=0.85 vs < 0.85) matched with gender, sleep duration (including naps, >=7 hours vs < 7 hours)

(d) Sleep regularity (top 40% vs bottom 40 %) matched with gender, sleep duration (sans naps, >=7 hours vs < 7 hours) and Perceived Stress Scale (post study, >=14 vs < 14)

(2) Sleep duration

Second, we analyzed the effect of sleep duration on GPA for the semester, PCS, MCS, and PSS at the end of the study, and the daily average subjective wellbeing-related measures. We matched our samples with gender and monthly sleep regularity. For sleep regularity, we compared the following two criteria ((a) top, bottom 40% and mid 20%, (b) >= 0.85 (mean), < 0.85). For sleep duration, we compared the following criteria (a) >= 7hours, < 7 hours for sleep duration excluding naps (b) top vs bottom 40% and (c) >= 7hours, < 7 hours including naps.

7.1.2 Results

First, we show the comparison between regular and irregular sleepers with different coarsening in Tables 7.1, 7.2, and 7.3.

Table 7.1 Comparison between regular and irregular sleepers (matched with gender (67%: male) and 4 week average sleep duration sans naps (>=6 hours or < 6 hours))

	Regular (top 40% N=19)	Irregular (bottom 40% N=21)	p-value
GPA for the semester	4.38	4.35	0.87
PSS Score_post	16.37	20.05	0.10
PCS_post	55.16	57.35	0.12
MCS_post	44.40	36.73	**0.01
Alertness_Morning	60.16	48.09	**0.01
Happiness_Morning	67.02	55.96	*0.03
Energy_Morning	62.15	47.45	***0.00
Health_Morning	70.90	61.54	0.12
Calmness_Morning	57.96	49.90	0.14
Alertness_Evening	47.10	39.43	0.16
Happiness_Evening	65.27	55.88	0.08
Energy_Evening	52.13	43.50	0.08
Health_Evening	69.13	61.84	0.24
Calmness_Evening	57.07	48.98	0.17

Table 7.2 Comparison between regular and irregular sleepers (matched with gender and 4 week average sleep duration sans naps (>=7 hours or < 7 hours))

	Regular (top 40% N=16)	Irregular (bottom 40% N=16)	p-value
GPA for the semester	4.50	4.30	0.37
PSS Score_post	16.19	18.50	0.28
PCS_post	55.09	58.19	0.06
MCS_post	45.10	37.93	*0.02
Alertness_Morning	61.08	50.53	*0.05
Happiness_Morning	68.01	59.75	0.09
Energy_Morning	64.14	50.48	**0.01
Health_Morning	73.31	65.92	0.24
Calmness_Morning	58.72	53.30	0.31
Alertness_Evening	51.25	42.13	0.11
Happiness_Evening	66.83	59.56	0.15
Energy_Evening	56.61	46.08	*0.04
Health_Evening	71.98	65.63	0.32
Calmness_Evening	58.44	52.13	0.28

Table 7.3 Comparison between regular and irregular sleepers (matched with gender and 4 week average sleep duration including nap (>=7 hours or < 7 hours))

	Regular (regularity>=0.85 N=20)	Irregular (regularity < 0.85 N=34)	p-value
GPA for the semester	4.35	4.35	0.95
PSS Score_post	16.75	19.26	0.19
PCS_post	55.56	57.89	0.08
MCS_post	43.87	37.31	**0.01
Alertness_Morning	59.58	48.05	**0.01
Happiness_Morning	66.63	58.09	*0.05
Energy_Morning	61.57	48.83	**0.01
Health_Morning	69.81	61.17	0.09
Calmness_Morning	57.97	52.06	0.23
Alertness_Evening	45.99	41.88	0.42
Happiness_Evening	64.87	58.71	0.17
Energy_Evening	51.33	45.90	0.25
Health_Evening	68.07	62.26	0.28
Calmness_Evening	56.96	51.51	0.31

With the three different coarsening methods, we found that regular sleepers have statistically significant higher mental health score, higher alertness and energy level in the morning than irregular sleepers.

Morning happiness was also either higher ($p < 0.05$) or trended toward being higher in the morning ($p < 0.10$) for regular sleepers.

In order to further examine the influence of sleep regularity on mental health score, we also controlled for stress level using the perceived stress scale at the end of the study.

Table 7.4 Comparison between regular and irregular sleepers (matched with gender, average sleep duration (>=7 hours or < 7 hours) and PSS (post) (>=14, < 14))

	Regular (top 40% N=18)	Irregular (bottom 40% N=21)	p-value
GPA for the semester	4.38	4.32	0.83
PCS_post	55.21	56.80	0.32
MCS_post	44.06	36.31	*0.02
Alertness_Morning	60.74	49.53	*0.05
Happiness_Morning	67.13	54.71	*0.04
Energy_Morning	62.17	46.44	**0.01
Health_Morning	70.12	62.08	0.25
Calmness_Morning	58.56	49.01	0.11
Alertness_Evening	47.06	38.35	0.14
Happiness_Evening	65.31	54.76	0.09
Energy_Evening	51.64	41.52	0.06
Health_Evening	68.01	62.44	0.43
Calmness_Evening	57.51	48.04	0.15

With stress level matching, gender matching and sleep duration matching, mental health was still statistically higher in regular sleepers than irregular sleepers. Morning alertness, happiness, and energy were also still higher in the regular than in the irregular sleepers ($p <= 0.05$).

Next, we analyzed the influence of sleep duration on mental health, stress and other health measures.

We also asked "how long do you need to sleep regularly for it to make a difference?" Thus, we analyzed the influence of the most recent 4, 3, 2, and 1 weeks of sleep regularity on the mental health score at the end of the study while matching for gender and sleep duration. We found that with 4 weeks and 3 weeks of sleep regularity, regular sleepers showed higher mental health than irregular sleepers; however the difference was not statistically significant when narrowed to only 2 and 1 weeks of sleep regularity.

To make sure that the test of the most recent "1 week of regularity" was not insignificant simply because 7 nights is not enough to provide statistical significance, we tested one other "single" week that we thought

might be significant. We analyzed how the first 1 week of the 30 day period's regularity influenced mental health at the end of the study. The regular sleepers based on the first 1 week of sleep data showed higher mental health than the irregular sleepers based on the first 1 week of sleep data. Thus, we see one week of data is adequate to have statistical significance. In addition, we expected that sleep regularity got lower as the semester went on; however, we did not find any statistically significant difference among the 1st, 2nd, 3rd and 4th weeks of regularity. We also analyzed if those who were regular sleepers at the first week were more likely to be regular sleepers in the later weeks. Sixty-five % of the first week's regular sleepers were regular sleepers at the following 2nd-4th week and 56 % of the first week's irregular sleepers were irregular sleepers for the following 2nd-4th weeks; therefore, for the majority of our participants, sleep regularity/irregularity was consistent over the semester.

Next, we considered the impact of sleep duration, while controlling for other factors. Tables 7.5-7.8 show the comparison between short and long sleepers with different coarsening. Our results showed higher PCS (physical health) scores at the end of the study (Table 7.6) and higher morning healthiness in long sleepers (Table 7.7); however these results were not consistent over the different ways of coarsening the control variables.

Table 7.5 Comparison between short and long sleepers (matched with gender and sleep regularity (top 40% or bottom 40%))

	short (bottom 40% N=15)	long (top 40% N=17)	p value
GPA for the semester	4.28	4.50	0.35
PSS Score_post	17.40	17.29	0.96
PCS_post	55.49	57.65	0.21
MCS_post	40.70	42.23	0.64
Alertness_Morning	53.29	58.02	0.39
Happiness_Morning	61.69	65.82	0.40
Energy_Morning	55.76	58.68	0.61
Health_Morning	65.89	72.91	0.27
Calmness_Morning	55.23	56.70	0.78
Alertness_Evening	47.70	45.80	0.75
Happiness_Evening	61.14	65.01	0.45
Energy_Evening	51.51	51.20	0.95
Health_Evening	65.55	71.69	0.33
Calmness_Evening	54.24	56.21	0.74

Table 7.6 Comparison between short and long sleepers (matched with gender and sleep regularity (top 40%, bottom 40% or mid 20%))

	Short (< 7 hours N=25)	Long (>= 7 hours N=29)	p value
GPA for the semester	4.29	4.42	0.40
PSS Score_post	18.90	17.68	0.54
PCS_post	55.86	58.37	*0.03
MCS_post	39.29	40.26	0.71
Alertness_Morning	51.05	53.79	0.54
Happiness_Morning	58.49	64.46	0.16
Energy_Morning	52.20	55.12	0.52
Health_Morning	60.14	69.27	0.06
Calmness_Morning	51.22	57.76	0.15
Alertness_Evening	45.50	40.97	0.34
Happiness_Evening	57.84	64.65	0.12
Energy_Evening	49.51	46.06	0.43
Health_Evening	60.76	68.65	0.11
Calmness_Evening	50.26	57.32	0.16

Table 7.7 Comparison between short and long sleepers (matched with gender and sleep regularity (>=0.85 or < 0.85))

	Short (< 7 hours N=26)	Long (>= 7 hours N=22)	p value
GPA for the semester	4.35	4.46	0.31
PSS Score_post	16.38	15.88	0.43
PCS_post	56.53	58.14	0.10
MCS_post	43.10	41.26	0.59
Alertness_Morning	56.38	54.25	0.69
Happiness_Morning	64.08	65.22	0.23
Energy_Morning	58.12	54.93	0.58
Health_Morning	64.97	73.19	*0.04
Calmness_Morning	56.30	60.55	0.23
Alertness_Evening	51.52	41.45	0.33
Happiness_Evening	62.27	65.37	0.19
Energy_Evening	55.65	45.00	0.53
Health_Evening	63.79	72.78	*0.05
Calmness_Evening	54.81	60.46	0.19

Table 7.8 Comparison between short and long sleepers (matched with gender and sleep regularity (>=0.85 or < 0.85))

	Short (< 7 hours including nap N=22)	Long (>= 7 hours N=32)	p value
GPA for the semester	4.31	4.38	0.62
PSS Score_post	16.27	19.75	0.07
PCS_post	56.71	57.24	0.69
MCS_post	42.10	38.12	0.11
Alertness_Morning	56.09	49.72	0.17
Happiness_Morning	65.23	58.52	0.12
Energy_Morning	58.86	49.90	0.05
Health_Morning	68.79	61.34	0.15
Calmness_Morning	58.75	51.15	0.11
Alertness_Evening	47.81	40.37	0.13
Happiness_Evening	63.99	58.92	0.25
Energy_Evening	52.92	44.47	0.07
Health_Evening	68.39	61.68	0.19
Calmness_Evening	58.06	50.41	0.14

Therefore, our results showed that sleep regularity in the 3-4 weeks preceding evaluation has a significant influence on mental health scores and on morning alertness, happiness, and energy scores. Moreover, sleep duration does not have a statistically significant influence on these factors after controlling for sleep regularity. Sleep duration, after controlling for regularity and gender, does show a slight trend toward possibly influencing some physical health measures, but this is not a strong result although it may become significant as we collect more cohorts.

It is important to note that because of the matching on regularity/irregularity and gender, not all the participants are in these comparisons. The average sleep duration in the short sleep group (after matching on gender and regularity) was 6.2 hours (median: 6.3, SD: 0.34), while the average sleep duration in the long sleep group was 7.9 hours (median: 7.7, SD: 0.55). Thus, we did not have hugely sleep deprived people in this comparison.

Overall, in these college student data, sleep regularity appears to be more important for mental health than sleep duration.

7.2 Stress Recognition using Multi-modal Data

Next, we investigate physiological, behavioral and personality markers for stress and how accurately we can classify stress level using surveys, personality types and objective measurement.

In order to understand which features from which modality work best in classifying high and low stress groups (top and bottom 20%, each N=55) based on their PSS score at the end of the study and find the classification accuracies, we applied sequential forward feature selection to find the best combinations of 1-3 features and support vector machines (SVM) (linear) and SVM (radial basis function kernel) classifiers. We compared the classification accuracies using the following modalities of features (Table 7.9). For each classification, we examined the accuracy using a leave-one-participant-out approach. We selected features and trained models from all except one participant's data and tested the model against the left-out participant's data.

Table 7.9 Features examined for influence

Personality type (5 features)	Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism
Sleep parameters (10 features)	Regularity, PSQI Score, bed time, sleep latency, wake time, sleep duration, nap duration, last caffeine time, sleep quality, no sleep%
Activity (6 features)	Total academic duration, total study duration, total exercise duration, total extra curricular activity duration, %sit, %walk
Social (5 features)	% memorable positive interaction % somewhat negative interaction % very negative interaction
EDA (49 features)	Mean, median, SD of EDA amplitude and # of EDA peaks (sitting, walking, running and 1-4Q sleep), percentages of EDA epochs for sitting, walking, running, and 1-4Q sleep
ACC (24 features)	Mean, median, SD of ACC (day, sitting, walking, running and 1-4Q sleep
ST (6 features)	Mean, median, SD of day and sleep skin temperature
Call (7 features)	# of calls, total duration of calls, mean, median and SD of timestamp, entropy, incoming %
SMS (7 features)	# of SMS, mean, median and SD of timestamp, entropy, % of # of outgoing SMS (0-3am), % of # of outgoing SMS (3-6am), incoming %
Screen (11 features)	mean, median and SD of timestamp, total duration of screen on, % of screen on (0-3am, 3-6am, 6-9am, 9-12am, 12-15pm, 15-18pm and 18-21pm)

Table 7.9 Features examined for influence (cont)

Location (4 features)	Total travel, daily mean, median, SD
Email (16 features)	median_daily_received, std_daily_received, median_daily_sent, std_daily_sent, median_time, stamp_sent, median_timestamp_received, std_timestamp_sent (in seconds), std_timestamp_received (in seconds), sent_0a_3a (%), sent_3a_6a (%), sent_6a_9a (%), sent_9a_12p (%), sent_12p_3p (%), sent_3p_6p (%), sent_6p_9p (%)
Light (8 features)	Lux mean and SD for 0-6am, 6am-12pm, 12pm-18pm, 18-24pm
Survey (10 features)	GPA for the semester, BMI, PSQI Score, State Score, Trait Score, PSS Score_post (not for stress recognition), PCS_post, MCS_post, delta PCS, delta MCS
Wrist	EDA, ACC, ST, light
Mobile phone	Call, SMS, screen, location

Table 7.10 shows selected features and their accuracy discriminating the two groups. Since we used a leave-one-participant-out approach, we sometimes observed a subset of features which worked for a certain percentage of participants and another subset of features which worked for the rest of the participants. Percentages in brackets show the percentages of people that the feature was selected for. For example, in Table 7.10, we obtained 80.0 % accuracy with Extraversion, PCS (post) and percentage of sitting for 57% of the participants and Extraversion, PCS (post) and MCS (post) for 43% of the participants.

The classification accuracy was highest (82.4%) with surveys (PSQI, PCS and MCS) and skin temperature (but not its rhythmicity) showed the lowest accuracy of all modalities of features. Sleep features (low sleep regularity, high PSQI (poor sleep) and long sleep duration (based from diary, but actiwatch based sleep duration was shorter in the high stress group) contributed to the high stress group. Among objective measures, features from an accelerometer worked best to classify high vs low stress group: the participants who spent more time on walking-level activities or who had a higher SD of running-level activity percentages over 30 days were more stressed. We also found a higher standard deviation of EDA amplitude while sitting was related to higher stress level. Some of the objective features such as higher percentages

of 3-6am outgoing SMS over the total number of outgoing SMS and 3-6am screen on were indicators for the high stress group staying up late. Social interaction features worked more accurately than activity features (total academic activity duration, not total exercise duration).

Table 7.10 Stress recognition: arrows (↑ ↓) show how each feature contributes to the high stress group

Modality	Accuracy	Features
Survey	82.4	PSQI score ↑ + PCS_post ↓ + MCS_post ↓
ALL	80.0	Extraversion ↓ + PCS_post ↓+ {median sit % ↓ (57%) or MCS_post ↓}
Sleep	77.2	Regularity ↓ + PSQI ↑ + sleep duration ↑
Personality	75.2	{Conscientiousness ↓ or agreeableness ↓ (52%)} + Extraversion ↓
ACC	72.8	Mean % walk ↑ or SD % run ↑ (51%)
Social	70.9	Very negative interaction % ↑ + {family contact % ↓ or positive interaction ↓ (54%)}
Phone + Wrist	68.3	3-6 am SMS outgoing % ↑ + lux SD (12-18pm) ↓ + mobility SD ↑
Email	67.3	Sent timestamp SD ↓ + {sent_6p_9p (%)↓ or sent_9p_0a (%)↑ (49%)}
Wrist	67.0	SD EDA sit ↑ + median # of peaks walk ↓ + {SD exposure (0-6am) ↑ or day ST SD [51%]↓}
Screen	67.0	3-6am % ↑ + 15-18pm % ↑
EDA	64.7	SD EDA sit ↑ + median # of peaks walk ↓
Light	57.3	SD exposure (0-6am) ↑
SMS	56.8	Timestamp SD ↓
Call	56.4	# of calls ↑
Location	55.6	Travel sum ↑+ median ↑ + SD ↑
ST	55.3	Day SD ↓ + sleep mean ↑+ sleep SD ↑
Activity	54.8	Academic activity total duration ↑

Chapter 8
Discussions, Limitations and Contributions

This chapter provides discussion based on our results in the previous chapters, along with limitations and contributions of this thesis.

8.1 Discussion

8.1.1 Characteristics of our population

How similar is our population to other college student populations? Here we discuss similarities based on measures (mostly surveys) that have been conducted both inside and outside of MIT.

Our participants' averaged sleep duration, based on a combination of daily self-report online diary entries and daily actigraphy from their wrists, was 7.0 hours excluding naps (6.9 hours on weekdays and 7.3 hours on weekends). A total of 78% of the participants slept less than 8 hours on average over all nights. Fifty-six % of the participants slept less than 7 hours on average on weeknights and 40% did on weekends. The Healthy Minds Study [Massachusetts Institute of Technology, 2015], hosted out of the University of Michigan and given to over 100 colleges and universities, was also run at MIT during the timeframe of our study with nearly 3000 students responding, including asking what time they typically go to sleep on weeknights and what time they typically wake up on weekdays. Based on these two self-reported typical sleep/wake times, the study concluded that more than half (54%) of MIT respondents get 8 or more hours of sleep, while 17% allow themselves to get less than 7 hours of sleep of weeknights. The same study,

conducted in 2013–14, concluded that national respondents were able to sleep an average 7.4 hours on weeknights. Another sleep survey done using self-reported bedtime and sleep duration on the MIT campus in 2012 showed that MIT undergraduate students' average typical weeknight sleep was 6.5 hours N=1441) [Lin, 2012]. Thus, our use of daily sensor data and daily online diary data (with a smaller population) shows lower sleep duration per weeknight than did Lin's survey method that infers sleep duration by asking for a single typical go-to-sleep time and single typical wake-up time in the Healthy Minds Study and also showed longer sleep duration than did Lin's survey done with bedtime and sleep duration (more than 10x our population). While it is possible that our participants were an unusual subset compared to the rest of the MIT population, it is also possible that the more precise methods we used, and the fact that people tend to think that the time they looked at the clock to go to bed is the same time they typically "go to sleep", could also explain why we found lower durations of sleep.

About sleep quality, prior studies from an urban Midwestern university (N=1000) reported that over 60% were categorized as poor quality sleepers by PSQI [Lund, 2010]. Our population showed a smaller population of poor sleepers (49%) than this data (more than 7x our population).

About stress, the majority of college students feel stressed on a daily basis: 85% of students are stressed on a daily basis from a single item in a survey of 2200 students at 40 randomly chose colleges in the United States [The Associated Press and MTV, 2009]. In our daily subjective data, 42% of the daily survey answers showed a bias toward being more stressed on the daily stress-calm scale. A previous study done at Southern Illinois University at Carbondale showed that the average college student PSS score was 18.95 (N=559) [Olpin, 1996]. Another study at a Turkish university (N=508) showed the average freshmen PSS score was 18.89 [Örücü, 2009]. Our average PSS score at the end of the study was 17.3, so our population was not extremely stressed compared to the other college student population.

Daily happiness was higher over weekends than during weekdays (Figure 6.13). While we don't know of studies of this self-reported data daily in college students, this pattern was consistent with previous studies measuring happiness by monitoring smiles with cameras at the university campus [Hernandez, 2012] and by using a subjective happy scale (extremely-not at all) on the iphone app in the U.K. [MacKerron 2012] (sample size each day > 100000). Alertness was higher upon waketime than upon bedtime, consistent with previous findings from laboratory alertness measurement using a visual analog scale [Van Dongen, 2010], although subjective alertness is influenced by various factors (circadian rhythm, naps, morningness and eveningness).

As I already described in the previous chapter, the mean physical health scores (pre and post study) in our population were higher than the average (53.0) for 18-24 year old populations in the U.S. [Utah Department of Health, 2001]. On the other hand, our participants' self-reported mental health scores (44.3 and 40.3 in the pre and post surveys) were slightly worse than the national average (46.0, 18-24 years old).

Some previous studies about mobile phone usage among college students reported that female college students spent about 10 hours/day on their phone [Roberts, 2014]; however that study was based on a questionnaire. Our objectively-gathered data showed that our population used mobile phones for a much shorter period of time (126 mins/ day on average) and we did not find any statistically significant difference in daily phone usage duration between male and female students. This proved the importance of collecting data using **O**bjective **T**echniques.

8.1.2 The influence of sleep duration and sleep regularity on academic performance, stress and wellbeing-related measures

Our results (section 7.1,2) showed that irregular sleepers had more negative outcomes (lower reports of alertness, and energy in the morning and lower mental health) compared to regular sleepers, even after controlling for sleep duration, gender and stress (PSS score) self-reported for the month. Moreover, this difference between regular and irregular sleepers is associated with a much larger effect on mental health than is a difference in sleep duration. Previous studies that monitored sleep patterns and stressful events over 9 consecutive nights (N=184) showed a significant relationship between the variability of sleep duration measured with actigraphy and stressful events (measured with a modified version of Psychiatric Epidemiology Research Inventory Life Events Scale)) [Mezick, 2009]. Another previous study showed that regularizing sleep-wake patterns reduced negative mood (e.g. tension-anxiety, anger-hostility, and fatigue) and the reduction was gone after regulation was taken away [Takasu, 2012]. One study compared healthy individuals and individuals with bipolar disorders and indicated the bipolar groups were more likely to have irregular bed-rise time [Baek, 2014]. The BWH team has conducted a 30-day sleep study for Harvard college students and reported the relationship between sleep irregularity, and poor sleep quality [Clerx, 2014, Clerx, 2015]. These previous studies implied negative health outcomes of irregular sleep similar to our results.

Previous studies showed the negative relationship between stress level and sleep duration (negative correlation in the StudentLife study [Wang, 2014] and daily stress recognition using sleep duration and other measurement and personalized models, Muaremi, 2013); however, our comparison between short sleepers and long sleepers did not show statistically significant difference in their stress levels. Wang's and Muaremi's papers did not show the statistics about sleep durations on their dataset; therefore, it is hard to compare their results to our dataset (short sleeper: 6.2 hours vs long sleepers: 7.9 hour sleep durations (one month average)). However, once again, our results are based on monthly averaged durations because we compare sleep duration to sleep regularity and we will analyze how daily sleep duration affects stress level on the following day as the next step.

Further causality effect analysis (e,g, [Bouwmans, 2015]) is required to understand for instance, whether low mental health causes irregular sleep or irregular sleep causes low mental health; however, short sleepers in our data showed weaker negative health impacts than irregular sleepers.

Some previous studies have shown an association between sleep and academic performance such as GPA increasing with early bedtime and wake time [Eliasson, 2010], and with regular sleep patterns [Medeiros, 2003, Clerx, 2014]). Our results did not show any statistically significant relationship between sleep duration or regularity and GPA; however, if we tried the sleep behavioral difference between weekdays and weekends as Medeiros, we might be able to find the association between the sleep behaviors and academic performance.

8.1.3 Stress Recognition using Multi-modal Data

The features with the highest accuracy for recognition of high vs low PSS (82.4%) were the combination of sleep quality, and physical and mental health. This is consistent with what previous stress studies showed ([National Institute of Mental Health, 2015], the relationship between high PSS and high PSQI (N= 187 community-dwelling adults) [Buysee, 2008], and high PSS and low MCS (N109, doctor of pharmacy students) [Marshall, 2008]).

Sleep parameters ranked 3rd place of all modalities. Low sleep regularity, high PSQI (poor sleep) and long sleep duration (based from diary, but actiwatch based sleep duration was shorter in the high stress group than in the low stress group) contributed to the high stress group. Sleep parameters, and physiological responses were used in one study [Muaremi, 2013, 2014] to classify 3 stress levels and sleep duration was one of the most important features.

Consistent with previous studies [Bogomolov, 2014][Sano, 2013b][Sano,2015], personality types were one of the most influential factors for stress. The combination of low extraversion and low conscientiousness or agreeableness contributed to the high stress group and these directions of the associations were consistent with the prior work [Ebstrup, 2011] (N=3471).

Another study "Student Life" which collected college student daily life data (N=48) for 10 weeks using smart phones investigated the relationship between the Perceived Stress Scale and mobile phone use patterns [Wang, 2014]. Their results found associations between higher conversation frequency (day and night) and longer conversation duration (day), and lower PSS, and longer sleep duration and lower PSS. We found that our high PSS (high stress) group has higher PSQI (poor sleep quality), more frequent screen-on between 0-3am and more phone calls in the high stress group. Higher percentages of 3-6am outgoing SMS over the total number of outgoing SMS and 3-6am screen on were indicators for the high stress group staying up late.

Our data also showed that EDA features worked better than skin temperature for recognizing stress. We found a higher standard deviation of EDA amplitude (but not the number of peaks) while sitting was related to higher stress level, which means the high stress group has EDA responses rather than constant low or high responses.

Even though we computed objective features that are relatively easy to interpret this time for this thesis, it is still hard to interpret the implication of some of the selected objective features. In addition, these results do not tell us the causality (e.g. sleep poorly because of high stress or high stress because of poor sleep). We need to continue our further analysis to understand the meaning and the causal relationship.

8.2 Limitations

Here we describe some limitations in our study. First, our study was designed to get data to build models and was not designed to be a randomized control trial to compare a control group and a treatment group; therefore, we cannot clarify causality among parameters. Instead, we observe real world data over time and find correlations and predictors in the complex datasets. For example we can obtain findings of correlations among variables, and also we can use machine learning algorithms to train and test different mappings between behaviors and outcomes. With results from one semester's cohort, we can then hypothesize that findings from that cohort will apply to an independent cohort from a future semester. Then we can test the findings on that future semester to see if they are predictive of outcomes in that group.

Second, our observation and population are limited. We have collected multimodal and long-term data; however, we have some factors we have not measured that could influence sleep, stress, mental health and other wellbeing-related measures. Our population is also limited; we targeted MIT undergraduate students who are Android users because of access to their mobile phone usage. We have a great number of iphone users we were unable to include in our study. Our participants also have social interactions with people outside the study. We collected who they spend time with in their daily lives; however, survey and sensor measurement have been conducted only with the study participants. We do not know how people outside the study influence the participants' health. About the relationship between their behaviors and GPA, we monitored a month of data in the middle of each semester, so the dataset does not necessarily explain their behaviors in the entire semester such as a period when they have final exams.

Wrist band measurement could limit the ambulatory data. For example, light exposure measured on the wrist might not be 100% accurate because the sensor can be covered by a sleeve. Physiology could be noisy with wrist movements; however, we select devices that the participants can wear 24/7 in their daily life with the hopes that substantial portions of the day provide valuable data.

Lastly, findings from the datasets might be only limited to the datasets we collect and be hard to generalize to other populations. However, more importantly, a key contribution of this thesis is its comprehensive approach to collecting and analyzing real-world continuous 24/7 physiological, behavioral, and social interaction dataset. The approach we developed and used here may be applied broadly.

8.3 Contributions

The contributions of this thesis are

- We presented the characterization of electrodermal activity (EDA) with dry wrist-worn electrodes during natural and lab sleep.
- We compared thresholds for detecting EDA peaks and establish criteria for EDA storms in natural sleep.
- We found that more than 80% of the EDA peaks occurred during slow-wave and non-REM stage 2 sleep and that EDA amplitude is higher in SWS than in other sleep stages.
- We examined sleep-wake classification with multi-modal wrist wearable sensor features (acceleration, skin temperature, and skin conductance), compared the accuracy for intra-subject and inter-subject classification and identified the best subset of features.
- We examined predictors for sleep-related memory consolidation using multi-modal wrist wearable sensor data and found that EDA features worked better than EEG and actigraphy features.
- We designed a long-term study for gathering rich multimodal data in daily life and established study procedures from preparation, recruiting, data cleaning and analysis.
- We collected 30-day physiological, social, and behavioral data from 168 participants, building a valuable dataset of over 100,000 hours of real world physiological and behavioral data.
- Our initial analysis revealed characteristics about sleep, stress, other wellbeing-related measures, personality type, and mobile phone usage about our population.
- We showed that irregular sleep over several weeks has statistically significantly more bad health outcomes (morning low alertness and energy, and low mental health) than regular sleepers, even after controlling gender and stress. Moreover, we showed that sleep duration (after controlling for irregularity) was not statistically significantly associated with these negative effects.
- We identified features useful for monthly reported perceived stress (high vs low): daily activities, personality, sleep, physiology, social interactions, phone usage, and mobility

8.4 Future Work

This study has a lot of future work and our analysis is being continued. Also, the SNAPSHOT study is ongoing and we will keep collecting data and applying different analyses techniques.

We are planning to recruit populations who are early risers for their sport practice (such as rowers) and analyze how their behaviors are different from other populations at MIT. One of our hypotheses is that early risers have a more regular sleep schedule, and higher wellbeing-related scores.

We also plan to examine the following additional research questions with this dataset:

- How are late night mobile phone usage and light exposure related to stress and wellbeing-related measures? (2.1.4)
- How is social interaction related to healthiness and happiness? (2.5)
- Can we find bio/behavioral/trait markers for mood and mental health? (2.4, 2.7)

This study focuses on collecting real world data and understanding the natural behaviors; however, we are also interested in adding interventions to the protocol to investigate how we can change behaviors. Visualization of the data or advice/feedback could help them to improve their sleep patterns, mental health and stress.

Another aspect of future work is making this study more deployable. This study collects so many things within one month. We are interested in scaling this type of study to collect long-term data in a larger scale student population and in different populations such as employees to identify causal and predictive features, especially for mood, stress and mental health.

Lastly, there are still different factors we have not measured, which could influence sleep, stress and wellbeing-related measures such as food intake, face to face interaction, and weather. We can also measure heart rate, respiration, blood data such as cortisol and blood sugar level, and even genome data to make our understanding richer and deeper.

8.5 Conclusions

This thesis develops a methodology to measure and analyze ambulatory physiological, behavioral and social data from wearable sensors and mobile phones with trait data such as personality types, for learning about behaviors and traits that impact human health and wellbeing. This thesis describes the development of software tools to measure and analyze ambulatory multi-modal data, study design and data collection and validation. This thesis also validates the methodology and tools on a selected subset of the questions that can be answered by the data collected.

First, I characterized sleep wrist EDA by comparing it with sleep palm EDA and sleep stages quantitatively. Next, I applied our EDA analysis methods to sleep-wake classification and sleep-related memory consolidation using wrist sensor data including EDA and found the best subsets of features for these tasks.

Lastly, with colleagues at MIT and the Brigham Women's Hospital, I designed and ran "the SNAPSHOT" study. We measured one month of multi-modal data from each of 168 undergraduate students including continuous physiology, behavioral and social data using wearable sensors and mobile phones as well as measurement in the laboratory and standardized surveys. As a first step toward showing the value of this new large datasets, I analyzed the effect of sleep regularity and sleep duration on academic performance, physical/mental health score (Physical Component Score and Mental Component Score from SF-12), Perceived Stress Scale and subjective daily wellbeing-related scores (alertness, happiness, sluggishness, healthiness and calmness) using coarsened exact matching to control several covariates. Our data showed that sleep irregularity was associated with low mental health and with negative health outcomes (low energy level and low alertness in the morning), when controlling for sleep duration. Moreover, sleep duration (after controlling for irregularity) was not statistically significantly associated with these negative effects. Overall, our data suggest that emphasizing duration of sleep for improving mental health is not as important as emphasizing keeping a regular sleep schedule. I also examined monthly reported perceived stress (high vs low) recognition using multi-modal data from surveys, wearable sensors and mobile phones and identified useful features: daily activities, personality, sleep, physiology, social interactions, phone usage, and mobility.

Appendix

A.Surveys

A.1 Pre-Study Screening Questionnaire

1) What is your full name?

_____

Are you 18 years of age?

Yes

No

2) What is your date of birth?

_____

((MM-DD-YYYY))

3) a) Please indicate your gender:

Male

Female

b) Are you currently pregnant?

Yes

No

4) Do you have a mobile/cell phone?

Yes

No

a) What is your mobile/cell phone number?

_____

b) Type of Phone:

Android

iPhone

Other

c) Please indicate the type of phone:

_____

d) Operating System (OS) Version: Note: - OS Version for iPhone can be checked by going to Settings-->General-->About-->Version - OS Version for Android can be checked by going to "About Phone" in the

"Settings" section

_____

e) Do you have a data plan on your mobile/cell phone?

Yes (unlimited)

Yes (limited)

No

f) Do you feel restricted in the use of your mobile/cell phone because you have a limited data plan?

Yes

No

g) Do you use secondary calling/messaging services?

(Example: Google Voice, Skype, Whatsapp, etc...)

Yes

No

h) Please indicate...

- all secondary calling/messaging services you use.

- what you use these services for (example: texting, voice calls, video chatting, etc...)

- what percentage of your TOTAL communication is done on

EACH of the indicated applications (example: 5% google voice & 10% Skype)

i) Would you be willing to discontinue use of ALL indicated services for the 30 day duration of the study?

Yes

No

5) Enter all email addresses you actively use.

6) a) Are you currently attending college?

Yes

No

b) What year of college are you currently in?

Freshman

Sophomore

Junior

Senior

Graduate

c) Please indicate what college you are currently attending:

Harvard University

Massachusetts Institute of Technology (MIT)

Other

7) a) Have you been in another time zone in the last 7 days?

Yes

No

b) Where were you?

_____

8) a) Are you planning to travel across more than one time zone between now and the end of the semester?

Yes

No

projectredcap.org

Confidential

b) How many times will you be traveling across more than one time zone between now and the end of the semester?

1

2-3

4 or more

1st Location:

Where:

_____

Start of travel:

_____

End of travel:

2nd Location:

Where:

Start of travel:

End of travel:

3rd Location:

Where:

Start of travel:

End of travel:

Please provide the location and dates for any traveling you will be doing across more than one time zone between now and the end of the semester

AFTER the 3rd trip:

A.2 Start of Study Questionnaire

Background Information

1) Academic major:

Aerospace Engineering

Aerospace Engineering with Info. Tech.

Anthropology

Archaeology & Materials

Architecture

Biological Engineering

Biology

Brain & Cognitive Sciences

Chemical Biological Engineering

Chemical Engineering

Chemistry
Civil & Environmental Engineering
Civil Engineering
Comparative Media Studies
Computer Science & Engineering
Computer Science & Molecular Biology
Atmospheric & Planetary Sciences
Economics
Electrical Engineering & Computer Science
Electrical Science & Engineering
ENG - Chemical Engineering
ENG – Engineering as recommended by the Dept of Mechanical Engineering
Environmental Engineering Science
Foreign Languages & Literatures
History
Humanities and Engineering
Humanities and Science
Interdisciplinary Major in Humanities
Linguistics & Philosophy
Literature
Management Science
Materials Science & Engineering
Mathematics
Mathematics with Computer Science
Mechanical & Ocean Engineering
Mechanical Engineering
Mechanical Engineering- Intern
Music
Nuclear Science & Engineering
Philosophy
Physics
Planning
Political Science
Writing

Undesignated

Other

Please specify your major:

_____

2) GPA for the last semester :

_____

(Range (0.0 - 5.0))

3) Where do you live?

Dorm

Independent living group (if applicable)

Off campus

Name of dorm or independent living group:

_____

4) Do you live alone or with roommates/flatmates?

Alone

With roommates/flatmates

How many roommates/flatmates do you live with?

_____

5) Do you share your bedroom with someone?

Yes

No

Sometimes

How many people do you share your bedroom with?

_____

6) What is the WIFI mac address (e.g.50:26:90:77:0d:67) on your smart phone?

Note: - For iPhone, go to Settings-->General-->About

- For Androids, go to Settings-->About Phone or About Device-->Status

- If you don't have a smart phone, enter "not applicable"

_____

Health Survey

This survey asks for your views about your health. This information will help you keep track of how you feel and how well you are able to do your usual activities.

1) In general, would you say your health is:

Excellent

Very good

Good

Fair

Poor

2) The following questions are about activities you might do during a typical day.

Does your health now limit you in these activities? If so, how much?

a) Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf

Yes, limited a lot

Yes, limited a little

No, not limited at all

b) Climbing several flights of stairs

Yes, limited a lot

Yes, limited a little

No, not limited at all

3) During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

a) Accomplished less than you would like

All of the time

Most of the time

Some of the time

A little of the time

None of the time

b) Were limited in the kind of work or other activities

All of the time

Most of the time

Some of the time

A little of the time

None of the time

4) During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

a) Accomplished less than you would like

All of the time

Most of the time

Some of the time

A little of the time

None of the time

b) Did work or activities less carefully than usual

All of the time

Most of the time

Some of the time

A little of the time

None of the time

5) During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

Not at all

A little bit

Moderately

Quite a bit

Extremely

6) These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the past 4 weeks...

a) Have you felt calm and peaceful?

All of the time

Most of the time

A good bit of the time

Some of the time

A little of the time

None of the time

b) Did you have a lot of energy?

All of the time

Most of the time

A good bit of the time

Some of the time

A little of the time

None of the time

c) Have you felt downhearted and blue?

All of the time

Most of the time

A good bit of the time

Some of the time

A little of the time

None of the time

7) During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?

All of the time

Most of the time

A good bit of the time

Some of the time

A little of the time

None of the time

Question 8 is for females only.

Please choose the first day of bleeding and the last day of bleeding for each of your menstrual cycles in the last 6 weeks.

Note: When selecting a date for the below questions, select the calendar icon and choose the month and year FIRST. Then, pick the day of the month.

8) a) Menstrual Cycle 1: FIRST day of bleeding

_____

a) Menstrual Cycle 1: LAST day of bleeding

_____

b) Menstrual Cycle 2: FIRST day of bleeding

_____

b) Menstrual Cycle 2: LAST day of bleeding

_____

c) Menstrual Cycle 3: FIRST day of bleeding

_____

c) Menstrual Cycle 3: LAST day of bleeding

Munich Questionnaire

What is your height?

Feet:

Inches:

What is your weight (POUNDS)?

I have a regular school schedule:

Yes

No

If 'YES', how many days per week?

Instructions: Please complete all of the following sections, regardless of whether you are in school on a regular basis or not. Use the 24 hour scale, for example 23:00 instead of 11:00PM!!!!

SCHOOL DAYS

On school days...

I go to bed at _____ o'clock. Note: Please use military time.

Note that some people stay awake for some time when in bed! I actually get ready to fall asleep at _____o'clock. Note: Please use military time.

I need _____ minutes to fall asleep.

I wake up at _____ o'clock. Note: Please use military time.

I wake up...

with an alarm clock

without an alarm clock

After _____ minutes, I get up.

FREE DAYS

On free days...

I go to bed at _____ o'clock. Note: Please use military time.

Note that some people stay awake for some time when in bed! I actually get ready to fall asleep at _____o'clock. Note: Please use military time.

I need _____ minutes to fall asleep.

I wake up at _____ o'clock. Note: Please use military time.

I wake up....

with an alarm clock

without an alarm clock

After _____ minutes, I get up.

Comment Field: Please leave a comment if you currently have NO possibility of freely choosing your sleep times (e.g. because of pet(s), child(ren) etc.). Use this field also to provide additional information, if the system asks for it:

DAYLIGHT EXPOSURE

On average, I spend the following amount of time outdoors in daylight (without a roof above my head):

On school days... HOURS:

MINUTES:

On free days... HOURS:

MINUTES:

PERSONAL DATA ON POTENTIAL RESEARCH SUBJECTS

Date of birth: Note: When selecting a date for this question, select the calendar icon and choose the month and year FIRST. Then, pick the day of the month.

_____

Gender:

Male

Female

Ethnic Categories: Note: Hispanic or Latino: A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race.

Hispanic or Latino

Not Hispanic or Latino

Race (please select one or more):

American Indian or Alaskan Native

Asian

Black or African American

Native Hawaiian or other Pacific Islander

White

Other

Please specify:

_____

NOTE: The categories that most closely reflect the individual's recognition in the community should be used for purposes of reporting mixed racial and/or ethnic origins. Definitions are as follows:

American Indian or Alaskan Native: A person having origins in any of the original peoples of North, Central, or South America, and maintains tribal affiliations or community attachment.

Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

Black or African American: A person having origins in any of the black racial groups of Africa.

Native Hawaiian or Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

White: A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.

A.3 End of Study Questionnaire

Health Survey

This survey asks for your views about your health. This information will help you keep track of how you feel and how well you are able to do your usual activities.

Please answer every question by selecting the answer as indicated. Although it is hoped that you will answer all of the questions in the survey, you may skip any question that you do not feel comfortable answering. If you are unsure about how to answer a question, please give the best answer you can.

1) In general, would you say your health is:

Excellent

Very good

Good

Fair

Poor

2) The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much? a) Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf

Yes, limited a lot

Yes, limited a little

No, not limited at all

b) Climbing several flights of stairs

Yes, limited a lot

Yes, limited a little

No, not limited at all

3) During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health? a) Accomplished less than you would like

All of the time

Most of the time

Some of the time

A little of the time

None of the time

b) Were limited in the kind of work or other activities

All of the time

Most of the time

Some of the time

A little of the time
None of the time

4) During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?
a) Accomplished less than you would like
All of the time
Most of the time
Some of the time
A little of the time
None of the time
b) Did work or activities less carefully than usual
All of the time
Most of the time
Some of the time
A little of the time
None of the time
5) During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?
Not at all
A little bit
Moderately
Quite a bit
Extremely
6) These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks...
a) Have you felt calm and peaceful?
All of the time
Most of the time
A good bit of the time
Some of the time
A little of the time

None of the time

b) Did you have a lot of energy?

All of the time

Most of the time

A good bit of the time

Some of the time

A little of the time

None of the time

c) Have you felt downhearted and blue?

All of the time

Most of the time

A good bit of the time

Some of the time

A little of the time

None of the time

7) During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?

All of the time

Most of the time

A good bit of the time

Some of the time

A little of the time

None of the time

A.4 Perceived Stress Scale- 10 Item

Instructions: The questions in this scale ask you about your feelings and thoughts during the last month. In each case, please indicate with a check how often you felt or thought a certain way.

* Required

Your User ID *

0=never, 1=almost never, 2=sometimes, 3=fairly often, 4=very often

In the last month, how often have you been upset because of something that happened unexpectedly?

In the last month, how often have you felt that you were unable to control the important things in your life?

In the last month, how often have you felt nervous and "stressed"?

In the last month, how often have you felt confident about your ability to handle your personal problems?

In the last month, how often have you felt that things were going your way?

In the last month, how often have you found that you could not cope with all the things that you had to do?

In the last month, how often have you been able to control irritations in your life?

In the last month, how often have you felt that you were on top of things?

In the last month, how often have you been angered because of things that were outside of your control?

In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?

A.5 SNAPSHOT Pre- and Post-experimental survey - social network survey

Please provide FIRST NAME and LAST NAME (no more than 3 people) for each question.(e.g. Tom Hedman).You can name the same people in more than one occasion, or none, if it's really the answer. This information is going to be used to help create a computer model of your social interactions.

* Required

1) Name your roommates/flatmates? (if any)

Roommate 1: First Name

Roommate 1: Last Name

Roommate 2: First Name

Roommate 2: Last Name

Roommate 3: First Name

Roommate 3: Last Name

2) If a natural disaster/tragedy affects Boston, who would you call first? second?

Person 1: First Name

Person 1: Last Name

Person 2: First Name

Person 2: Last Name

Person 3: First Name

Person 3: Last Name

3) Who do you talk to about personal matters (love life, concerns, family matters, etc)?

Person 1: First Name

Person 1: Last Name

Person 2: First Name

Person 2: Last Name

Person 3: First Name

Person 3: Last Name

4) Who do you talk to about work/research/classes?

Person 1: First Name

Person 1: Last Name

Person 2: First Name

Person 2: Last Name

Person 3: First Name

Person 3: Last Name

5) Who do you talk to about friends and other people you know?

Person 1: First Name
Person 1: Last Name
Person 2: First Name
Person 2: Last Name
Person 3: First Name
Person 3: Last Name

6) Who do you talk to about media/entertainment (sports, movies, tech gadgets, music, video games)?

Person 1: First Name
Person 1: Last Name
Person 2: First Name
Person 2: Last Name
Person 3: First Name
Person 3: Last Name

7) If you have to move to a new apartment/house, who would you ask for help?

Person 1: First Name
Person 1: Last Name
Person 2: First Name
Person 2: Last Name
Person 3: First Name
Person 3: Last Name

8) Who do you study with?

Person 1: First Name
Person 1: Last Name
Person 2: First Name
Person 2: Last Name
Person 3: First Name
Person 3: Last Name

9) Who do you go shop, party or play with (incl. video games)?

Person 1: First Name
Person 1: Last Name
Person 2: First Name
Person 2: Last Name
Person 3: First Name
Person 3: Last Name

10) Who do you share ideas with?

Person 1: First Name
Person 1: Last Name
Person 2: First Name
Person 2: Last Name
Person 3: First Name
Person 3: Last Name

11) Who do you often disagree with (real person, not media character)?

Person 1: First Name
Person 1: Last Name
Person 2: First Name
Person 2: Last Name
Person 3: First Name
Person 3: Last Name

12) Who spends the most time at your apartment/dorm excluding the people you live with?

Person 1: First Name
Person 1: Last Name
Person 2: First Name
Person 2: Last Name
Person 3: First Name
Person 3: Last Name

13) Do you hang out at someone else's apartment/dorm? If yes, whose?

Person 1: First Name

Person 1: Last Name

Person 2: First Name

Person 2: Last Name

Person 3: First Name

Person 3: Last Name

Your ID *

A.5 Big Five personality test

Take this psychology test to find out about your personality! This test measures what many psychologists consider to be the five fundamental dimensions of personality.

Learn more about the Big Five by reading answers to commonly asked questions.
Read our consent form, which explains the benefits of this free, anonymous test and your rights.
There are no "right" or "wrong" answers, but note that you will not obtain meaningful results unless you answer the questions seriously.
These results are being used in scientific research, so please try to give accurate answers.
Your results will be displayed as soon as you submit your answers.
As you are rating yourself, you are encouraged to rate another person. By rating someone else you will tend to receive a more accurate assessment of your own personality. Also, you will be given a personality profile for the person you rate, which will allow you to compare yourself to this person on each of five basic personality dimensions. Try to rate someone whom you know well, such as a close friend, coworker, or family member.
If you would like to compare your personality to another person's, please select how you are related to the other person.

Directions: The following statements concern your perception about yourself in a variety of situations. Your task is to indicate the strength of your agreement with each statement, utilizing a scale in which 1 denotes strong disagreement, 5 denotes strong agreement, and 2, 3, and 4 represent intermediate judgments. In the boxes after each statement, click a number from 1 to 5 from the following scale:

Strongly disagree

Disagree

Neither disagree nor agree

Agree

Strongly agree

There are no "right" or "wrong" answers, so select the number that most closely reflects you on each statement. Take your time and consider each statement carefully. Once you have completed all questions click "Submit" at the bottom.

I see myself as someone who...

1. ...Is talkative

Strongly Disagree

1 2 3 4 5

Strongly Agree

2. ...Tends to find fault with others

Strongly Disagree

1 2 3 4 5

Strongly Agree

3. ...Does a thorough job

Strongly Disagree

1 2 3 4 5

Strongly Agree

4. ...Is depressed, blue

Strongly Disagree

1 2 3 4 5

Strongly Agree

5. ...Is original, comes up with new ideas

Strongly Disagree

1 2 3 4 5

Strongly Agree

6. ...Is reserved

Strongly Disagree

1 2 3 4 5

Strongly Agree

7. ...Is helpful and unselfish with others

Strongly Disagree

1 2 3 4 5

Strongly Agree

8. ...Can be somewhat careless

Strongly Disagree

1 2 3 4 5

Strongly Agree

9. ...Is relaxed, handles stress well

Strongly Disagree

1 2 3 4 5

Strongly Agree

10. ...Is curious about many different things

Strongly Disagree

1 2 3 4 5

Strongly Agree

11. ...Is full of energy

Strongly Disagree

1 2 3 4 5

Strongly Agree

12. ...Starts quarrels with others

Strongly Disagree

1 2 3 4 5

Strongly Agree

13. ...Is a reliable worker

Strongly Disagree

1 2 3 4 5

Strongly Agree

14. ...Can be tense

Strongly Disagree

1 2 3 4 5

Strongly Agree

15. ...Is ingenious, a deep thinker

Strongly Disagree

1 2 3 4 5

Strongly Agree

16. ...Generates a lot of enthusiasm

Strongly Disagree

1 2 3 4 5

Strongly Agree

17. ...Has a forgiving nature

Strongly Disagree

1 2 3 4 5

Strongly Agree

18. ...Tends to be disorganized

Strongly Disagree

1 2 3 4 5

Strongly Agree

19. ...Worries a lot

Strongly Disagree

1 2 3 4 5

Strongly Agree

20. ...Has an active imagination

Strongly Disagree

1 2 3 4 5

Strongly Agree

21. ...Tends to be quiet

Strongly Disagree

1 2 3 4 5

Strongly Agree

22. ...Is generally trusting

Strongly Disagree

1 2 3 4 5

Strongly Agree

23. ...Tends to be lazy

Strongly Disagree

1 2 3 4 5

Strongly Agree

24. ...Is emotionally stable, not easily upset

Strongly Disagree

1 2 3 4 5

Strongly Agree

25. ...Is inventive

Strongly Disagree

1 2 3 4 5

Strongly Agree

26. ...Has an assertive personality

Strongly Disagree

1 2 3 4 5

Strongly Agree

27. ...Can be cold and aloof

Strongly Disagree

1 2 3 4 5

Strongly Agree

28. ...Perseveres until the task is finished

Strongly Disagree

1 2 3 4 5

Strongly Agree

29. ...Can be moody

Strongly Disagree

1 2 3 4 5

Strongly Agree

30. ...Values artistic, aesthetic experiences

Strongly Disagree

1 2 3 4 5

Strongly Agree

31. ...Is sometimes shy, inhibited

Strongly Disagree

1 2 3 4 5

Strongly Agree

32. ...Is considerate and kind to almost everyone

Strongly Disagree

1 2 3 4 5

Strongly Agree

33. ...Does things efficiently

Strongly Disagree

1 2 3 4 5

Strongly Agree

34. ...Remains calm in tense situations

Strongly Disagree

1 2 3 4 5

Strongly Agree

35. ...Prefers work that is routine

Strongly Disagree

1 2 3 4 5

Strongly Agree

36. ...Is outgoing, sociable

Strongly Disagree

1 2 3 4 5

Strongly Agree

37. ...Is sometimes rude to others

Strongly Disagree

1 2 3 4 5

Strongly Agree

38. ...Makes plans and follows through with them

Strongly Disagree

1 2 3 4 5

Strongly Agree

39. ...Gets nervous easily

Strongly Disagree

1 2 3 4 5

Strongly Agree

40. ...Likes to reflect, play with ideas

Strongly Disagree

1 2 3 4 5

Strongly Agree

41. ...Has few artistic interests

Strongly Disagree

1 2 3 4 5

Strongly Agree

42. ...Likes to cooperate with others

Strongly Disagree

1 2 3 4 5

Strongly Agree

43. ...Is easily distracted

Strongly Disagree

1 2 3 4 5

Strongly Agree

44. ...Is sophisticated in art, music, or literature

Strongly Disagree

1 2 3 4 5

Strongly Agree

45. ...Is politically liberal

Strongly Disagree

1 2 3 4 5

Strongly Agree

46. ...Has high self-esteem

Strongly Disagree

1 2 3 4 5

Strongly Agree

A.6 Morning daily diary

Please check if you did not sleep in the past 24 hours

What time did you try to fall asleep?

How long did it take you to fall asleep?

In the 60 minutes before trying to fall asleep, did you spend at least 5 minutes doing any of the following?

Interacting with people in person

Interacting with people through electronic media (e.g. emails, calls, SMS, skype, chat, online games)

Did not interact with people

How did you finally wake up?

Spontaneously awoke

Awoken by an alarm

Awoken by another disturbance

What time did you finally wake up?

How many times did you awaken?

Not counting your final awakening / wake up.

List each: when you woke and for how long? [put extra awakenings under 'Comments']

Awakening 1 start

How long were you awake for?

Awakening 2 start

How long were you awake for?

Did you nap yesterday?

Yes/No

How many times did you nap yesterday?

List each: when the nap started and for how long [put extra naps under 'Comments']

Nap 1 start

How long was the nap?

Did you remove your actiwatch?

Yes/No

When is your first scheduled academic or extracurricular event today, if any?

Check here for no scheduled activity

For each of the following, indicate how you feel right now by clicking on each line and adjusting the sliders.

Sleepy-Alert

Sad-Happy

Sluggish-Energetic

Sick-Healthy

Stressed Out-Calm Relaxed

Please enter any comments for today

A.7 Evening daily diary

Did you attend any academic activities today (including classes, e-classes, sections, seminars, labs, study groups)?

Yes/No

How many?

List each: when the activity started and for how long? [put extra events under 'Comments']

Activity 1 start

For how long?

Activity 2 start

For how long?

Activity 3 start

For how long?

How many hours did you study by yourself today, not including any of the academic activities in the previous question?

Did you engage in any exercise-based activities today (including sports, gym, cycling, etc.)?

Yes/No

How many times?

List each: when the activity started and for how long [put extra under 'Comments']

Activity 1 start

For how long?

Did you attend any other extracurricular activities today, besides academic activities and exercise-based activities?

Yes/No

Did you miss or were you late for any scheduled events (e.g., academic activities, exercise-based activities, other extracurricular activities) because you overslept today?

Yes/No

How many total servings of caffeine did you have today?

When is the latest hour you consumed caffeine?

Please refer to this guide to help calculate servings

Cola 12 oz (1 can)	1/2 serving
Tea (1 cup)	1/2 serving
Home brew coffee 8 oz (1 cup)	1 serving
Energy drink 16 oz (1 large can)	2 servings
5 hour energy	2 servings
Large coffee 16 oz (large or grande)	2 servings
Caffeine pill - 100mg	1 serving
Caffeine pill - 200mg	2 servings

Besides caffeine, did you use any other medications, drugs, or alcohol today?

Yes/No

For each of the following, indicate how you feel right now by clicking on each line and adjusting the sliders.

Sleepy-Alert

Sad-Happy

Sluggish-Energetic

Sick-Healthy

Stressed Out-Calm Relaxed

In addition, did you have an emotionally charged interaction with someone today?

No

Yes, a memorable positive interaction

Yes, a somewhat negative interaction

Yes, and it was very negative

Please enter any comments for today

B. Participant Demographics

Here, we summarize demographic information of our population (Tables B.1-B.6)

Table B.1 Gender

Gender	#
M	107
F	62

Table B.2 Age

Age	#
17	1
18	41
19	51
20	26
21	34
22	8
23	3
24	1
25	1
26	0
Not reported	3

Table B.3 School year

Year	#
1	50
2	43
3	35
4	31
Not reported	10

Table B.4 Major

Major	#
Aerospace Engineering	6
Atmospheric & Planetary Sciences	1
Biology	8
Biological Engineering	9
Brain & Cognitive Sciences	3
Chemistry	1
Chemical Engineering	9
Chemical Biological Engineering	5
Civil & Environmental Engineering	3
Computer Science & Engineering	16
Computer Science & Molecular Biology	1
Economics	1
Electrical Engineering & Computer Science	24
Electrical Science & Engineering	2
Management Science	2
Materials Science & Engineering	8
Mathematics with Computer Science	2
Mathematics	2
Mechanical Engineering	34
Mechanical & Ocean Engineering	1
Nuclear Science & Engineering	2
Philosophy	1
Physics	10
Political Science	1
Undesignated	14
Not reported	3

Table B.5 Self-Reported Ethnicity and Gender of All Enrolled Participants

Ethnic Category	Males	Females	Unknown	Total
Hispanic	32	10	1	43
Not Hispanic or Latino	71	52	0	123
Unknown	3	0	0	3
Total of all Enrolled Participants	106	62	1	169

Table B.6 Self-Reported Race and Gender of All Enrolled Participants

Ethnic Category	Males	Females	Unknown	Total
American Indian/Alaska Native	0	1	0	1
Asian	16	14	0	30
Native Hawaiian or Other Pacific Islander	0	0	0	0
Black or African American	16	7	0	23
White	51	34	0	85
More than one race	11	4	0	15
Unknown or not reported	12	2	1	15
Total of all Enrolled Participants	106	62	1	169

Table B.7 Answers to questions in SF-12 related to MCS (Mental Component Score) (%)

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?			
	[%]	a) Accomplished less than you would like	b) Did work or activities less carefully than usual
Pre	None of the time	40.2	40.8
	A little of the time	27.8	30.8
	Some of the time	22.5	23.1
	Most of the time	6.5	4.7
	All of the time	3.0	0.6
Post	None of the time	22.5	24.3
	A little of the time	29.6	30.8
	Some of the time	27.8	30.8
	Most of the time	10.1	5.9
	All of the time	4.7	3.0

		How much of the time during the past 4 weeks...			
		a) Have you felt calm and peaceful?	b) Did you have a lot of energy?	c) Have you felt downhearted and blue?	During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?
Pre	None of the time	0.6	0.0	10.8	51.5
	A little of the time	10.2	7.7	52.1	28.1
	Some of the time	35.3	37.5	24.6	14.4
	Most of the time	52.1	53.0	12.6	6.0
	All of the time	1.8	1.8	0.0	0.0
Post	None of the time	1.8	1.2	7.8	31.7
	A little of the time	23.4	14.9	33.5	31.1
	Some of the time	29.3	44.0	34.1	23.4
	Most of the time	39.5	32.7	19.8	8.4
	All of the time	2.4	2.4	0.6	1.2

C.Study Attrition and Completion Rate

As we mentioned in section 5.2, 169 participants consented to join the study and 7 participants dropped out. The reasons the 7 participants dropped out were (1) one participant was under 18 (2) one participant crossed more than one time zone (3) 5 participants decided not to continue the study because two participants were too busy with school, one did not want to wear the sensors, one broke his android phone and the other one concerned about the collected data security)

Our diary entry completion rate was 92% for the 2014 Spring cohort, 97% for the 2014 Fall cohort and 95% for the 2015 Spring cohort. We obtained 71% and 85% of full datasets (Q-sensor, actiwatch, overnight study, daily diaries, and phone application) for 2014 Fall and 2015 Spring. We obtained 82% of 30-day Q-sensor data (one day was counted if measurement was done for more than 18 hours per day) and 83% of 30-day phone data in 2014 Fall, and 85% for Q-sensor data and 91% for phone data in 2015 Spring Study completion rate was 97% in 2014 Fall and 96% in 2015 Spring.

References

K. Adam, M. Tomeny, and I. Oswald, "Physiological and psychological differences between good and poor sleepers," J. Psychiatr. Res., vol. 20, no. 4, pp. 301–316, Jan. 1986.

J. K. Aggarwal and M. S. Ryoo, "Human activity analysis," ACM Computing Surveys, vol. 43, no. 3. pp. 1–43, 2011.

N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland, "Social fMRI: Investigating and shaping social mechanisms in the real world," Pervasive Mob. Comput., vol. 7, no. 6, pp. 643–659, Dec. 2011.

S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak, "The role of actigraphy in the study of sleep and circadian rhythms.," Sleep, vol. 26, no. 3, pp. 342–392, 2003.

K. Asahina and K. Omura, "Phenomenological study of paradoxical phase and reverse of sleep," Jpn. J. Physiol., vol. 14, pp. 365–72, Aug. 1964.

The Associated Press and MTV, http://www.halfofus.com/wp-content/uploads/2013/10/mtvU-AP-2009-Economy-College-Stress-and-Mental-Health-Poll-Executive-Summary-May-2009.pdf, 2009.

J. H. Baek, J. S. Kim, M. J. Kim, S. Ryu, K. Lee, K. Ha, and K. S. Hong, "Lifetime Characteristics of Evening-Preference and Irregular Bed-Rise Time Are Associated With Lifetime Seasonal Variation of Mood and Behavior: Comparison Between Individuals With Bipolar Disorder and Healthy Controls," Behav. Sleep Med., pp. 1–14, Nov. 2014.

B. Barbini, F. Benedetti, C. Colombo, E. Guglielmo, E. Campori, and E. Smeraldi, "Perceived mood and skin body temperature rhythm in depression," Eur. Arch. Psychiatry Clin. Neurosci., vol. 248, no. 3, pp. 157–160, 1998.

B. Bei, J. F. Wiley, J. Trinder, and R. Manber, "Beyond the mean: A systematic review on the correlates of daily intraindividual variability of sleep/wake patterns," Sleep Med. Rev., Jul. 2015.

G. Belenky, N. J. Wesensten, D. R. Thorne, M. L. Thomas, H. C. Sing, D. P. Redmond, M. B. Russo, and T. J. Balkin, "Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study.," *J. Sleep Res.*, vol. 12, no. 1, pp. 1–12, Mar. 2003.

T. Blackwell, S. Ancoli-Israel, S. Redline, and K. L. Stone, "Factors that may influence the classification of sleep-wake by wrist actigraphy: the MrOS Sleep Study.," J. Clin. Sleep Med., vol. 7, no. 4, pp. 357–67, Aug. 2011.

A. Bogomolov, B. Lepri, F. B. Kessler, F. Pianesi, A. S. Pentland, and F. B. Kessler, "Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits," in ACM Multimedia 2014, 2014.

W. Boucsein, Electrodermal Activity. Springer, 1992.

M. E. J. Bouwmans, E. H. Bos, A. J. Oldehinkel, and P. de Jonge, "Temporal order of change in sleep quality and positive affect in major depressed patients and healthy controls", The 4th SAA Conference, Society for Ambulatory Assessment, 2015.

R. J. Broughton, R. Poire, and C. a Tassinari, "the Electrodermogram (Tarchanoff Effect) During Sleep.," Electroencephalogr. Clin. Neurophysiol., vol. 18, pp. 691–708, Jun. 1965.

J. Van den Bulck, "Television viewing, computer game playing, and Internet use and self-reported time to bed and time out of bed in secondary-school children.," *Sleep*, vol. 27, no. 1, pp. 101–4, Mar. 2004.

N. Burch, "Data Processing of psychophysiological recordings," in Symposium on the Analysis of Central Nervous System and Cardiovascular Data using Computer Methods, pp. 165–180, 1965.

D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer, "The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.," Psychiatry Res., vol. 28, no. 2, pp. 193–213, May 1989.

J. T. Cacioppo, J. H. Fowler, and N. A. Christakis, "Alone in the crowd: the structure and spread of loneliness in a large social network.," J. Pers. Soc. Psychol., vol. 97, no. 6, pp. 977–91, Dec. 2009.

C. Cajochen, K. Kräuchi, and A. Wirz-Justice, "Role of melatonin in the regulation of human circadian rhythms and sleep," Journal of Neuroendocrinology, vol. 15, no. 4. pp. 432–437, 2003.

C. Cajochen, S. Frey, D. Anders, J. Späti, M. Bues, A. Pross, R. Mager, A. Wirz-Justice, and O. Stefani, "Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance.," *J. Appl. Physiol.*, vol. 110, no. 5, pp. 1432–8, May 2011.

A.-M. Chang, D. Aeschbach, J. F. Duffy, and C. A. Czeisler, "Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness," *Proc. Natl. Acad. Sci.*, vol. 112, no. 4, p. 201418490, Dec. 2014.

L. G. Chepenik, L. A. Cornew, and M. J. Farah, "The influence of sad mood on cognition.," Emotion, vol. 7, no. 4, pp. 802–811, 2007.

S. Chernbumroong, A. S. Atkins, and H. Yu, "Activity classification using a single wrist-worn accelerometer," 2011 5th Int. Conf. Software, Knowl. Information, Ind. Manag. Appl. Proc., pp. 1–6, 2011.

N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years.," N. Engl. J. Med., vol. 357, no. 4, pp. 370–9, Jul. 2007.

J. K. Clark, "Relationship of personality and sleep to academic success in the United States Military Academy : a perspective utilizing the five factor model of personality," 2007.

W. Clerx, A. J. K. Phillips, S. Lockley, C. O'Brien, E. Klerman, and C. Czeisler. "Impact of irregularity of sleep-wake schedules on circadian phase and amplitude in college undergraduates," The 2014 Meeting of the Society for Research on Biological Rhythms, 2014.

W. Clerx, A. J. K. Phillips, S. Lockley, C. O'Brien, E. Klerman, and C. Czeisler. "Irregular Sleep in College Students: Consequences for Sleep Consolidation, Circadian Rhythms and Performance", SLEEP 2015, the 29th Annual Meeting of the Associated Professional Sleep Societies, 2015.

D. A. Cohen, W. Wang, J. K. Wyatt, R. E. Kronauer, D.-J. Dijk, C. A. Czeisler, and E. B. Klerman, "Uncovering residual effects of chronic sleep loss on human performance.," Sci. Transl. Med., vol. 2, no. 14, p. 14ra3, 2010.

S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress.," J. Health Soc. Behav., vol. 24, no. 4, pp. 385–96, Dec. 1983.

R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity.," Sleep, vol. 15, no. 5, pp. 461–9, Oct. 1992.

S. Daan, D. G. Beersma, and A. A. Borbély, "Timing of human sleep: recovery process gated by a circadian pacemaker.," Am. J. Physiol., vol. 246, no. 2 Pt 2, pp. R161–R183, 1984.

M. M. David, A. W. MacLean, J. B. Knowles, and M. E. Coulter, "Rapid eye movement latency and mood following a delay of bedtime in healthy subjects: do the effects mimic changes in depressive illness?," *Acta Psychiatr. Scand.*, vol. 84, no. 1, pp. 33–9, Jul. 1991.

M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system", in: T. Cacioppo, L.G.T.& G.G.B. (Ed.), Handbook of Psychophysiology. Cambridge University Press, pp. 159–181.2007.

S. S. Dickerson and M. E. Kemeny, "Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research.," *Psychol. Bull.*, vol. 130, no. 3, pp. 355–91, May 2004.

D.-J. Dijk and S. W. Lockley, "Integration of human sleep-wake regulation and circadian rhythmicity.," J. Appl. Physiol., vol. 92, no. 2, pp. 852–862, 2002.

R. K. Dishman, Y. Nakamura, M. E. Garcia, R. W. Thompson, A. L. Dunn, and S. N. Blair, "Heart rate variability, trait anxiety, and perceived stress among physically fit men and women," Int. J. Psychophysiol., vol. 37, no. 2, pp. 121–133, Aug. 2000.

S. Doberenz, W. Roth, and E. Wollburg, "Methodological considerations in ambulatory skin conductance monitoring," Int. J. ..., vol. 80, no. 2, pp. 87–95, 2011.

J. F. Duffy and D.-J. Dijk, "Getting through to circadian oscillators: why use constant routines?," J. Biol. Rhythms, vol. 17, no. 1, pp. 4–13, 2002.

J. F. Ebstrup, L. F. Eplov, C. Pisinger, and T. Jørgensen, "Association between the Five Factor personality traits and perceived stress: is the effect mediated by general self-efficacy?," *Anxiety. Stress. Coping*, vol. 24, no. 4, pp. 407–19, 2011.

S. Eggermont and J. Van Den Bulck, "Nodding off or switching off? The use of popular media as a sleep aid in secondary-school children," J. Paediatr. Child Health, vol. 42, no. 7, pp. 428–433, 2006.

A. H. Eliasson and C. J. Lettieri, "Early to bed, early to rise! Sleep habits and academic performance in college students," Sleep Breath., vol. 14, no. 1, pp. 71–75, 2010.

S. Elsenbruch, M. J. Harnish, and W. C. Orr, "Heart rate variability during waking and sleep in healthy males and females.," Sleep, vol. 22, no. 8, pp. 1067–71, Dec. 1999.

S. Folkard, J. Arendt, and M. Clark, "Can melatonin improve shift workers' tolerance of the night shift? Some preliminary findings.," *Chronobiol. Int.*, vol. 10, no. 5, pp. 315–20, Oct. 1993.

D. E. Ford and D. B. Kamerow, "Epidemiologic study of sleep disturbances and psychiatric disorders. An opportunity for prevention?," *JAMA*, vol. 262, no. 11, pp. 1479–84, Sep. 1989.

J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study.," BMJ, vol. 337, p. a2338, Jan. 2008.

D. C. Fowles, M. J. Christie, R. Edelberg, W. W. Grings, D. T. Lykken, and P. H. Venables, "Committee report. Publication recommendations for electrodermal measurements.," Psychophysiology, vol. 18, no. 3, pp. 232–9, May 1981.

E. Frank, H. A. Swartz, and E. Boland, "Interpersonal and social rhythm therapy: an intervention addressing rhythm dysregulation in bipolar disorder.," *Dialogues Clin. Neurosci.*, vol. 9, no. 3, pp. 325–32, Jan. 2007.

E. Freixa i Baqué, "Reliability of spontaneous electrodermal activity in humans as a function of sleep stages.," Biol. Psychol., vol. 17, no. 2–3, pp. 137–43, 1983a.

E. Freixa i Baqué, B. Chevalier, J. C. Grubar, C. Lambert, A. Lancry, P. Leconte, H. Mériaux, and F. Spreux, "Spontaneous electrodermal activity during sleep in man: an intranight study.," Sleep, vol. 6, no. 1, pp. 77–81, Jan. 1983b.

N. H. Frijda, "Mood," in The Oxford Companion to Emotion and the Affective Sciences, Oxford University Press, pp. 258–259, 2009.

M. M. Garrison, K. Liekweg, and D. A. Christakis, "Media use and child sleep: the impact of content, timing, and environment.," Pediatrics, vol. 128, no. 1, pp. 29–35, Jul. 2011.

D. Goldberg, K. Bridges, P. Duncan-Jones, and D. Grayson, "Detecting anxiety and depression in general medical settings.," BMJ, vol. 297, no. 6653, pp. 897–899, 1988.

A. Grünerbl, A. Muaremi, V. Osmani, G. Bahle, S. Ohler, G. Tröster, O. Mayora, C. Haring, and P. Lukowicz, "Smartphone-based recognition of states and state changes in bipolar disorder patients.," IEEE J. Biomed. Heal. informatics, vol. 19, no. 1, pp. 140–8, 2015.

N. K. Gupta, W. H. Mueller, W. Chan, and J. C. Meininger, "Is obesity associated with poor sleep quality in adolescents?," Am. J. Hum. Biol., vol. 14, no. 6, pp. 762–768, 2002.

F. Harb, M. P. Hidalgo, and B. Martau, "Lack of exposure to natural light in the workspace is associated with physiological, sleep and depressive symptoms.," Chronobiol. Int., vol. 32, no. 3, pp. 368–75, 2015.

J. A. Healey and R. W. Picard, "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors," IEEE Trans. Intell. Transp. Syst., vol. 6, no. 2, pp. 156–166, Jun. 2005.

J. Hernandez, R. R. Morris, and R. W. Picard, "Call Center Stress Recognition with Person-Specific Models," in *Affective Computing and Intelligent Interaction*, vol. 6974, pp. 125–134, 2011.

J. Hernandez, M. (Ehsan) Hoque, W. Drevo, and R. W. Picard, "Mood meter," in Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12, p. 301, 2012.

T. H. Holmes and R. H. Rahe, "The social readjustment rating scale," J. Psychosom. Res., vol. 11, no. 2, pp. 213–218, Aug. 1967.

T. Hori, "Electrodermal and electro-oculographic activity in a hypnagogic state.," Psychophysiology, vol. 19, no. 6, pp. 668–72, Nov. 1982.

T. Hori, A. Miyasita, and Y. Niimi, "Skin potential activities and their regional differences during normal sleep in humans.," Jpn. J. Physiol., vol. 20, no. 6, pp. 657–71, Dec. 1970.

J. Horne and O. Östberg, "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms," *International Journal of Chronobiology,* no.4, pp. 97-100, 1976.

S. M. Iacus, G. King, and G. Porro, "Cem: Software for Coarsened Exact Matching." Journal of Statistical Software 30. 2009.

Y. Janjhua and Chandrakanta, "Behavior of Personality Type Toward Stress and Job Performance: A Study of Healthcare Professionals," J. Fam. Med. Prim. Care, vol. 1, no. 2, p. 109, 2012.

N. Jaques, S. Taylor, A. Azaria, A. Ghandeharioun, A. Sano, R. Picard, "Predicting students' happiness from physiology, phone, mobility, and behavioral data" In Proc. Affective Computing and Intelligent Interaction (ACII), Xi'an, China, September 2015.

G. Jean-Louis, D. F. Kripke, R. J. Cole, J. D. Assmus, and R. D. Langer, "Sleep detection with an accelerometer actigraph: Comparisons with polysomnography," *Physiol. Behav.*, vol. 72, no. 1–2, pp. 21–28, 2001.

C. Jenaro, N. Flores, M. Gómez-Vela, F. González-Gil, and C. Caballo, "Problematic internet and cell-phone use: Psychological, behavioral, and health correlates," Addict. Res. Theory, vol. 15, no. 3, pp. 309–320, Jan. 2007.

O. P. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," In L. A. Pervin, & O. P. John (Eds.), Handbook of personality: Theory and research (2nd ed., pp. 102-138). New York: Guilford, 1999.

M. W. Johns, B. A. Cornell, and J. P. Masterton, "Monitoring sleep of hospital patients by measurement of electrical resistance of skin.," J. Appl. Physiol., vol. 27, no. 6, pp. 898–901, Dec. 1969.

L. C. Johnson and A. Lubin, "Spontaneous electrodermal activity during waking and sleeping.," Psychophysiology, vol. 3, no. 1, pp. 8–17, Jul. 1966.

A. Karni, D. Tanne, B. S. Rubenstein, J. J. Askenasy, and D. Sagi, "Dependence on REM sleep of overnight improvement of a perceptual skill.," Science, vol. 265, no. 5172, pp. 679–82, Jul. 1994.

M. Kay, E. K. Choe, J. Shepherd, B. Greenstein, N. Watson, S. Consolvo, and J. A. Kientz, "Lullaby," in Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12, p. 226, 2012

S. B. S. Khalsa, M. E. Jewett, C. Cajochen, and C. A. Czeisler, "A Phase Response Curve to Single Bright Light Pulses in Human Subjects," *J. Physiol.*, vol. 549, no. 3, pp. 945–952, Jun. 2003.

W. D. S. Killgore, J. M. Richards, D. B. Killgore, G. H. Kamimori, and T. J. Balkin, "The trait of Introversion-Extraversion predicts vulnerability to sleep deprivation.," J. Sleep Res., vol. 16, no. 4, pp. 354–63, Dec. 2007.

R. Kobayashi, Y. Koike, M. Hirayama, H. Ito, and G. Sobue, "Skin sympathetic nerve function during sleep—a study with effector responses," *Auton. Neurosci.*, vol. 103, no. 1–2, pp. 121–126, Jan. 2003.

V. Kolodyazhniy, J. Späti, S. Frey, T. Götz, A. Wirz-Justice, K. Kräuchi, C. Cajochen, and F. H. Wilhelm, "An Improved Method for Estimating Human Circadian Phase Derived From Multichannel Ambulatory Monitoring and Artificial Neural Networks," Chronobiology International, vol. 29, no. 8. pp. 1078–1097, 2012.

A. J. Koumans, B. Tursky, and P. Solomon, "Electrodermal levels and fluctuations during normal sleep.," Psychophysiology, vol. 5, no. 3, pp. 300–6, Nov. 1968.

K. Kräuchi, "How is the circadian rhythm of core body temperature regulated?," *Clin. Auton. Res.*, vol. 12, no. 3, pp. 147–149, Jun. 2002.

K. Kräuchi, C. Cajochen, and A. Wirz-Justice, "Waking up properly: is there a role of thermoregulation in sleep inertia?," J. Sleep Res., vol. 13, no. 2, pp. 121–7, Jun. 2004.

S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," Biological Psychology, vol. 84, no. 3. pp. 394–421, 2010.

K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9," J. Gen. Intern. Med., vol. 16, no. 9, pp. 606–613, Sep. 2001.

M. Kusserow, O. Amft, and G. Tröster, "Monitoring Stress Arousal in the Wild.," IEEE Pervasive Comput., vol. 12, no. 2, pp. 28–37, 2013.

L. C. Lack, M. Gradisar, E. J. W. Van Someren, H. R. Wright, and K. Lushington, "The relationship between insomnia and body temperatures," Sleep Medicine Reviews, vol. 12, no. 4. pp. 307–317, 2008.

T. Ledowski, J. Bromilow, J. Wu, M. J. Paech, H. Storm, and S. A. Schug, "The assessment of postoperative pain by monitoring skin conductance: Results of a prospective study," Anaesthesia, vol. 62, no. 10, pp. 989–993, 2007.

B. K. Lester, N. R. Burch, and R. C. Dossett, "Nocturnal EEG-GSR profiles: the influence of presleep states.," Psychophysiology, vol. 3, no. 3, pp. 238–48, Jan. 1967.

A. Lewicke, E. Sazonov, M. J. Corwin, M. Neuman, and S. Schuckers, "Sleep versus wake classification from heart rate variability using computational intelligence: consideration of rejection in classification models.," IEEE Trans. Biomed. Eng., vol. 55, no. 1, pp. 108–118, 2008.

A. J. Lewy, "Melatonin and Human Chronobiology," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 72, no. 1, pp. 623–636, Jan. 2007.

R. Liguori, V. Donadio, E. Foschini, V. Di Stasi, G. Plazzi, E. Lugaresi, and P. Montagna, "Sleep stage-related changes in sympathetic sudomotor and vasomotor skin responses in man," *Clin. Neurophysiol.*, vol. 111, no. 3, pp. 434–439, Mar. 2000.

R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "MoodScope," in Proceeding of the 11th annual international conference on Mobile systems, applications, and services - MobiSys '13, p. 389, 2013.

X. Long, P. Fonseca, J. Foussier, R. Haakma, and R. M. Aarts, "Sleep and wake classification with actigraphy and respiratory effort using dynamic warping," IEEE J. Biomed. Heal. Informatics, vol. 18, no. 4, pp. 1272–1284, 2014.

H. G. Lund, B. D. Reider, A. B. Whiting, and J. R. Prichard, "Sleep Patterns and Predictors of Disturbed Sleep in a Large Population of College Students," J. Adolesc. Heal., vol. 46, no. 2, pp. 124–132, 2010.

G. MacKerron, and S. Mourato, Mappiness. http://www.mappiness.org.uk, 2012

P. C. Mackinnon, "Variations in the number of active palmar digital sweat glands during the human menstrual cycle.," *J. Obstet. Gynaecol. Br. Emp.*, vol. 61, no. 3, pp. 390–3, Jun. 1954.

A. Malhotra and J. Loscalzo, "Sleep and Cardiovascular Disease: An Overview," Prog. Cardiovasc. Dis., vol. 51, no. 4, pp. 279–284, 2009.

C. A. Mangina and J. H. Beuzeron-Mangina, "Direct electrical stimulation of specific human brain structures and bilateral electrodermal activity," *Int. J. Psychophysiol.*, vol. 22, no. 1–2, pp. 1–8, Apr. 1996.

E. Marcelli, Holmes L, Troncosco M, Granberry P, and O. Buxton, "Permanently temporary: The health and socioeconomic integration of Dominicans in metropolitan. Boston," San Diego, California: Center for Behavioral and Community Health Studies. San Diego State University, 2009.

L. L. Marshall, A. Allison, D. Nykamp, and S. Lanke, "Perceived stress and quality of life among doctor of pharmacy students.," *Am. J. Pharm. Educ.*, vol. 72, no. 6, p. 137, 2008.

J. L. Martin and A. D. Hakim, "Wrist actigraphy," Chest, vol. 139, no. 6. pp. 1514–1527, 2011.

T. R. Martin, G. L. Flett, P. L. Hewitt, L. Krames, and G. Szanto, "Personality correlates of depression and health symptoms: A test of a self-regulation model," J. Res. Pers., vol. 31, pp. 264–277, 1996.

A. Martinez-Nicolas, E. Ortiz-Tudela, M. A. Rol, and J. A. Madrid, "Uncovering Different Masking Factors on Wrist Skin Temperature Rhythm in Free-Living Subjects," PLoS One, vol. 8, no. 4, 2013.

Massachusetts Institute of Technology, "The Healthy Minds Study", http://chancellor.mit.edu/sites/default/files/pdf/HMS_MIT_2015_Results.pdf, 2015

P. Maquet, P. Peigneux, S. Laureys, M. Boly, T. Dang-Vu, M. Desseilles, and A. Cleeremans, "Memory processing during human sleep as assessed by functional neuroimaging.," *Rev. Neurol. (Paris).*, vol. 159, no. 11 Suppl, pp. 6S27–9, Nov. 2003.

S. C. Mednick, N. A. Christakis, and J. H. Fowler, "The spread of sleep loss influences drug use in adolescent social networks.," *PLoS One*, vol. 5, no. 3, p. e9775, Jan. 2010.

E. J. Mezick, K. A. Matthews, M. Hall, T. W. Kamarck, D. J. Buysse, J. F. Owens, and S. E. Reis, "Intra-individual variability in sleep duration and fragmentation: Associations with stress," *Psychoneuroendocrinology*, vol. 34, no. 9, pp. 1346–1354, Oct. 2009.

D. G. McDonald, H. D. Shallenberger, R. L. Koresko, and B. G. Kinzy, "Studies of spontaneous electrodermal responses in sleep.," Psychophysiology, vol. 13, no. 2, pp. 128–34, Mar. 1976.

D. McNair, M. Lorr, and L. Droppleman, "Profile of Mood States (POMS)." 1989.

A. L. D. Medeiros, D. B. F. Mendes, P. F. Lima, and J. F. Araujo, "The Relationships between Sleep-Wake Cycle and Academic Performance in Medical Students," Biol. Rhythm Res., vol. 32, no. 2, pp. 263–270, 2003.

J.-K. Min, A. Doryab, J. Wiese, S. Amini, J. Zimmerman, and J. I. Hong, "Toss 'n' turn," in Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14, 2014, pp. 477–486.

F. Mokhayeri, M.-R. Akbarzadeh-T, and S. Toosizadeh, "Mental stress detection using physiological signals based on soft computing techniques," in 2011 18th Iranian Conference of Biomedical Engineering (ICBME), 2011, pp. 232–237.

W. Morris, "A functional analysis of the role of mood in affective systems," in Review of personality and social psychology, N. Park, Ed. 1992, pp. 256–293.

S. T. Moturu, I. Khayal, N. Aharony, W. Pan, and A. (Sandy) Pentland, "Using Social Sensing to Understand the Links between Sleep, Mood, and Sociability," in 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing, pp. 208–214, 2011.

A. Muaremi, B. Arnrich, and G. Tröster, "Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep," Bionanoscience, vol. 3, no. 2, pp. 172–183, 2013.

A. Muaremi, A. Bexheti, F. Gravenhorst, B. Arnrich, and G. Tröster, "Monitoring the Impact of Stress on the Sleep Patterns of Pilgrims using Wearable Sensors," in IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 2014.

R. T. Nadler, R. Rabi, and J. P. Minda, "Better mood and better performance. Learning rule-described categories is enhanced by positive mood.," Psychol. Sci. a J. Am. Psychol. Soc. / APS, vol. 21, no. 12, pp. 1770–1776, 2010.

National Institute of Mental Health, Fact Sheet on Stress, http://www.nimh.nih.gov/health/publications/stress/index.shtml, 2015

E. E. Noftle and R. W. Robins, "Personality predictors of academic outcomes: big five correlates of GPA and SAT scores.," J. Pers. Soc. Psychol., vol. 93, pp. 116–130, 2007.

H. Oginska and J. Pokorski, "Fatigue and mood correlates of sleep length in three age-social groups: School children, students, and employees.," Chronobiol. Int., vol. 23, no. 6, pp. 1317–1328, 2006.

A. Ohman, "Electrodermal activity and vulnerability to schizophrenia: A review," Biol. Psychol., vol. 12, no. 2–3, pp. 87–145, 1981.

M. N. Olpin," Perceived Stress Levels and Sources of Stress Among College Students: Methods, Frequency, and Effectiveness of Managing Stress by College Students", Southern Illinois University at Carbondale, 1996

M. Ç. Örücü and A. Demir, "Psychometric evaluation of perceived stress scale for Turkish university students," *Stress Heal.*, vol. 25, no. 1, pp. 103–109, Feb. 2009.

E. J. Paavonen, M. Pennonen, M. Roine, S. Valkonen, and A. R. Lahikainen, "TV exposure associated with sleep disturbances in 5- to 6-year-old children.," *J. Sleep Res.*, vol. 15, no. 2, pp. 154–61, Jun. 2006.

K. Petrie, A. G. Dawson, L. Thompson, and R. Brook, "A double-blind trial of melatonin as a treatment for jet lag in international cabin crew.," *Biol. Psychiatry*, vol. 33, no. 7, pp. 526–30, Apr. 1993.

H. Piosczyk, N. Landmann, J. Holz, B. Feige, D. Riemann, C. Nissen, and U. Voderholzer, "Prolonged sleep under Stone Age conditions.," *J. Clin. Sleep Med.*, vol. 10, no. 7, pp. 719–22, Jul. 2014.

M. Z. Poh, N. C. Swenson, and R. W. Picard, "A wearable sensor for unobtrusive, long-term assessment of electrodermal activity.," IEEE Trans. Biomed. Eng., vol. 57, no. 5, pp. 1243–52, May 2010.

M. Z. Poh, T. Loddenkemper, C. Reinsberger, N. C. Swenson, S. Goyal, M. C. Sabtala, J. R. Madsen, and R. W. Picard, "Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor," Epilepsia, vol. 53, no. 5, 2012.

C. P. Pollak, W. W. Tryon, H. Nagaraja, and R. Dzwonczyk, "How accurately does wrist actigraphy identify the states of sleep and wakefulness?," Sleep, vol. 24, no. 8, pp. 957–65, Dec. 2001.

A. E. Poropat, "A meta-analysis of the five-factor model of personality and academic performance.," Psychol. Bull., vol. 135, pp. 322–338, 2009.

M. R. Ralph, R. G. Foster, F. C. Davis, and M. Menaker, "Transplanted suprachiasmatic nucleus determines circadian period.," Science, vol. 247, no. 4945, pp. 975–978, 1990.

A.Rechtschaffen and A. Kales, "A Manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects," Brain Information Service/Brain Research Institute, University of California; 1968.

J. A. Roberts, L. H. P. Yaya, and C. Manolis, "The invisible addiction: Cell-phone activities and addiction among male and female college students," J. Behav. Addict., vol. 3, no. 4, pp. 254–265, 2014.

I. M. Rosen, P. A. Gimotty, J. A. Shea, and L. M. Bellini, "Evolution of sleep quantity, sleep deprivation, mood disturbances, empathy, and burnout among interns.," *Acad. Med.*, vol. 81, no. 1, pp. 82–5, Jan. 2006.

J. N. Rosenquist, J. H. Fowler, and N. A. Christakis, "Social network determinants of depression.," *Mol. Psychiatry*, vol. 16, no. 3, pp. 273–81, Mar. 2011.

A. Sadeh, K. M. Sharkey, and M. A. Carskadon, "Activity-based sleep-wake identification: an empirical test of methodological issues.," Sleep, vol. 17, no. 3, pp. 201–7, Apr. 1994.

J. C. Sagot, C. Amoros, V. Candas, and J. P. Libert, "Sweating responses and body temperatures during nocturnal sleep in humans.," *Am. J. Physiol.*, vol. 252, no. 3 Pt 2, pp. R462–70, Mar. 1987.

B. Sander, J. Markvart, L. Kessel, A. Argyraki, and K. Johnsen, "Can sleep quality and wellbeing be improved by changing the indoor lighting in the homes of healthy, elderly citizens?," Chronobiol. Int., pp. 1–12, 2015.

A.Sano and R. W. Picard, "Toward a taxonomy of autonomic sleep patterns with electrodermal activity.," Conf. Proc. IEEE Eng. Med. Biol. Soc., vol. 2011, pp. 777–80, Jan. 2011.

A. Sano and R. W. Picard, "Recognition of sleep dependent memory consolidation with multi-modal sensor data," in *2013 IEEE International Conference on Body Sensor Networks*, pp. 1–4, 2013a.

A. Sano and R. W. Picard, "Stress Recognition Using Wearable Sensors and Mobile Phones," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 671–676, 2013b.

A. Sano, R. W. Picard, and R. Stickgold, "Quantitative analysis of wrist electrodermal activity during sleep.," Int. J. Psychophysiol., vol. 94, pp. 382–9, 2014a.

A. Sano and R. W. Picard, "Comparison of sleep-wake classification using electroencephalogram and wrist-worn multi-modal sensor data," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 930–933, 2014b.

A. Sano, A. J. Phillips, A. Z. Yu, A.W. McHill, S. Taylor, N. Jaques, C.A. Czeisler, E.B Klerman, R.W. Picard, "Recognizing Academic Performance, Sleep Quality, Stress Level, and Mental Health using Personality Traits, Wearable Sensors and Mobile Phones," In the proceedings of Body Sensor Networks, Cambridge, USA, June 2015.

C. B. Saper, T. E. Scammell, and J. Lu, "Hypothalamic regulation of sleep and circadian rhythms.," Nature, vol. 437, no. 7063, pp. 1257–1263, 2005.

J. A. Sarabia, M. A. Rol, P. Mendiola, and J. A. Madrid, "Circadian rhythm of wrist temperature in normal-living subjects A candidate of new index of the circadian system.," *Physiol. Behav.*, vol. 95, no. 4, pp. 570–80, Nov. 2008.

A. M. Schell, M. E. Dawson, A. Rissling, J. Ventura, K. L. Subotnik, M. J. Gitlin, and K. H. Nuechterlein, "Electrodermal predictors of functional outcome and negative symptoms in schizophrenia," Psychophysiology, vol. 42, no. 4, pp. 483–492, 2005.

C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device.," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 2, pp. 410–7, Mar. 2010.

Y. Shiihara, A. Umezawa, Y. Sakai, N. Kamitamari, and M. Kodama, "Continuous recordings of skin conductance change during sleep.," *Psychiatry Clin. Neurosci.*, vol. 54, no. 3, pp. 268–9, Jun. 2000.

T. Shochat, O. Flint-Bretler, and O. Tzischinsky, "Sleep patterns, electronic media exposure and daytime sleep-related behaviours among Israeli adolescents," Acta Paediatr. Int. J. Paediatr., vol. 99, no. 9, pp. 1396–1400, 2010.

T. Shochat, "Impact of lifestyle and technology developments on sleep.," *Nat. Sci. Sleep*, vol. 4, pp. 19–31, Jan. 2012.

J. M. Siegel, "The REM sleep-memory consolidation hypothesis.," Science, vol. 294, no. 5544, pp. 1058–63, Nov. 2001.

A. M. Soehner, K. S. Kennedy, and T. H. Monk, "Personality correlates with sleep-wake variables.," Chronobiol. Int., vol. 24, no. 5, pp. 889–903, Jan. 2007.

C. D. Spielberger, R. L. Gorsuch, R. E. Lushene, P. R. Vagg, and G. A. Jacobs, "Manual for the State-Trait Anxiety Inventory (Form Y)," vol. IV, Jan. 1983.

R. Stickgold, D. Whidbee, B. Schirmer, V. Patel, and J. A. Hobson, "Visual discrimination task improvement: A multi-step process occurring during sleep.," J. Cogn. Neurosci., vol. 12, no. 2, pp. 246–54, Mar. 2000.

H. Storm, "Changes in skin conductance as a tool to monitor nociceptive stimulation and pain.," Curr. Opin. Anaesthesiol., vol. 21, no. 6, pp. 796–804, 2008.

N. Sugawara, T. Kikuchi, K. Yanagi, S. Yamamura, H. Morishima, H. Adachi, T. Kumano-Go, A. Mikami, Y. Sugita, and M. Takeda, "Using electronic media before sleep can curtail sleep time and result in self-perceived insufficient sleep," Sleep Biol. Rhythms, vol. 5, no. 3, pp. 204–214, Jul. 2007.

N. N. Takasu, Y. Takenaka, M. Fujiwara, and M. Toichi, "Effects of regularizing sleep-wake schedules on daytime autonomic functions and psychological states in healthy university students with irregular sleep-wake habits," Sleep Biol. Rhythms, vol. 10, no. 2, pp. 84–93, Apr. 2012.

D. J. Taylor, K. L. Lichstein, H. H. Durrence, B. W. Reidel, and A. J. Bush, "Epidemiology of insomnia, depression, and anxiety.," *Sleep*, vol. 28, no. 11, pp. 1457–64, Nov. 2005.

S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, "Automatic Identification of Artifacts in Electrodermal Activity Data" In Proc. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, August 2015.

S. Thomée, A. Härenstam, and M. Hagberg, "Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study.," BMC Public Health, vol. 11, p. 66, Jan. 2011.

J. Tilmanne, J. Urbain, M. V Kothare, A. Vande Wouwer, and S. V Kothare, "Algorithms for sleep-wake identification using actigraphy: a comparative study and new results.," J. Sleep Res., vol. 18, no. 1, pp. 85–98, Mar. 2009.

Utah Department of Health, 2001 Utah Health Status Survey, http://health.utah.gov/opha/publications/2001hss/sf12/SF12_Interpreting.pdf, 2001.

H. P. A. Van Dongen, G. Maislin, J. M. Mullington, and D. F. Dinges, "The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation.," 2003.

H. P. A. Van Dongen, "Circadian Rhythm in Sleepiness, Alertness and Performance", Chapter 38 of Principles and Practice of Sleep Medicine, Saunders, 2010.

M. van Dooren, J. J. G. G.-J. de Vries, and J. H. Janssen, "Emotional sweating across the body: comparing 16 different skin conductance measurement locations.," Physiol. Behav., vol. 106, no. 2, pp. 298–304, May 2012.

M. van Eck, H. Berkhof, N. Nicolson, and J. Sulon, "The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol.," Psychosom. Med., vol. 58, no. 5, pp. 447–58, 1996.

M. Vollrath, "Personality and stress.," Scand. J. Psychol., vol. 42, no. 4, pp. 335–47, Sep. 2001.

T. G. Vrijkotte, L. J. van Doornen, and E. J. de Geus, "Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability.," Hypertension, vol. 35, no. 4, pp. 880–6, Apr. 2000.

R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "StudentLife," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*, pp. 3–14, 2014.

N. G. Ward, H. O. Doerr, and M. C. Storrie, "Skin conductance: A potentially sensitive test for depression," Psychiatry Res., vol. 10, no. 4, pp. 295–302, 1983.

J. C. Ware, I. Karacan, P. J. Salis, J. Thornby, and M. Hirshkowitz, "Sleep-related electrodermal activity patterns in impotent patients.," Sleep, vol. 7, no. 3, pp. 247–54, Jan. 1984.

J. Ware, M. Kosinski, and S. D. Keller, "A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity.," *Med. Care*, vol. 34, no. 3, pp. 220–233, 1996.

D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales.," *J. Pers. Soc. Psychol.*, vol. 54, no. 6, pp. 1063–70, Jun. 1988.

J. B. Webster, D. F. Kripke, S. Messin, D. J. Mullaney, and G. Wyborney, "An activity-based sleep monitor system for ambulatory use.," Sleep, vol. 5, no. 4, pp. 389–399, 1982.

A. Weiss, F. Xu, A. Storfer-Isser, A. Thomas, C. E. levers-Landis, and S. Redline, "The association of sleep duration with adolescents' fat and carbohydrate consumption.," Sleep, vol. 33, no. 9, pp. 1201–1209, 2010.

A. Wirz-Justice and R. H. Van den Hoofdakker, "Sleep deprivation in depression: what do we know, where do we go?," *Biol. Psychiatry*, vol. 46, no. 4, pp. 445–453, Aug. 1999.

A. R. Wolfson and M. A. Carskadon, "Understanding adolescents' sleep patterns and school performance: A critical appraisal," Sleep Medicine Reviews, vol. 7, no. 6. pp. 491–506, 2003.

F. Zizi, G. Jean-Louis, C. D. Brown, G. Ogedegbe, C. Boutin-Foster, and S. I. McFarlane, "Sleep duration and the risk of diabetes mellitus: Epidemiologic evidence and pathophysiologic insights," Current Diabetes Reports, vol. 10, no. 1. pp. 43–47, 2010.