2020

Open Source Quantitative Stress Prediction Leveraging Wearable Sensing and Machine Learning Methods

Blake Hewgill
University of Vermont

Open Source Quantitative Stress Prediction Leveraging Wearable Sensing and Machine Learning Methods

A Thesis Presented

by

Blake Hewgill

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Electrical Engineering

October, 2020

Defense Date: August 14, 2020
Thesis Examination Committee:

Jeff Frolik, Ph.D., Advisor
Ryan S. McGinnis, Ph.D., Advisor
Nick Cheney, Ph.D., Chairperson
Ellen W. McGinnis, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of Graduate College

Abstract

The ability to monitor physiological parameters in an individual is paramount for the evaluation of physical health and the detection of many ailments. Wearable technologies are being introduced on a widening scale to address the absence of low-cost and non-invasive health monitoring as compared to medical grade equipment and technologies. By leveraging wearable technologies to supplement or replace traditional gold-standard measurement techniques, the research community can develop a deeper multifaceted understanding of the relationship between specific physiological parameters and particular health conditions. One particular research area in which wearable technologies are beginning to see application is the quantification of physical and mental stress levels in individuals through brainwave and physiological feature monitoring. At present, these methods are time consuming, invasive, expensive, or some combination of the three.

This thesis chronicles the development and application of a novel open source wearable sensing platform to the field of stress and fatigue estimation and quantization. More specifically, the garment in its current configuration monitors heart rate, blood oxygen saturation, skin temperature, respiration rate, and skin conductivity parameters to explore the relationship between these parameters and various self-reported stress measures. Utilizing machine-learning methods, subject-specific models were generated in an n=1 study which predicts the self-perceived stress level of the subject with an accuracy of between 92% and 100%. The garment was developed with a modular interface and open source code base to allow and encourage reconfiguration and customization of the sensor array for other research applications. The dataset generated in this effort spans the early stages of the COVID-19 pandemic as the subject experienced increasing levels of isolation and tracks physiological parameters across two months via daily measurements.

Citations

Material from this thesis has been published in the following form:

Hewgill, B., McGinnis, R., Frolik, J.. (2020) *A Low-Cost Modular Health Monitoring Garment.* 2020 IEEE Engineering in Medicine and Biology Conference (EMBC), Montreal, Canada, 2020.

Acknowledgements

My mother fostered in me a passion for learning and I will always be grateful to her for everything she has done for me and the opportunities she has given me. Without her support I might have mistakenly become a Business Major. My wonderful girlfriend, Maggie, also deserves a great deal of thanks for her usually unwavering support. I'm sorry our coffee table became a circuit laboratory during the COVID-19 pandemic, but I think the solder burns give the table more character. I will not be thanking the cat, who made it her daily agenda to prevent as much scholarly work as was within her power. Fortunately, she cannot read.

I must thank both of my advisors. Dr. Jeff Frolik and Dr. Ryan McGinnis have supported me for many years; without their guidance and abounding patience I would not have been able to complete this work. In both my undergraduate and graduate studies, the lessons I've taken from their courses as well as our meetings have shaped the engineer I have become and hope to continue to be. Dr. Nicholas Cheney also deserves a sizable share of gratitude for agreeing to chair my committee before having ever met me. These three gentlemen reflect a dedication to engineering, learning, and research that goes beyond that which their titles would suggest. I am a better man for having met each of them.

Dr. Ellen McGinnis also deserves enormous thanks for agreeing to join my thesis committee in the eleventh hour, despite having two small children relying on her already. Were I in her shoes, I likely would have reserved that time for catching up on sleep rather than lending assistance to a third small child.

This effort chronicled in this thesis was funded by the Vermont Space Grant Consortium under NASA Cooperative Agreement NNX15AP86H.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

From early vital sign monitoring during the Gemini space program in the 1960's [1] to current wireless consumer grade electromyography (EMG) and electrocardiography (ECG) technologies [2, 3], there persists a clear requirement for low profile, non-intrusive wearable sensing that does not interfere with users' activity or movement. Torso-based wearables which allow multiple simultaneous sensor modalities have been developed in the past [4] and several are currently available on the consumer market. These wearable devices can be utilized in a wide range of clinical and scientific applications, from activity monitoring in healthy subjects to fall-risk analysis in elderly or neurocompromised subjects.

One subject area in which wearables are only sparingly deployed is in mental stress detection. Mental stress, defined by Merriam-Webster as "the burden on one's emotional or mental well-being created by demands on one's time" [5], is an inescapable part of daily life. While all individuals experience stress, the severity, duration, and

perception of this stress can play a role in an individual's overall health. Chronic stress can lead to a multitude of other health issues including obesity, heart disease, sleep disorders, anxiety, and depression [6]. Chronic and acute stress have both been shown to increase the risk of workplace accident and can impair both short term and long term memory [7, 8, 9]. All of these side effects of stress pose a risk not only individuals, but also to the companies and corporations that employ them. Employees suffering from stress and its side effects pose a risk to themselves, their coworkers, and the assets of their employer. Stress has also been linked to workplace absenteeism [10], directly affecting the work output of the employee.

Stress is clearly an issue in workplace actors, the employees across all levels of an organizational hierarchy. However, if the stress level of workplace actors were to be monitored, interventions could be staged in order to remove those actors from actively stressful situations and mitigate risk. These benefits would extend to society as a whole, as opposed to the workplace exclusively. To date, no such monitoring system exists.

1.2 Problem Statement

According to the American Institute of Stress, the direct cost of mental stress to the American economy is approximately $300 Billion annually and stress adversely affects eight in ten working-age adults [11]. Particularly in the current economic climate in which the national deficit is more than $25 Trillion, this cost represents an unsustainable drain on an already straining financial system. If the effects of mental stress on the economy were completely eliminated, the current national deficit could

be eliminated in 83 years without making other changes.

In order to address and mitigate the costs of stress, there must first be a reliable and reproducible method of ascertaining the level of stress being experienced. To date, few efforts have been made to quantify individual stress levels beyond self-reporting. In clinical settings, the most widely accepted stress quantization method is the Perceived Stress Scale [12], though other methods are used such as the Ardell Wellness Stress Self-Assessment, the Matheny-McCarthy Stress Coping Resources Inventory, and the Symptoms of Stress Frequency self evaluation [13]. While these measures offer some insight to personal stress levels, they remain an extremely subjective manner of evaluation. By developing models which have the ability to quantize stress levels on a sliding scale based on empirical measurement alongside open source hardware capable of providing these measures, individual stress levels could be monitored in near-real time. With an up-to-date characterization of current and historical stress levels, intervention could be strategically planned in populations before heightened stress levels interfere negatively with quality of work. By providing insight as to when intervention may be necessary, workers in mission-critical environments could be relieved before their actions begin to pose a risk to personnel or property.

This thesis presents a wearable device developed with the intended application of stress evaluation, but with a modular platform intended to allow the device to be utilized in a broad range of other clinical research applications such as ailment detection and health metric tracking. Furthermore, this work presents models developed utilizing machine learning techniques to predict the level of stress in a subject based on physiological parameters detected using the aforementioned wearable device.

1.3 Contributions

The work proposed in this thesis consists of a novel open source wearable sensing device and its application to mental stress estimation. There are three main contributions from this effort.

1. *An Open Source and Modular Health Monitoring Hardware Solution*

   The measurement system proposed herein represents a novel approach to wearable sensing by introducing an open source platform to encourage further development unaffected by fiscal motivation. This system is modular and scalable by nature of the custom board layout. The development of this system is detailed in Chapter 4.

2. *A Public Dataset of Self-Estimated Stress Level and Measured Biosignals*

   The effort chronicled in this thesis resulted in a dataset consisting of 80 data collections of 11 minutes each in length. Each of these collections contains both skin and ambient temperature measurements, electrocardiogram, photoplethysmography, galvanic skin response, and respiration measurements. These data collections were compiled over a period of three months in a consistent environment. The data collection is detailed in Chapter 4.5.

3. *Machine Learning Models for Prediction of Self-Perceived Mental Stress Levels*

   Utilizing the dataset described in *Contribution 2*, machine learning methods are applied to develop quantitative predictive models of self-perceived mental stress in a subject. These models attempt to predict the resultant scores of four independent clinically used self evaluations of mental stress. The creation and

training of these models is described in Chapter 6.

While there is significant prior work in wearable sensing, there has not been an open source multimodal platform to date. There is also a distinct lack of publicly accessible long term data linking multiple physiological parameters to self perceived stress levels. Predictive models of mental stress do not appear to have been published in prior work. The three contributions presented in this thesis are novel, and will hopefully inspire further work into the long term characterization and prediction of mental stress.

1.4 Thesis Outline

This thesis follows the following template: Chapter 2 describes varying methods and approaches used in the estimation of stress and fatigue, as well as a brief history of stress as a health ailment. In Chapter 3, wearable sensing is introduced and several specific use cases for the technology are described. Several specific commercial wearable devices are also discussed. Chapter 4 features an in-depth examination of the open source hardware platform developed in this work. This section presents both the hardware and software structures of the device, as well as a brief cost overview. In Chapter 5, the methodology for data collection and feature extraction is laid out. The specifics of the machine learning approach are described in Chapter 6. Finally, Chapter 7 serves to reflect on the work contained herein by proposing future work and closing remarks.

Chapter 2

Stress and Fatigue Estimation

2.1 Introduction

The stress response is an inescapable and necessary part of daily life. It is an evolutionary trait that has manifested as a reaction to uncertainty and danger, and grew from the classical concept of the 'Fight-or-Flight' response. It is important to note that not all stress is negative; tempered acute stress responses can encourage motivation and focus, and improves physical performance in some populations. The negative mental and physical effects typically associated with stress are generally the outcome of either chronic stress situations or events of excessive acute stress. Every individual has a different tolerance level for these types of stress, and the point at which stressful situations begin to cause ill effects will vary from person to person [15].

Moreover, the experience of stress is not a permanent one. There are many methods available to cope with stress and, over time, lower the level of chronic stress being experienced, even without directly addressing the exposure to triggering stimuli but by instead altering the psychological perception of those stimuli and the reaction to

them [16]. In this chapter, we will examine the history of stress identification and research as well as the various practical methods of quantifying the extent of stress being experienced by an individual.

2.2 History of Stress Estimation

The concept of stress as it is understood today is a recent contribution to the field of science and academic research. During his time as a medical student from 1925 to 1929, Hungarian physician Hans Selye noted that patients suffering from from widely varying afflictions often exhibited identical symptoms. After graduating medical school and earning a doctorate in chemistry, Selye began work with the McGill University Department of Biochemistry. Researching with rats, he once again noticed that rats subjected to physical injury, uncomfortable cold, and intoxication would respond in the same manner for the duration of exposure to these stimuli. In 1936, Selye published his theory on General Adaptation Syndrome (GAS). In this theory, he posited that these responses could be characterized by three successive stages: alarm, resistance, and exhaustion. In this publication, the GAS response was repeatedly claimed to be the direct response to 'stress,' the first time the word had been used in this sense (contributing to Selye's moniker, 'The Father of Stress'). During the Alarm stage, the body distributed increased levels of oxygen and glucose to the organs in order to increase the ability of the subject to respond immediately to threats, and blood pressure also increased. This condition lasted for up to 48 hours, at which point the extent of glucose and oxygen distribution gradually returned to pre-stimulus levels. In this second stage, Resistance, the body was able to adapt to

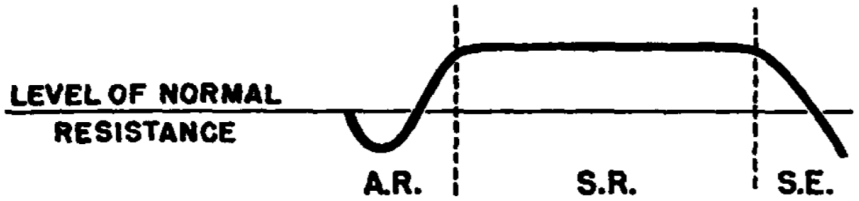

Figure 2.1: Three-phase stress response pver time according to the General Adaptation Syndrome. A.R.: Alarm Reaction, S.R.: Resistance Stage, S.E.: State of Exhaustion. Reproduced from [17]

the external threat and blood pressure returned to normal. While the rats were able to endure exposure to a specific unpleasant or harmful stimulus, their susceptibility to other harmful substances or events increased. The adaptation was, however, not permanent. After a period of resistance which varied in length by subject the rats entered the Exhaustion phase, where resistance to the original threat decreased and the animals eventually died. If the stimulus was removed from the rats, they entered an alternate and self explanatory third stage: Recovery. *As an editorial side note, these experiments would be considered downright barbaric and unethical by modern standards; Selye would regularly leave caged rats outside for extended periods in Montreal winters or practice unnecessary surgeries. At the time these were considered acceptable conditions for animal experimentation.*

A depiction of the GAS response can be seen in Figure 2.1. The academic community spurned the GAS theory, as the concepts were vague and unmeasurable. Selye took his theory to the general public, and later went on to document that stress reactions in humans are identical in the case of receiving good news or bad news. He called stress reactions to negative stimuli 'distress,' and reactions to positive stimuli 'eustress,' titles which are still utilized in psychological fields [17].

Several decades later in the 1960's, psychologists began to adopt Selye's theories. Extensive research was undertaken to examine the connection between stress and disease. These studies determined links between stress and the neuroendocrine system, the system that releases hormones into the blood based on messages from the brain [18]. This system forms the bridge between psychological stress and its physical manifestations, which will be examined further in Section 2.4. In 1980, Post Traumatic Stress Disorder (PTSD) was added to the standardized Diagnostical and Statistical Manual of Mental Disorders, characterized by an ongoing and severe reaction to a psychological stressor. By the 1990's the terms 'stress' and 'stressor' had entered the American vernacular, and attention was turned to addressing stress in specific areas of life such as workplace stress and the development of stress coping mechanisms [19].

The 1980's also saw the development of the Transactional Model of stress and coping, developed by Richard Lazarus and Susan Folkman. This five stage model is still considered relevant and applicable today, and contends that not all stressors contribute to stress levels. In the first model stage after stimulus exposure, an individual evaluates the stressor during a Primary Appraisal (Stage 1). If the stressor is deemed to be positive or irrelevant, it is disregarded and does not compound stress, ending the reaction. If instead the stressor is deemed to be a negative or dangerous influence, it undergoes Second Appraisal (Stage 2) where the individual analyzes their personal coping resources. If resources are sufficient, stress is not compounded, but if an individual does not have adequate resources then the third stage of the model is reached: Stress (Stage 3). The Stress stage invariably leads to the Coping stage (Stage 4), wherein the individual may either change the stressful situation in a problem-focused approach or change his or her perception of the situation via an

emotion-focused approach. The length of time spent before transitioning to stage four is dependent on the individual as well as the specific situation. Finally in the fifth stage, Reappraisal (Stage 5), the individual may garner knowledge from the first four stages and find themselves better equipped to respond to the same stressor in the future. The greatest difference between the Transactional model and the GAS model is the concept that humans can acquire and develop coping skills which allow varying degrees of freedom when responding to a stressor [20].

Self-coping methods for stress de-escalation have existed to a certain extent for decades, likely long before stress was formally defined as a concept. The first known usage of the term 'coping' occurred in the 1950's in a medical context, when psychiatric patients were encouraged to engage in self-care practices such as shaving and exercising in order to maintain a sense of self worth and independence. In the 1960's, emergency responders were first advised to engage in self-care acts to balance physical, emotional, and spiritual needs. Addressing these three focus areas might involve improvement of sleep hygiene, personal journaling, and meditation practices respectively [21].

Racial injustice also played an important role in the expansion of self-care practices. In many communities, African American citizens did not have access to the quality of healthcare enjoyed by their white counterparts. To address this shortcoming, these communities encouraged self-care as preventative medicine and encouraged wholistic health practices. The Black Panther Party recognized this lack of healthcare equality and incorporated free healthcare into their platform base. In her 1988 book *A Burst of Light*, African American activist Audre Lorde wrote, 'Caring for myself is not self-indulgence, it is self-preservation, and that is an act of political warfare,'

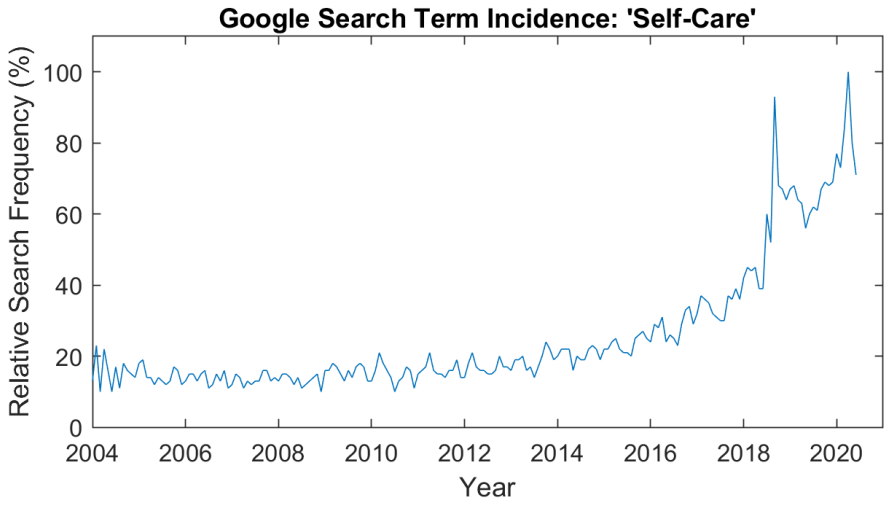

Figure 2.2: Google search popularity of the term 'Self-Care' since January 2004

reaffirming the notion to her readers and community that self-care practices could double as civil disobedience [22].

Unfortunately, these stress coping methods are not reaching a broad enough audience. A 2015 study by the American Psychological Association determined that stress levels are rising across the nation, noting financial responsibilities, family responsibilities, and stressors in the workplace as leading contributing causes. Younger generations reported more extreme stress levels than more elderly generations, and a large portion of study participants indicated they did not feel they were doing enough to address their stress [23]. A brief examination of Google internet search trends from the company's inception through June, 2020 reveals that interest in self-care topics has drastically increased over the course of the last 16 years (Figure 2.2). This trend may be interpreted from two differing viewpoints; either the general public is increasingly aware of self-care methodology or the necessity for self-care has sharply increased.

2.3 Self Estimation of Mental Stress

Over the years, there have been many different methods to self-evaluate personal stress levels. Perhaps the earliest of these is the 1967 Holmes and Rahe Scale, deployed to investigate the possible relationship between stressful life events and physical illness. In this scale, subjects tally stressful life events occurring within the past year and each event is assigned a point value. If the sum of these points is greater than 300, the subject is considered at risk of developing illness. Scores between 150 and 299 were considered to indicate moderate risk of illness, while scores below 150 were considered to indicate a low risk of illness [24]. While this scale was widely used around the time of its inception, the Holmes and Raye Scale has generally been replaced clinically with more modern methodologies.

One such clinical scale of mental stress is the Cohen Perceived Stress Scale (PSS) introduced in 1983. The PSS consists of ten self-evaluation questions reflecting on the extent to which daily life has been affected by stress over the previous month. Each of these questions is answered on a scale of 0 to 4, with larger values corresponding to increasing levels of disruption. These scores are then tallied for a final score between 0 and 40. Scores from 0 to 12 are considered a 'Low Stress' outcome, 13 to 26 are considered a 'Moderate Stress' outcome, and scores of 27 to 40 are considered a 'High Stress' outcome [12]. This scale is currently the most frequently used clinical measure of mental stress.

The Ardell Wellness Stress Test (1977) is another stress index which is still widely used today. In this test, subjects are asked to reflect on 25 items by assigning values to each of them in the range of -3 to +3. Lower numbers indicate greater dissatisfaction

with the item in question while higher numbers indicate greater satisfaction. The 25 items on the assessment are aspects of daily life such as home life and career choice. Responses to each item are added together to obtain a final score between -75 and 75. These scores are then assigned to one of six categories separated by 25 point intervals, ranging from 'No coping capacity - candidate for major psychological care' to 'Fully self actualized - nearly immune from the ravages of stress.' The Ardell test is unique in that it was the first assessment to take a wholistic approach to stress, inquiring as to physical, mental, emotional, spiritual, and social aspects [26].

Another self-evaluation measure is the Depression Anxiety Stress Scale (DASS) from 1995. This evaluation method consists of 42 statements each of which is responded to with a rating between 0 and 3. If a subject assigns a 0 to an item, he or she is indicating that the statement in question does not apply, whereas an answer of 3 indicates that the statement strongly applies [25]. Of these 42 questions, 14 relate each to scales of depression, anxiety, and stress which are then broken down into a number of subscores. The depression subscales are dysphoria, devaluation of life, hopelessness, anhedonia, and social inertia. Anxiety is separated into situational anxiety, skeletal muscle effects, and autonomic arousal. The stress scale assigns subscores to difficulty relaxing, nervous arousal, irritability, and impatience.

The Stress Coping Resources Inventory (SCRI) was first published in October of 2000. In this scale there are 32 questions asking subjects how they feel about certain situations or how they believe others perceive them. Answers to these questions range in specific wording but are consistently assigned a value between 1, which indicates the minimum accordance with the question, and 4, which indicates a strong alignment with the answer that the SCRI deems the most positive. The subject is then assigned

seven scores based on their answers. The mean score from questions 1 through 7 is deemed the Wellness score. Questions 8 through 13 average to become the Thought Control score, and questions 14 through 20 indicate Active Coping skills. 21 through 26 are a measure of Social Ease, while 27 and 28 return a Tension Reduction skill. The final answers, 29 through 32, indicate the subject's Spiritual Practice score. An Overall Stress Coping score is calculated by averaging these previous six scores. Each of the final seven scores falls into one of four classifications for coping abilities. Scores less than 1.5 indicate less than average coping ability, whereas scores between 1.5 and 2.4 indicate average coping. 2.5 to 3.4 indicate an above average set of coping skills, and scores 3.5 or above are assigned the label of 'Superior Stresscoper.' By representing different aspects of the subject's coping ability, it is proposed that he or she may be able to focus specifically on improving those skills found to be underdeveloped. [27].

The Symptoms of Stress Frequency Index takes a slightly different approach to self stress evaluations in that subjects are not categorized in any way based on their final score. In this assessment subjects are presented 12 symptoms of excessive stress such as headaches, fatigue, and insomnia. A value between 1 and 7 is assigned to each symptom according to the frequency with which the subject experiences it, with 1 indicating 'Never,' and 7 indicating 'Almost all day, every day.' These scores are then added together and fall somewhere in the range between 12 and 84, with higher scores indicating a greater stress level [13].

All of these methods attempt to assess mental stress and its perceived effects on the individual. However none of these tests are capable of assessing stress levels in real time, and all require intrusive interventions into daily tasks to complete. If

mental stress could be detected through methods which do not require self-estimation, the potential exists to create a long-term characterization of a subject's mental state which could be used to monitor changes in an individual's level of stresscoping.

2.4 Quantitative Measurement

Attempts have been made in recent years to quantify mental stress and fatigue based on external physiological responses, with varying success. Some biosignals have been proven to be associated with chronic or acute stress response through signal features present in either the time domain or in the frequency domain representation of the signal. Some of these characterizing features are exclusively manifested internally to the body, such as changes in the digestive process or the concentrations of stress hormones in the bloodstream [14]. These responses are difficult to track without resorting to invasive monitoring methods such as drawing blood or inserting cameras into the body. Other responses are manifested externally and can easily be monitored with basic equipment, such as the galvanic skin response or changes in the heart beat.

2.4.1 Heart Rate Parameters

There has been substantial research into the effect of chronic and acute stress as it manifests itself in the behavior of the heart, specifically in measures of heart rate variability. Heart rate variability (HRV) is the variance in time between successive heart beats. More specifically it is the variance of the RR-interval. The RR-interval is the length of time between successive R waves of the heart beat, ocurring when ventricles depolarize. The R wave serves as a convenient fiducial mark because it

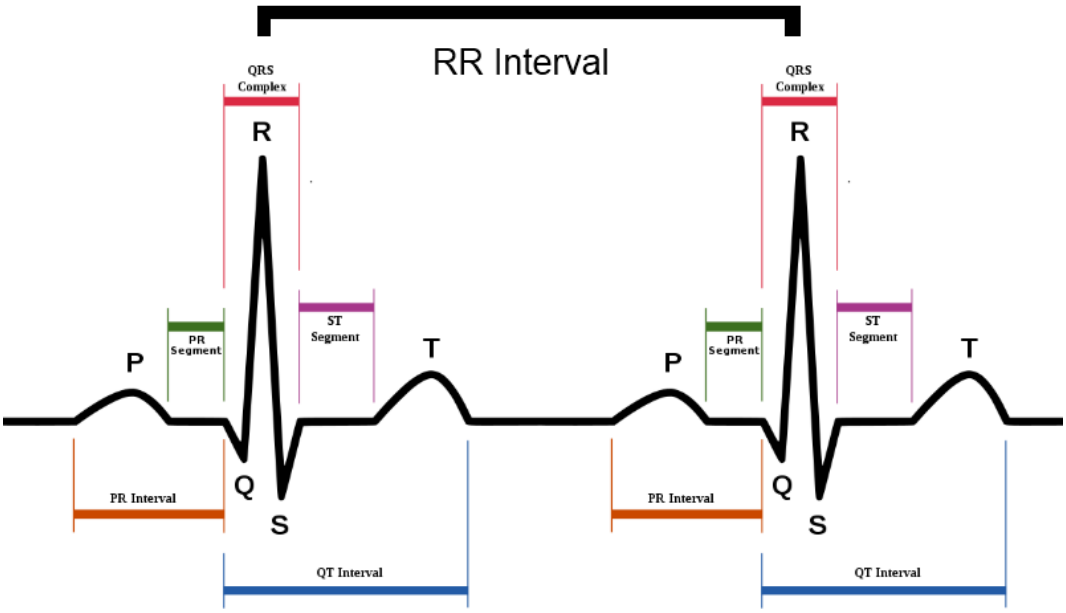

Figure 2.3: Labeled diagram of a QRS wave as it appears on an ECG signal. The interval between R waves (RR Interval) is a common metric for quantifying heart rate variability. Reproduced from [28].

produces a sharp and well defined peak in an electrocardiogram signal, as can be seen in Figure 2.3. This peak occurs as part of the QRS complex, which represents the electrical impulse of the heart as it permeates the ventricles and indicates the depolarization. Not all heart beats feature Q or S waves, but should contain an R wave. RR-intervals vary from one heartbeat to the next due to interactions between the two components of the autonomic nervous system, the sympathetic and parasympathetic nervous systems. The parasympathetic nervous system signals the heart to decrease its rate, while the sympathetic nervous system sends signals to increase the heart rate. Heart rate detection equipment varies, but electrocardiogram and photoplethysmography signals are frequently used for this purpose. Both devices operate on well documented principles and have been in use for more than 80 years [29].

Two other metrics most widely used to detect stress from heart rate are the Low Frequency to High Frequency Ratio (LF/HF) and the Root Mean Square of Successive Differences (rMSSD), both extracted from HRV data. In the case of the LF/HF ratio, the frequency domain of the HRV data is divided into regions of 'low frequency' content from 0.05-0.15 Hz and 'high frequency' content from 0.15-0.4 Hz. The LF region must be constructed from a minimum of 2 minutes of HRV data and is thought to contain information from both the sympathetic nervous system and the parasympathetic nervous system. The HF region requires at least one minute of data and is affected solely by the parasympathetic nervous system. Decreased levels of HF frequency content have been linked to stress and panic. The ratio between these two regions is thought to provide an indicator of the ratio between sympathetic and parasympathetic nervous system activation. It has been noted however that the overall length of recorded data used to calculate LF/HF does influence the calculated value, as different processes contribute to LF and HF regions at different times in the day [30].

rMSSD is traditionally calculated using 5 minutes of HRV data, though some research has investigated operating on lengths of between 10 seconds and 1 minute. Because it is a measure of variance, this metric is unaffected by trending changes in time series heart rate measurements. This is the primary metric used in time domain HRV analysis, and has been shown to strongly correlate with self-reported stress levels [31]. Other metrics positively associated with mental stress levels are mean heart rate, mean RR-interval, normalized low frequency content, and the difference between low frequency and high frequency content of HRV [32].

2.4.2 Galvanic Skin Response

Galvanic Skin Response (GSR), also known as Electrodermal Activity (EDA), is the measure of varying conductance on the skin surface. Levels of skin conductance vary according to many parameters, such as environmental humidity and barometric pressure, but also directly as a result of physiological processes. There are three processes which contribute to GSR: muscular activity, vascular dilation, and secretory changes. Sweat response is included in the category of secretory change and has been demonstrated to be associated with acute stress response, following stimulus exposure within five or fewer seconds [33]. One practical demonstration of this principle is the polygraph, or lie detector test. GSR measures were added to the standard polygraph device in 1938 on the assumption that lying under pressure would activate an acute sweat response in a subject. To this day, GSR sensor data is considered the most reliable sensor channel for lie detection during polygraph sessions [34].

There are two main methods for acquiring GSR data: exosomatic and endosomatic. Exosomatic methods apply an AC or DC current to the skin surface, and measure the voltage drop across the electrodes. Through a simple Ohm's Law calculation (V/I=R), skin conductance can be determined as the inverse of the calculated resistance. DC current methods are the simplest to deploy but risk electromagnetic interference induced by the applied current. AC current exosomatic methods avoid this limitation, but require the addition of an inverter to the equipment. In endosomatic GSR measurements, external current is not applied to the subject. Two electrodes are placed on the skin and the electrical potential between them is amplified. These systems are able to detect smaller GSR events than their exosomatic

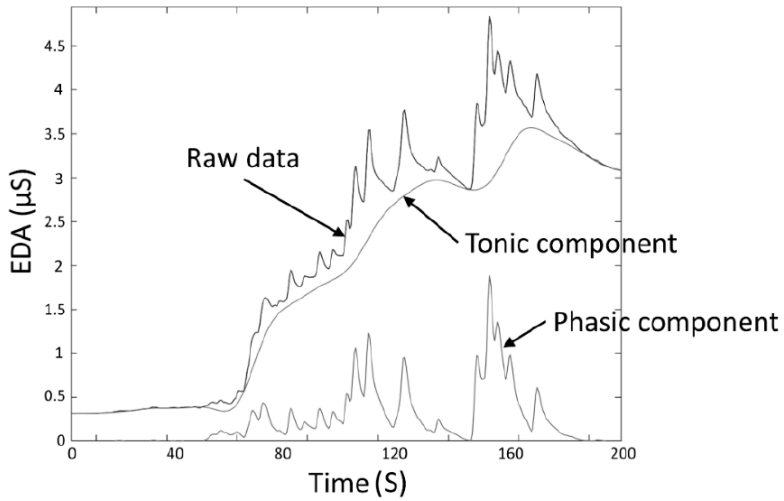

Figure 2.4: Raw GSR data deconstructed into tonic and phasic components, from [36]

counterparts, but require more complex circuitry [35].

Raw GSR data can be deconstructed into two main components: the phasic and the tonic. Phasic GSR components adapt rapidly to stimulus exposure, resulting in a well defined peak in the data generally taking place over the course of 0.5 to 5 seconds. These events have been linked strongly with acute stress responses. In contrast, the tonic component produces a slower trend in the GSR data which lasts for the duration of the presence of the stressor. These events can generally be extracted from raw data simply by taking the mean of the raw data, resulting in a computationally inexpensive measure of GSR. More advanced detrending approaches may provide a more complete characterization of chronic stress [36]. A visualization of the two components can be seen in Figure 2.4.

2.4.3 Temperature

Body temperature has also been shown to fluctuate with mental stress. Specifically, the temperature of the skin surface is known to fluctuate with vascular constriction by nature of reduced blood flow near the skin [37]. This vasular constriction is caused by the activation of the sympathetic nervous system. Due to this relationship, it is believed that the experience of either acute or chronic stressors will lower the skin temperature of a subject for the duration of exposure to those stressors. The extent of skin temperature drift as a result of stress has been attributed to the level of severity of the stressor, from barely perceptable to more than one degree Celsius. The intensity of the stress reaction has also shown to delay the rate at which the body returns to pre-stimulus temperature once the stressor is eventually removed, with greater stress responses resulting in a longer refractory period [38].

2.5 Conclusion

In this chapter, we explored the history of mental stress awareness and measurement. Various methods of self-estimation were identified and elaborated upon, as well as the physiological responses that accompany different levels of stress.

As our planet becomes more connected and world news can be widely distributed within seconds of being finalized and published, the number of stressful stimuli we are exposed to on a daily basis is only likely to increase. Fortunately, the stigma surrounding mental health treatment and discussion is rapidly dwindling, and methods of stress-coping and self care are being more widely validated and disseminated.

Stress management and self-care training, if widely taught, could reduce the strain on a burdened healthcare system and encourage a more highly motivated workforce.

While the availability of stress management and self-care programs is beneficial to the wellbeing of a population, progress cannot be made without creating methods to identify those in the population who are experiencing unhealthy amounts of stress. Self evaluation methods are available for this purpose, but remain a subjective measure for the already inexact science of stress management. For this reason, effective quantization methods of mental stress are a necessity. By identifying physical indicators of mental stress levels in a population, intervention can be staged before these levels reach critical thresholds for longterm physical and mental health of an individual. In the next chapter, we will explore portable methods of measuring biosignals, and begin to consider how they might be applied to this field.

Chapter 3

Wearable Technology

Wearable devices have begun to permeate daily life across many cultures. As semiconductor technologies continue becoming smaller and smaller, manufacturers increasingly find themselves with the ability to integrate greater functionality into their devices. As a result, devices are constantly evolving and advancing in terms of sensor modalities and capabilities.

3.1 History

Mankind has forever sought to make technology portable, which has inevitably resulted in the advent of wearable devices. Arguably the earliest foray into the field of wearable technology ocurred when Peter Henlein began to produce tiny mechanical clocks to be worn as necklaces in 1505 [39]. This concept of portable time-telling advanced to the development of pocket watches in the late 1600's, and the eventual development of wristwatches in the late 1800's. Today the wristwatch is by far the most common development platform for consumer wearables, incorporating activity

Figure 3.1: Examples of wearable devices. Left to Right: Akouphone hearing amplification device (1900), Polar Sport Tester PE 2000 (1982), Fitbit (2013), BioStamp RC (2017).

tracking, heart rate detection, and even social media into some wrist-worn devices.

Another prime example in the early adoption and evolution of wearable devices is the hearing aid. The now ubiquitous hearing amplification devices were first introduced as early as the year 1634, in the form of a handheld trumpet-like device [40]. These devices lacked any form of electronic amplification until Miller Reese Hutchison invented the Akouphone in 1898, which would be refined and manufactured by the Siemens company from 1908 onward. One of these devices can be seen in Figure 3.1. The first *wearable* hearing aid was introduced in 1936, though it was based on vacuum tube technology and was markably larger than similar devices we consider portable today. Transistor technology advanced the minimization of hearing aids but were not reliable until Zenith Electronics introduced the Maico Transist-Ear in 1952. When microprocessors were introduced in the 1970's, onboard digital processing allowed hearing aids to improve the quality of sound delivered to the user while also decreasing the size of the enclosure. Today, many hearing amplification devices offer wireless cellphone connectivity and multiple audio profiles, to tailor the user experience based on ambient sound conditions [41].

One subcategory of wearable technology, and the one upon which this thesis is focused, is Wearable *Sensing*. Devices in this category are worn on or near the

skin, where they detect, process, and transmit signals of interest. These devices may present information to the user in real time, or may simply store data for later use. Activity trackers are a prime example of this type of device. The earliest form of digital activity tracking came in the form of the pedometer, a small device generally worn on the hip, which counts the number of steps taken and distance traveled by an individual over the course of a day. The first of these devices was designed in 1780 and was based on the mechanism of a self winding clock. This concept was reinvigorated in 1965 by the introduction of the step-counting 'manpo-kei,' which translates in English to '10,000 steps counter' [42].

In 1982, Polar released the world's first wireless heart rate monitor, the Sport Tester PE 2000, which consisted of a sensing transmitter band worn across the chest and a wrist-worn receiver with a display. These devices continued to increase in accuracy and decrease in size, and Polar remains an active manufacturer in fitness tracking today [43]. The American company Fitbit has also released a series of dedicated activity trackers, beginning in 2008 with the Fitbit Tracker. This device was the first activity tracker to feature wireless charging, and also monitored users' sleep cycles based on periods of restlessness. In 2013 the company introduced the waterproof Fitbit Flex, the wrist worn device intended for 24/7 wear [44]. Modern devices may incorporate accelerometers, gyroscopes, and sometimes pulse oximeters and altimeters, however their intended purpose of monitoring and reporting health parameters to the user remains unchanged.

In 2006, Nokia became the first manufacturer to produce a cellular phone which included an accelerometer. This allowed the phone to act as an activity tracker without additional peripherals. Other manufacturers took notice, and today all smartphones

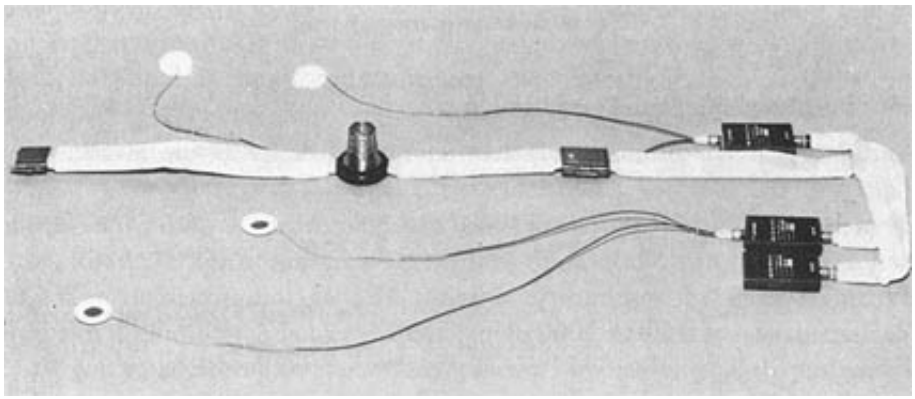

Figure 3.2: Early Biobelt astronaut bioinstrumentation unit. Two ECG leads pictured upper left, phonocardiography device pictured center, two lead impedance pneumogram pictured lower left. Reproduced from [1]

feature at the minimum a triaxial accelerometer while many also provide for optical-based methods of heart rate monitoring via photoplethysmography (PPG) sensor. Public application repositories such as Google Play or the Apple App Store allow third parties to develop and distribute software which leverages these hardware resources for a multitude of purposes including gaming, health monitoring, and activity tracking [45]. As a result, individuals with cellphone access have greater access to activity tracking than ever before.

Wearable multimodal monitoring systems are commonplace today, but symbolized a great advancement in scientific instrumentation at their introduction. Prior to the space program, instrumenting human subjects was not a concept that had been explored outside of hospital patients. No sensing package existed at the time which could monitor astronaut health to the extent desired by flight surgeons. To address this shortcoming NASA developed the Biobelt, a cotton and teflon pouch worn under astronaut flight suits which contained sensors, signal conditioning circuitry, and telemetry devices. An image of one of these Biobelts can be seen in Figure 3.2. The

unit featured a two lead ECG and phonocardiography system, as well as an impedance pneumogram to deliver heart rate and respiration rates of astronauts. During ground tests the device also contained a skin temperature sensor, but this was removed prior to any active mission usage [1].

3.2 Modern Wearables

Health monitoring applications have also inspired large scale research and development efforts over the years, to the extent that there is a specific term for portable health monitoring solutions: m-health. While the term initially applied to the broadening use of mobile phones in healthcare, it has expanded to include the use of wearable devices. These devices may utilize any number of sensor modalities in order to successfully track parameters of interest. One such example is the BioStamp RC platform produced by MC10 [46]. The BioStamp RC is an adhesive backed sensor applied directly to the skin in one of 25 validated body locations. The sensor is capable of detecting accelerometry, gyroscopic rotation, ECG, and EMG signals. Data is stored locally, and each device can operate remotely for 36 hours at a time. A bluetooth radio is used to offload data to a mobile phone or tablet, and because each BioStamp stores data rather than actively streaming, multiple units can be worn simultaneously to gather movement and muscle activation data of certain limbs or even the entire body. This platform allows researchers to methodically study subject activity and body kinematics, even if a subject is unable to be physically present in a clinician's office.

There have also been wearable sensors marketed directly to consumers and hob-

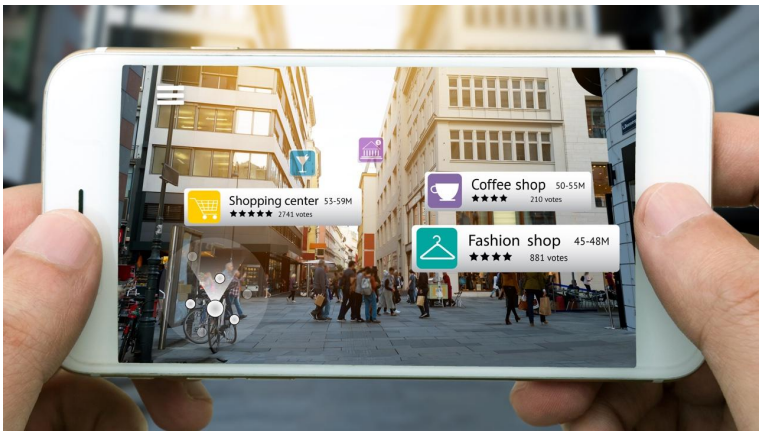

Figure 3.3: Augmented reality implementation: details about the locations in the camera's field of view are presented to the user.

byists. The MYO Armband from Thalmic Labs [47] was a publicly available EMG sensing armband which wirelessly streamed muscle activation data to a nearby computer. The user could then apply these signals at will, mapping activation patterns to desired outcomes such as control of flying drones or the articulation of a smart prosthetic device. The $200 device is no longer marketed as the company has shifted focus to producing augmented-reality smart eyeglasses. Augmented reality, differing from virtual reality in that items are overlaid on the real world, as opposed to entering a simulated world, is another novel, though somewhat controversial, application of wearable sensing. An example of how this may appear on a personal mobile device can be seen in Figure 3.3.

In 2013, Google marketed the Google Glass, a pair of augmented reality eyeglasses which were quickly removed from the market over both safety and privacy concerns [48]. Google continues to offer Glass to businesses as a workplace solution to increase worker productivity by provided hands-free access to reference materials. Thalmic Labs has determined that the market climate is now more conducive to this type of

device (likely due to the distinct lack of a camera compared to Glass), and aims to release the product at some point in 2020. Both of these devices feature screens, visible only to the individual wearing the device, which present information about objects in the user's field of view. In the same manner in which the increased hardware resources of mobile phones have allowed those devices to be used as activity trackers, the presence of a camera on the phone and increased processing power allow third party developers to introduce augmented reality applications to the masses.

Garments of clothing offer a convenient and approachable method of attaching sensors to a subject. Many companies offer training solutions to athletes in addition to the wrist-worn devices already mentioned previously. The Myontec mBody [49] and Athos Coach [50] are two such systems. Both platforms consist of electromyography sensor-laden athletic garments with wireless datalogging and exporting. In the case of the mBody, muscle activation is monitored in the hamstrings, quadriceps, and gluteals alongside inertial data. The Myontec software then generates a workout plan for the athlete to meet targeted training goals. The Athos system offers the same monitoring of lower body muscle groups and training recommendations while also allowing EMG monitoring of the pectoral, bicep, tricep, deltoid, and latissimus dorsi muscle groups.

Many studies have also investigated the presence of IMU's embedded in footwear to monitor stride characteristics and the progression of fatigue in athletes [51]. Gait, fatigue, and muscle activation analysis are all methods athletic trainers utilize to isolate areas of improvement for their athletes to focus on, as well as quantitatively track physical improvements over time.

In recent years, smart garments have increased in popularity both from an aca-

demic research perspective as well as in the field of hobbyist electronics. This is in part due to the increasing availability of conductive thread, allowing researchers and hobbyists alike to introduce tactile interfaces into a garment. An extension of this resource is the creation and refinement of textile electrodes. These electrodes generally feature a cotton pad treated with PEDOT:PSS, a conductive polymer which maintains its properties through many machine wash cycles [52]. By replacing traditional disposable electrodes with PEDOT:PSS-based textile electrodes and traditional cabling with conductive thread, wearable garments can be made which are washable, more flexible, and more comfortable than previous wearable devices.

3.3 Special Measurement Cases

Very few wearable technologies are applicable in all situations; special cases exist wherein either the hardware must be adapted or the processing flow must change in order to account for these conditions. Perhaps the most intuitive of these special cases is microgravity. When the human body enters orbit around the planet, there are fundamental differences as compared to the Earth's surface. Accelerometers cannot be normalized to a 9.8 m/s^2 gravity vector because an orbiting device is consistently in free fall. Orbital vehicles are oriented by tracking the location of stars on the horizon and comparing them to astrological maps, but humans within these vehicles (and as an extension, any wearable devices on their person) do not have access to the same field of view. It becomes very difficult as a result to determine the absolute orientation of an individual using inertial measurement sensors, though the ability to determine relative orientation after movement within a spacecraft is unimpeded.

Computer vision technologies can be used within spacecraft to identify crew members and their absolute orientation, although these systems are incredibly expensive to implement. The International Space Station features a Phoenix Technologies motion capture system for this purpose at a material cost of between $150,000 and $250,000, noninclusive of the $10,000/pound cost to elevate a payload to orbit [53].

In space, bodily fluids also tend to rise towards the head, confounding optical and surface health measurement techniques such as PPG and EMG by altering the fluid content at the measurement site [54]. This change can be accounted for via calibration of sensors compared to gold standard measurement methods, but sensors models must be calibrated each time they are used until fluid movement equalizes in the body. Unfortunately, non-surface EMG methods involve inserting a needle through the skin layer and into the muscle, an incredibly invasive method of obtaining baseline muscle activation levels.

There are two specific torso-based devices which have been used in microgravity environments. The first, the MagIC-Space, is a torso-based wearable developed specifically to monitor cardiac patterns during sleep in microgravity environments [55]. The garment consists of an instrumented vest featuring a two-lead ECG with textile electrodes, a textile plethysmograph for respiration measurement, a temperature probe, and two accelerometers for the purpose of seismocardiography measurements of the upper chest. In 2015, the MagIC-Space was used aboard the International Space Station to collect seven overnight sleep recordings during a validation trial. The limitations described above inherently do not afflict the MagIC-Space by nature of the device's intended usage; there is no attempt at estimating astronaut orientation or location within a vehicle and the device does not attempt to collect PPG or EMG

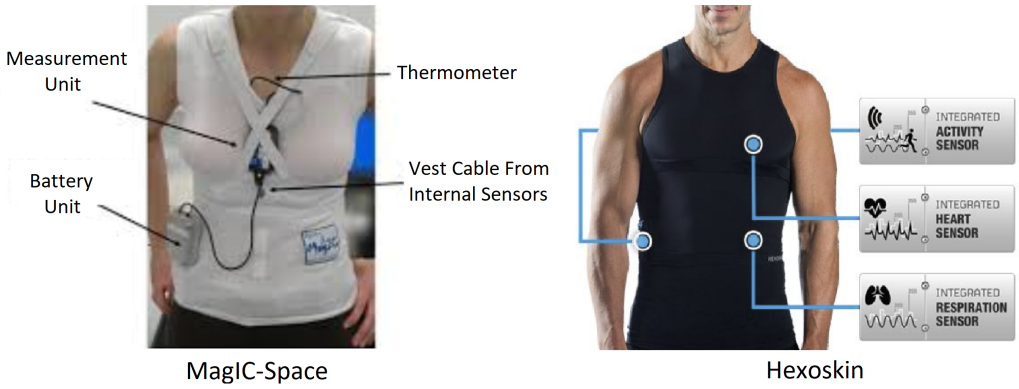

Figure 3.4: Modern examples of torso-based wearable systems. Left: MagIC-Space sleep monitoring system ([55]). Right: Hexoskin health monitoring device, precursor of the Astroskin wearable platform ([83]).

measurements. This garment was capable of collecting quality data in which 96% of the recordings were able to be used for subsequent analysis. By measuring a subject during sleep periods, the device was able to avoid the common occurrence of motion artifacts in sensor data resulting from normal daily activities.

The second wearable sensing garment developed for microgravity environments is the Astroskin device developed by Canadian company Carre Technologies. The Astroskin garment is an extension of the commercially available Hexoskin wearable device, modified for usage onboard the International Space Station. Both the Hexoskin system and the MagIC-Space can be seen in Figure 3.4. The Astroskin device features a three-lead ECG, a plethysmography sensor, forehead located PPG, a skin temperature sensor, and an accelerometer. From these sensors, the company claims to be able to measure heart rate variability, lung capacity and respiration, pulse oximetry, skin temperature, activity levels, and blood pressure [56]. The garment was worn on several occasions in 2019 aboard the space station to prove its viability as a biosignal monitoring device in orbital environments. An independent validation

of the Astroskin indicated that while the device is certainly capable of recording data and presenting it to users in real time, the accuracy and precision of some measurements, particularly blood pressure and respiration rate, was severely lacking at the time [57].

3.4 Conclusion

Wearable technologies allow greatly enhanced portability of multiple sensor modalities as compared to gold-standard measurement methodologies such as the 12-lead ECG (heart rate), the turbine flowmeter (respiration), or motion capture technology (movement). These devices can also be considerably less invasive than their traditional counterparts. Wearable device sales have increased from $16 Billion in 2016 to $44 Billion in 2020 worldwide, not only because of the reduced cost of sensor components as manufacturing technologies advance but also because of an increased acceptance of m-health methodology in clinical research environments [58].

There are also specific measurement cases which must be considered when deploying wearable platforms. Some of these situations are readily apparent, but some are population specific and there is a risk they may not be addressed before data collection methodology is in place and possibly already completed. Subject populations and collection environments must be carefully evaluated before finalizing any data collection implementation.

In the next chapter, we will propose a multimodal measuring system which leverages several of the technologies described in this chapter and in the previous chapter. This upcoming chapter chronicles the hardware development efforts of this thesis.

Chapter 4

Proposed Measurement System

4.1 Measurement Methodology

The Wearable Integrated Health Monitoring System (WIHMS) is proposed as an open source, multimodal modular health monitoring wearable device. The software and all custom design files are available under the MIT open source license [59] in a public GitHub respository [60].

Sensors are sewn into the fabric of a form-fitting polyester long sleeved cycling

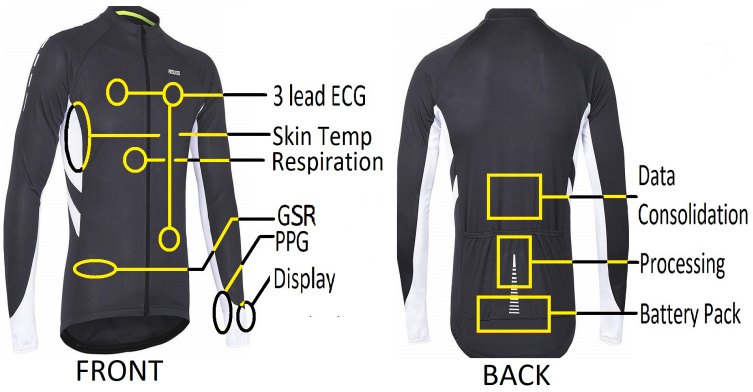

Figure 4.1: Visual representation of the WIHMS layout

shirt, as depicted in Figure 4.1. One galvanic skin response sensor is sewn into the garment's left flank. ECG leads are anchored along the hems of the shirt with the pads protruding through the front of the fabric to place leads on either side of the user's chest and upper left leg. The skin temperature sensor is sewn along the serratus anterior in order to shield the device to rapid skin temperature changes resulting directly from the ambient environment, while the ambient temperature sensor is mounted to the exterior face of the Arduino. PPG information is collected from the left index finger from both 880 nm and 660 nm wavelengths, and the air gap from the sensor surface to the skin surface is minimized by stitching a band of $\frac{1}{4}$ inch elastic around the wrist and finger. A Bosch BNO055 IMU is mounted to the lower sternum in order to return gyroscope-based respiration measurement.

Incorporating low-cost photoplethysmography methods with surface biopotential measures of heart activity may improve processing of artifacts [61] and can provide additional information regarding the user's health such as cuffless blood pressure measurement [62]. Moreover, inclusion of additional sensor modalities, like those for characterizing galvanic skin response and skin temperature may enable further characterization of overall health status [63].

The WIHMS garment was worn by a single subject (male, 25 years old) for 80 collection periods of 11 minutes each in length to determine the system's initial validity for use in collecting physiological data. The collection length was established based on the assertion in [64] that frequency-based HRV metrics require between 5 and 10 minutes to return reliable results. These datasets were collected over the course of two months in order to qualify both the ruggedness of the garment as well as track self-perceived stress over a reasonably large span of time. Before each collection, the

subject first completed four or the self-evaluation methods of stress outlined in Section 2.3. The first of these is a modified version of the Cohen Perceived Stress Scale (PSS), where the recall period of the PSS was modified from one month to one day so as to provide a more granular characterization of stress over time. The subject also completed the Ardell Wellness Stress Test, the Coping Resources Index for Stress, and the Symptoms of Stress Frequency Index. These three tests were not modified from their published form. Responses to these self-evaluations were recorded in a Microsoft Excel document alongside the date on which the collection took place, the time at which the data collection began, and the filename of the collection as generated by the WIHMS garment.

4.2 Selected Sensors

Hobbyist grade sensors were selected to perform the sensing requirements proposed in the WIHMS platform. By making this decision, the total cost of the sensor array was kept to below $170 and the variety of sensors available for use in the garment was expanded greatly. These advantages are explained more thoroughly in section 4.3.

A MAX30102 particle sensor breakout board from MAXIM [65] monitors photoplethysmography (PPG) from two light emitting diodes of different wavelengths. A red LED emits light at 660 nm, and an infrared LED emits light at 880 nm. The MAX30102 operates from a 5 V supply, but a 3.3 V logic level. The cabling to the PPG sensor runs along the underside of the subject's left arm, and is attached to the pad of the left index finger in accordance with the assertion by Longmore et. al that this location is the most reliable for heart rate, blood oxygen saturation, and respi-

Parameter	Default	As Set	Maximum
LED Brightness	255	60	255
Sample Averaging	8	4	32
LED Mode	1	2	3
Sample Rate (Hz)	100	3,200	3,200
Pulse Width (μs)	411	69	411
ADC Range	4096	8192	16,384

Table 4.1: MAX30102 Default settings and settings as used in data collection. Settings were altered to avoid sensor saturation and to increase data rate of the device.

ration rate [66]. The setup parameters were modified to prevent signal saturation at the collection site and to increase the data rate of the sensor data. Settings used for data collection are shown in Table 4.1 alongside the default values:

Electrocardiography (ECG) sampling is performed using a 3-lead Gravity brand heart rate monitor. This device was selected due to the design based on a low noise Analog Devices AD8232 chip [67]. This device is a single-chip signal conditioning block developed specifically for differential monitoring of biopotential signals. As a result, the device by nature filters signal noise resulting from small movements or non-ideal electrode placement. Additionally the AD8232 was designed for low power applications, requiring less than 0.2 mA of current from a 5 V supply. Leads are configured in a Type I configuration [68], with the positive and negative leads on either side of the subject's upper chest. The ground lead is fixed to the subject's lower left leg. All measurements considered in this thesis were performed using 3M Red Dot disposable ECG electrodes [69].

A BNO055 Absolute Orientation Sensor [70] resides on the subject's lower sternum for non-intrusive respiration measurement. This device was selected as it not only features both an accelerometer and a gyroscope, but also natively runs several onboard fusion and filtering algorithms in order to reject noisy data. Gyroscopic

respiration sensors have been shown to provide reliable measurements, even in some cases where accelerometry based respiration is unsuccessful due to noise [71]. This sensor is connected to a 5 V supply, and operates at a 5 V logic level.

Two custom thermistor-based temperature sensors are used to track skin surface temperature and environmental ambient temperature to infer the extent of the subject's temperature regulation. These analog sensors consist of a simple resistive voltage divider circuit, wherein one component in the divider varies its resistance based on the temperature of the silicon substrate. The output voltage of this divider can then be converted to a temperature measurement by considering the Steinhart-Hart Equation

$$T = (A + B * ln(R) + C * (ln(R))^3)^{-1} \tag{4.1}$$

where T signifies the temperature in Kelvin, R signifies the resistance of the semiconductor substrate, and parameters A, B, and C are the Steinhart-Hart coefficients. These coefficients are specific to each individual model and type of semiconductor. In the case of the Vishay NTCLE100E3474GB0 device utilized in the WIHMS garment [72], these parameters are calculated to be 7.86×10^{-4}, 1.81×10^{-4}, and 8.84×10^{-8} respectively. One of these sensors is located in the left armpit to measure body temperature. The second sensor is located on the back of the garment in free space, with approximately two inches of standoff from the top of the microcontroller in order to measure environmental ambient temperature. These sensors are connected to the 5 V supply of the garment.

A custom electrodermal activity sensor is used to collect galvanic skin measurements (GSR) from the subject's left flank in order to monitor the electrodermal activity in accordance with stress reactions [63]. This sensor consists of two flat con-

ductive pads spaced $\frac{1}{2}$ inch apart. These pads are connected to the collector and the emitter of a Micro Commercial S8050 NPN transistor. The gate of this transistor is pulled to ground with a 5.1 kΩ resistor, and a 5 V supply voltage is applied to the collector. Based on the level of electrodermal activity of the subject's skin, the current through the transistor will vary and the voltage at the emitter can be measured as an indicator of galvanic skin response.

4.3 Communications and Hardware Interface

The WIHMS platform utilizes a combination of three communication methods: analog voltage, Inter-Integrated-Circuit (I2C) communication, and Serial Peripheral Interface communication. These three methods merge at a custom hardware interface board before being passed to a microcontroller for data file logging. Before discussing the hardware communication interface (Section 4.3.3), each of the digital communication methods must be explained in more detail.

4.3.1 I2C Communication

The I2C protocol is a 2-wire serial interface standard first introduced by Philips Semiconductor in 1982 [73]. This specification was introduced for applications favoring simplicity and low expense. Compared to many other communications protocols, I2C is considered a slow solution. The original I2C standard data rate is a mere 100 Kbps, while most modern devices operate at 400 Kbps or even up to 3.2 Mbps in some high

speed devices [74].

The two wires utilized by the I2C protocol are the clock line and the data line, usually designated “SCL” and “SDA” respectively. Each line features a pull-up resistor to a “high” logic voltage, as the lines are connected either in either an open-drain or open-collector configuration depending on the transistor technology used in the design of the I2C device. A basic I2C bus layout is shown in Figure 4.2. The clock line is generated by a single master device, and connected to all Slave devices present on the I2C bus. However, devices may alternate between Master and Slave roles in between transmission events. Each device on the bus is assigned a unique 7 bit identification address, though in some cases this address is extended to 10 bits. In theory, 127 Slave devices may be accessible by a single Master device using 7 bit addressing, or 1024 devices when using 10 bit addressing. In reality, the number of devices on the bus is limited by bus capacitance. According to the official standard, only 400 pF of capacitance is allowed on the line in order to prevent a false latch condition. At high levels of capacitance, RC characteristics dominate the behavior of the rising and falling edges of the SCL and SDA signals, in some cases to the extent that the line voltage does not cross the logic low/high threshold. In this situation the Slave devices sit idle, waiting for a data signal that can never come. Additionally, because of the open-drain/open-collector configuration of the transmission lines, any device on the bus can pull either line to ground voltage at will.

The I2C standard was developed for communication between multiple integrated circuits on a single circuit board, with off-board communication handled by more rugged industrial standards such as RS-232 or RS-485 [76]. Due to the simplicity and low cost of implementing the I2C standard, many inexpensive digital sensors make

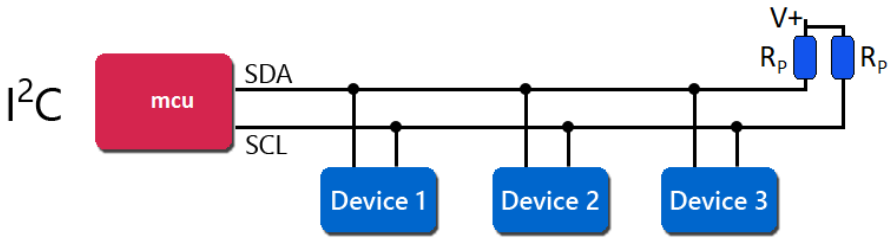

Figure 4.2: Wiring overview for an I2C bus. Reproduced from [75]

use of I2C across cable runs and rely on pull-up resistors to mitigate the RC effects of extending the communication bus, as shown by the extent of I2C device addresses compiled by Adafruit at [77].

There are three main solutions to the issue of I2C bus capacitance when going off-board. First and most obvious, the number of devices on the bus should be kept to a minimum and excess cable length should be eliminated. Every additional device on the bus adds not only the input capacitance of the device, but also of the necessary cabling to link the Master and Slave.

The second solution is to strengthen the pull-up resistors by lowering their value. The I2C spec sets an upper limit of 3 mA on the drain current of either line. This corresponds to a resistance value of 1.67 kΩ on a 5 V line, or 1.1 kΩ on a 3.3 V line. Many devices populate the I2C pull-up resistors with values between 4.7 kΩ and 10 kΩ with the assumption that multiple devices will be placed on the same bus, all with their own resistors which will add in parallel and reduce the equivalent series resistance. Increasing the pull-up value introduces a safeguard against violating the 3 mA current limit, but also inherently slows the rising edge of the I2C signals. With this knowledge, all pull-up resistors can be removed from all devices on the bus and a single set of correctly sized resistors can be populated on the line. One downside of

this approach is that in power-sensitive applications, lowering the value of the pull-up resistance inherently increases the power draw of the system during data transmission events.

Finally, active I2C buffers or I2C isolation chips can be used to isolate the bus into smaller sections, each of which may have an individual capacitance up to the 400 pF limit. These chips inherently introduce a constant propagation delay to the signal (the value of which depends on the specific device but is generally in the range of tens of nanoseconds), but by reducing the RC time constant of the bus may reduce the transition time between logic levels. Compared to the first two solutions, this third option is the most expensive. Not only do the buffers themselves increase material expense, but considerations must be made in circuit board layout to accommodate the IC and its supporting components. The buffer also requires a small quiescent current whether or not it is transmitting data. Depending on the pull-up resistors selected, the power consumption of a system implementing this solution is likely to be greater than one which simply opts for strong pull-up values.

It should be noted that some I2C buffers, such as the NXP PCA9615 [78] are available specifically for transmitting I2C signals over long distances up to 3 meters. This is achieved by converting the I2C lines to differential signals to be transmitted, then converting them once again to a single ended signal at the end of the cable run. This method of signal buffering is the most complicated solution to I2C over distance as it requires the most external components, including a second PCA9615 (or similar) to receive the differential signal.

4.3.2 SPI Communication

The Serial Periferal Interface (SPI) communications protocol is another single ended serial communication method intended for short distance data transfer [79]. Unlike the I2C protocol, an features a single Master device directing one or more Slave devices. At minimum, the protocol requires four wires to function. These connections are typically denoted as follows: "SCLK" is a clock signal which is generated exclusively by the Master device. "MOSI," short for "Master Output Slave Input," is the data line from the Master to the Slaves, generally containing an instruction set. "MISO," or "Master Input Slave Output," is a data line for Slave devices to return data to the Master device. Finally "CS" stands for "Chip Select." This line is an active-low signal controlled by the Master device, meaning in order to query a specific Slave device the voltage on this line must be pulled down from a logic-high state. Any Slave device detecting a falling edge on the CS line will become active and attempt to send data back to the Master. As a result, a unique selection line is necessary for every Slave device on the bus. The basic layout of an SPI bus can be seen in Figure 4.3. The number of Slave devices on an SPI bus is therefore limited by the number GPIO pins available on the Master device. Unlike I2C, however, Slave devices do not require unique addressing in order to interact with the Master.

The SPI clock is set considerably faster than one used for I2C devices, often between 3 and 8 MHz on embedded devices. The lack of pull-up resistors on data lines allows SPI devices to operate at higher data rates with less concern for RC charging effects on the lines. SPI protocol is traditionally utilized for very short range communication between devices; as the distance between devices increases, the clock

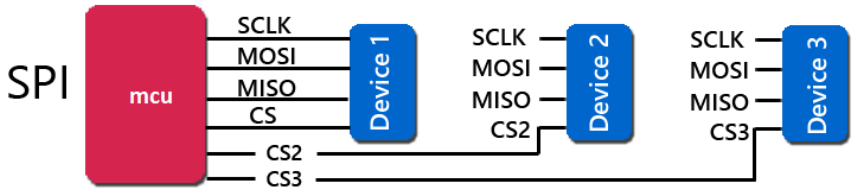

Figure 4.3: Wiring overview for an SPI bus. Reproduced from [75]

frequency must be lowered to combat parasitic and stray capacitance interfering with rising and falling signal edges. 3 meters of cable-length separation between Master and Slave devices is considered to be the maximum reliable transmission distance for the protocol, and at this distance the clock frequency may need to be reduced to as low as 125 kHz to guarantee reliable data transfer.

4.3.3 Hardware Resources

An Arduino Mega 2560 is used to collect and store sensor data. The ECG sensor, GSR sensor, and temperature sensors provide analog voltage outputs, which are connected directly to individual analog input pins on the microcontroller (pins A1, A4, A7, and A11). The remaining sensors utilize the common Inter-Integrated Circuit (I2C) protocol, discussed in detail in Section 4.3.1 above. This system of communication is simple to both implement and troubleshoot, but is very sensitive to bus capacitance slowing the rising edge of the digital signal. To work around this limitation, strong pull-up resistors (2.2 kΩ - 4.7 kΩ) are used in combination with Texas Instruments TCA9517 I2C buffer chips on a custom printed circuit board to isolate the line to each sensor and prevent the bus from latching up during operation at the 400 kbps standard I2C "fast" data rate. The WIHMS system is designed to be powered via a

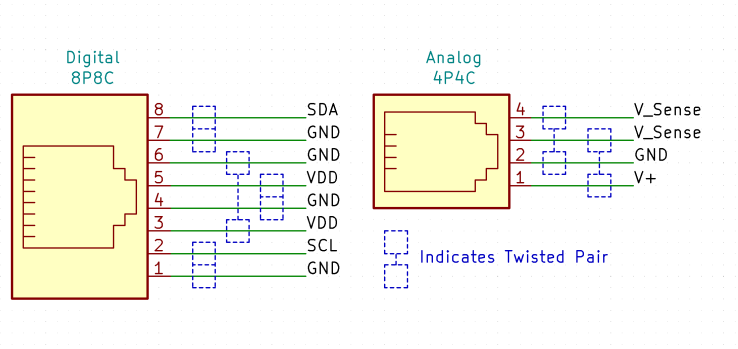

Figure 4.4: Wiring convention for digital and analog sensors over twisted pair cabling

common 5 V DC power bank, but the onboard regulators are capable of driving the system from an input of between 2.7 - 11.8 Volts. Design files and sensor resources are located in the WIHMS Open Source file repository on GitHub [60].

The custom designed data-aggregation board served to interface the Arduino to each of the sensors. This board is mounted to the small of the back to pass sensor data to the microcontroller located in a pocket on the lower rear of the shirt. Connection between the sensors and the data-aggregation board is accomplished using common off-the-shelf connectors. Analog sensors utilize a 4-pole, 4-contact RJ11 telephone jack, while I2C-based sensors utilize an 8-pole, 8-contact RJ45 connector. I2C communications are carried between the sensors and the data-aggregation board over common CAT6 twisted pair cable. Twisted pair conductors were assigned in order to minimize crosstalk and electromagnetic interference. SDA, SCL, and postitive voltage lines of the sensor's I2C are each twisted around their own ground wire. The wiring convention for both the analog and digital communication, including twisted pair assignments as described above, is shown in Figure 4.4. A block diagram for the entire system layout is located in Appendix Section C.

The CAT6 cable currently in use on WIHMS is not shielded as a capacitance limiting measure, but performance of the I2C bus in its current state suggests that the lines could be shielded for enhanced noise resistance without sacrificing bus reliability. The capacitance of the unshielded twisted pair is characterized by the equation

$$C = \frac{0.7065}{ln(\frac{s}{d})} * e_r \tag{4.2}$$

where C is the capacitance in pF/inch, s is the separation of the two conductors in mm, d is the diameter of the conductors in mm, and e_r is the dielectric constant of the conductor insulation. In the case of the 24 AWG cabling used in WIHMS, the equation can be populated as follows:

$$C = \frac{0.7065}{ln(\frac{1.27}{0.510})} * 2.26 = 1.75\, pF/inch \tag{4.3}$$

The maximum individual cable run in WIHMS is the 1 meter run from the data board to the PPG sensor. From the above calculation, it can be calculated that the parasitic capacitance of this cable run is 73.5 pF. Additionally, IEEE standard 802.3 [80] states that the maximum capacitance allowable in the connection between a manufactured 8p8c jack and 8p8c recepticle is 4 pF. The TCA9517 datasheet specifies a maximum input/output capacitance of 13 pF. As a result, the capacitance of the longest CAT6 interconnection is assumed to be 93.5 pF or less, well below the permissible 400 pF limit.

Each port on the data-aggregation board is configurable for devices requiring 3.3 or 5 V logic levels, as well as 3.3 V or 5 V power supplies. This level of configuration allows other sensors and sensor modalities to be incorporated into the WIHMS system

without hardware redesign or rework. The board also features an onboard MAXIM DS3231 precision real time clock to timestamp each data file as it is created.

The data file is stored using the SPI protocol to communicate with and write to a 64 GB Samsung EVO high speed MicroSD card formatted using the FAT32 standard [81]. Filenames are assigned based on the date provided by the real time clock in the format "MMDDHHmm.bin." Filenames originally utilized the format "YYYYMMDDHHmm.bin," but the standard 8.3 format convention utilized in FAT filing systems limits filenames to eight characters, with a three character file type extension.

4.3.4 Software

The open source nature of the WIHMS project means that the embedded processing flow may be adapted at will to suit a particular set of requirements as needed. For the work presented in this thesis, the garment was configured to collect 11 minutes of data from the six sensors described in Section 4.2. At power-on, the microcontroller first queries the digital sensors in order to ensure the availability and proper operation of the I2C bus. After verifying the operation of the I2C bus and presence of all anticipated sensors, configuration parameters are written to the digital sensors and the real time clock is queried for the current date and time. The system clock notes the moment that data collection is initiated, to the millisecond. During this boot process, a splash screen displays the following text to the subject on the wrist mounted OLED display:

WIHMS 1.1

Blake Hewgill 2020

Rev. 20200405

Welcome

The microcontroller then verifies the presence and proper formatting of the micro SD memory card for data storage. If the card is not present, the subject is notified *"SD NOT Connected"* on the OLED display and WIHMS enters a standby mode while the SD card fitment is verified. Otherwise, a binary file is created on the card under the name "MMDDHHmm.bin" as noted above in Section 4.3.3. The data collection process then begins, and the prompt *"Collecting Data"* is presented to the subject.

Sensors are queried in a loop for the duration of the data collection. ECG, PPG, temperature, and GSR measurements are queried every cycle of the loop. The BNO055 used for respiration measurements is queried only every 16 cycles of the loop. Because the normal resting respiration rate of adults is between 12 and 20 breaths per minute [82], or 0.20 Hz to 0.33 Hz, this lower sampling rate is still more than adequate to characterize breathing patterns, complete with inflection points and the possibility of hyperventilation (30 breaths per minute or a 0.5 Hz respiration rate). The sensors are queried at 200 Hz, with the exception of the 12.5 Hz respiration sampling frequency. Every 16 measurements, the temperature sensor to be queried alternates between the skin temperature sensor and the ambient temperature sensor. After 11 minutes of data collection, WIHMS transitions to a standby mode, and displays the message *"Complete"* to the subject.

A simplified software process flow for the WIHMS data collection script can be seen in Figure 4.5, and the full code currently used to perform data collection is

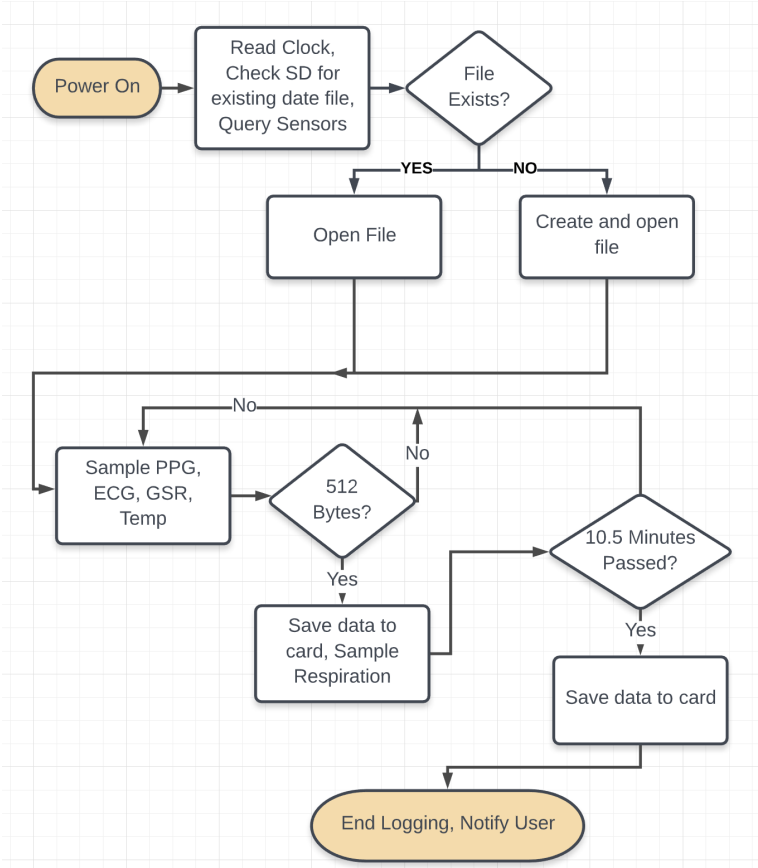

Figure 4.5: Flow chart outlining data collection process

included in Appendix D.1.

Each set of measurements from the entire sensing array is stored in a data line 16 bytes long. An array of 32 16-byte structures is created in the Arduino memory. The format of this data structure is shown in Table 4.2, which also provides the latency time to query each sensor. The length of the array is selected to be 32 due to a requirement in the SD standard which specifies that the SD card be written to in 512 byte blocks. Each time the 512 byte memory block fills with data, it is written to the SD card and overwritten with the next 32 sets of measurements. After 10 such blocks are written to the SD card the file is closed and reopened to finalize the write process

Data	Time	ECG	IR PPG	Red PPG	Temp.	GSR	Respiration
Type	Long	Int	Int	Int	Int	Int	Int
Length	4 Byte	2 Byte	2 Byte	2 Byte	2 Byte	2 Byte	2 Byte
Latency	1.8 μS	0.1 mS	1.25 mS	1.25 mS	0.1 mS	0.1 mS	0.3 mS

Table 4.2: Data structure format in Arduino memory and sensor response times

and permanently write the data blocks onto the SD card. This redundancy ensures that in the event of power loss or other system failure, only up to 1.6 seconds of data may be irretrievably lost. A completed 11 minute data file occupies 2001 kilobytes of memory on the card. On the currently used 64 GB Samsung memory card, 33,537 files may be collected before filling the available space. This equates to approximately 256 days of continuous logging.

4.4 Cost Comparison

As shown in the full bill of materials provided in Section B of the Appendix, the final material cost for the WIHMS platform is $169.38. This material cost does not take into account any economies of scale, as only a single prototype was constructed. The bill of materials also does not consider development labor, as it is impossible to predict the exact value or required assembly time of subsequent iterations or reproductions of WIHMS. As it stands, the component cost of the garment makes it an attractive wearable sensing option as compared to other consumer devices available on the market.

Table 4.3 presents various existing consumer oriented wearable devices and their sensing capabilities. Many unimodal options exist, but the introduction multi-sensor arrays unvaryingly leads to increased unit cost.

Device	Mfr.	ECG	PPG	GSR	EMG	IMU	Temp.	Cost
WIHMS	-	X	X	X			X	$169
mBody	Myontec				X	X		$1040
AppleWatch	Apple		X			X		$399
Ring	Oura		X					$299
Hexoskin	Hexoskin	X	X			X		$579
Biostamp (3)	MC-10	X			X	X		$2500
Zephyr	Medtronic	X				X		$667

Table 4.3: Features and cost comparison of some available wearable sensing devices

The Hexoskin unit [83] represents the most similar commercial offering to WIHMS, but at a unit cost of more than three times the component cost of WIHMS. A variant of the Hexoskin, the Astroskin [84], offers several additional sensor modalities to the base garment, but is not available for purchase on the consumer market at the time of this writing.

As it stands, the WIHMS platform offers greater versatility in sensor modalities than many current market offerings, whilst also allowing construction at a fraction of the purchase price of other devices. Current market trends seem to be reducing the overall size of wearable devices, but are doing so at the cost of sensor placement. The Oura Ring [85], for example, consists of a non-intrusive finger-worn ring but relies solely on photoplethysmography-based sensing.

4.5 Conclusion

This chapter examined the physical implementation of the WIHMS platform as a data collection device. Individual sensor modalities and devices were discussed, as was their placement on the physical garment. The communication protocols employed

were also presented and discussed with regard to their role in the system. The custom data interface hardware at the core of the system was explained, and all design files are available in the project's GitHub repository. Material cost of the WIHMS system was compared to the retail price of certain other wearable devices available on the consumer market.

This chapter primarily served to justify the completion of Contribution 1: *"An Open Source and Modular Health Monitoring Hardware Solution."* In the next chapter the methodology behind data collection is examined, as is the extraction of features from the data collected using WIHMS.

Chapter 5

Data Collection

5.1 Data Collection Methodology

The WIHMS prototype garment was employed to collect all data examined for the remainder of this thesis. The data rate of the system was set to 200 Hz during collection to encourage high resolution results, as well as conform to the claim from [86] that 50 Hz is a high enough frequency for HRV analysis with a reasonable margin of safety. A deeper investigation by Beres et. al [87] reaffirmed the validity of a 200 Hz sampling rate as sufficient for analysis with a low margin of error of less than 2%.

Binary log files generated by WIHMS were passed through a custom MATLAB file reader to convert the raw bytes into numerical data, at which point each sensor was processed separately to determine several characterizing parameters for each data type. For each of the 80 data collections considered, a single value corresponding to each parameter of interest was calculated using the methods described below in Section 5.2. Each of these parameters was then compared to the self-reported stress measures associated with the data collection. 82 different parameters were attributed

to each data collection, exclusive of self estimation measures of stress.

The 80 data collections considered in analysis were collected between April 6th and June 26th, 2020. For each collection the subject adorned the WIHMS garment, completed all four self evaluations, then remained still for the completion of the recording. Collections were performed during one of three daily time ranges: Morning (4:01 to 12:00), Afternoon (12:01 to 18:00), and Evening (18:01 to 3:59). These ranges were selected to evaluate whether self-reported stress values or the physiological parameters recorded by the system varied based on the time of day at which they were taken. Stress values over the entire dataset annotated with personal and global events can be seen plotted in Appendix C.2.

5.2 Feature Identification

All data from each of the sensors deployed as part of the WIHMS array was processed individually to extract features descriptive of each physiological parameter. Prior to these features being extracted, the data was resampled to 200 Hz. This step is necessary because while the WIHMS garment averages a 200 Hz sample rate, each individual sample is not necessarily spaced evenly at 5 ms apart in time due to the specific processes executed by the microcontroller. Sensors are not necessarily queried at the instant new data is available, and software loop iterations including a flush to the SD card added a nonuniform delay to the time required for completion of the loop. Because each sample is timestamped, however, the data can easily be resampled using shape-preserving piecewise cubic interpolation.

5.2.1 ECG Features

Several ECG features were extracted in both the time and frequency domains. Before features can be extracted, however, the data must be filtered to remove baseline wander, high frequency artifacts, and 60 Hz AC noise. To remove these frequency regions from the data, a two stage filter was implemented. The first stage consisted of a Butterworth bandpass filter with frequency cutoffs of 0.5 Hz and 100 Hz and a stopband attenuation of 80 dB. The 0.5-100 Hz region of ECG is considered to be the standard bandwidth of an ECG signal, and is widely accepted to eliminate low frequency baseline drift as well as harmonics in the data above the 100 Hz region of useful frequency content [88]. A second Butterworth filter was then applied as a notch filter to reject any noise in the data resulting specifically from 60 Hz coupling from the mains. The stopband of this filter was 59.25-60.75 Hz, with a stopband attenuation of 60 dB. The results of these filters on the frequency spectrum of the raw ECG data can be seen in Figure 5.1.

From the filtered data, the QRS complex of each heartbeat (explained in Section 5.2.1) was extracted. This is performed in two ways by leveraging both a custom peak detecting algorithm as well as the traditional Pan-Tompkins algorithm [90]. In the case of the custom algorithm, the ECG data was first transformed to a wavelet series using an inverse maximal overlap discrete wavelet transform with the 'Symlet 4' style wavelet. This process applied to an ECG signal emphasizes the QRS complex of the heart beat, and maximizes the R wave, as outlined by Lahmiri [89]. R-R intervals were then determined by recording the timestamps of peaks with a height greater than 90% of the ECG signal and a peak distance of $\frac{1}{3}$ second or greater, allowing

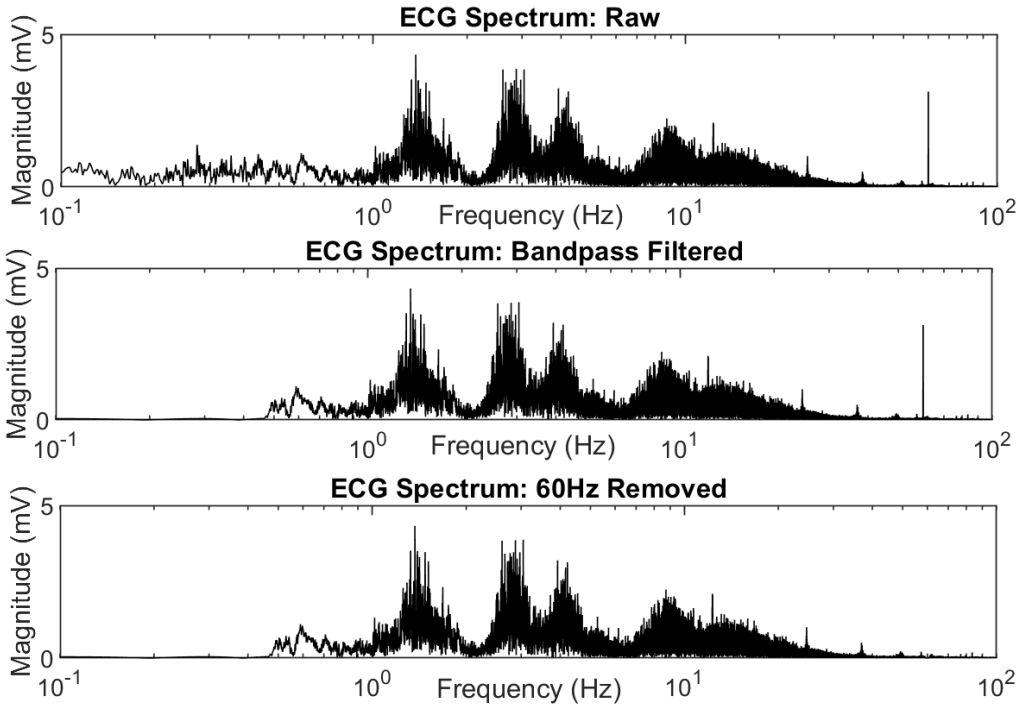

Figure 5.1: Effect of bandpass and notch filters on ECG frequency spectrum

heart rates of up to 180 beats per minute to be detected.

In the case of the Pan-Tompkins algorithm, a 5-15 Hz bandpass filter was applied to the data to amplify the QRS complex and reduce noise. A derivative filter was then applied to amplify the R wave above the Q and S portions of the ECG. The resultant signal was then squared to amplify these peaks further, and peaks were then detected to locate the timestamped location of each R wave.

From the RR-interval data, mean heart rate and standard deviation were calculated, as was the maximum average heart rate over a 1-minute sliding window. Heart rate variability was examined using two parameters: the root-mean-square of successive differences (rMSSD) of heart beats and the LF/HF power ratio of RR intervals [30], both of which are described in further detail in Appendix Section A. The low frequency and high frequency spectrums were isolated by linearly interpolating the RR vector by a factor of five to enhance resolution, and computing the Lomb-Scargle periodigram [91] of the resultant data. Samples corresponding to 0.05-0.15 Hz and 0.15-0.40 Hz were categorized as low and high frequency respectively, and the LF/HF ratio was calculated by dividing the integral of the low frequency region by that of the high frequency region.

5.2.2 PPG Features

From the PPG measurements, a two stage median filter [92] was applied to reduce noise and smooth the raw data prior to feature extraction. After this smoothing, a sliding window was used to extract features in 1 second intervals across the data collection. When the window reached the end of the data collection, statistical features were extracted from the vectors of 1 second calculations. In the current hardware

configuration and PPG sensor location on the left index finger, a 20 second window width was found to provide an acceptable balance between time-resolution and limited baseline wander. Over each 11 minute (660 second) data collection, the 20 second sliding window returned 634 calculations for each parameter of interest.

A peak detection algorithm was applied to the PPG data to locate local minima and maxima to extract features from the dataset. The positive and negative first derivative of each pulse in the PPG data was calculated to determine the rate of change of blood volume in the skin. A peripheral capillary oxygen saturation parameter (SPO2) was calculated using two different methods from equations found in the manufacturer's datasheet for the MAX30102 sensor, reproduced below:

$$PPGratio = \frac{AC(Red)/DC(Red)}{AC(IR)/DC(IR)} \tag{5.1}$$

$$SPO2(1) = -45.06 * (PPGratio)^2 + (30.35 * PPGratio) + 94.99 \tag{5.2}$$

$$SPO2(2) = 104 - (17 * PPGratio) \tag{5.3}$$

where the DC component of the signal was defined as the mean of the signal over the window length, and the AC component was defined as the root mean square of the signal over the window length. There is doubt regarding the accuracy of the calibration values provided by MAXIM, as most of the calculated SPO2 values in the dataset were between 75% and 85%. In a healthy adult, these values should be on the order of 95% to 100% [94].

The dominant frequencies of the red and infrared PPG vectors were also calculated using emsemble empirical mode decomposition as in [93]. Maximum, minimum, median, mean, and standard deviation values were extracted from the 634 sliding

window calculations for the following parameters: Red LED positive rate of change, Red LED negative rate of change, Infrared LED positive rate of change, Infrared LED negative rate of change, heart rate (based on peak-peak intervals), SPO2 (1), and SPO2 (2).

5.2.3 Respiration Features

Because the respiration data was sampled at 12.5 Hz in a "sample-and-hold" methodology compared to the 200 Hz of the other sensors, it first had to be interpolated before features could be extracted along the same time intervals of the other sensors. The data was first filtered with a moving average Gaussian window to interpolate between adjacent sample-and-hold steps. A moving average filter was applied to the data which reduced the amplitude of the respiration signal, but smoothed the data such that the timestamped locations of extrema from inhalation and exhalation could be determined by using the same peak detection algorithm applied to the PPG data. The effect of these two filters on time-series data can be observed in Figure 5.2. Peak and valley time-series locations are identified and indicated by black circles on the figure after filtering.

Duration between inhalations (peak locations) was calculated, as was the difference between exhalations (valley locations). Amplitude of respiration was also calculated by subtracting the values of exhalation points from their adjacent inhalation points. Mean, median, maximum, minimum, and standard deviation of each of these measures were calculated and utilized as features. The average respiration rate in breaths per minute is also calculated by dividing the length of the data collection in minutes by the number of detected inhalation points.

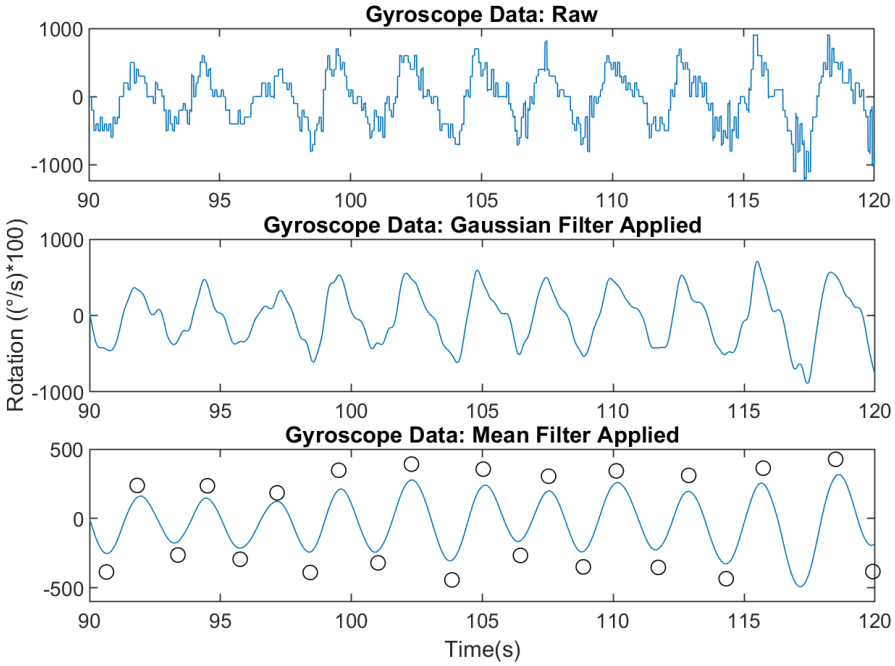

Figure 5.2: Top: Raw 12.5 Hz "sample and hold" respiration data. Middle: Respiration data after the application of a moving average Gaussian window. Bottom: Respiration data after application of both a moving average Gaussian window and a mean filter. Inhalation and exhalation events indicated by black circles.

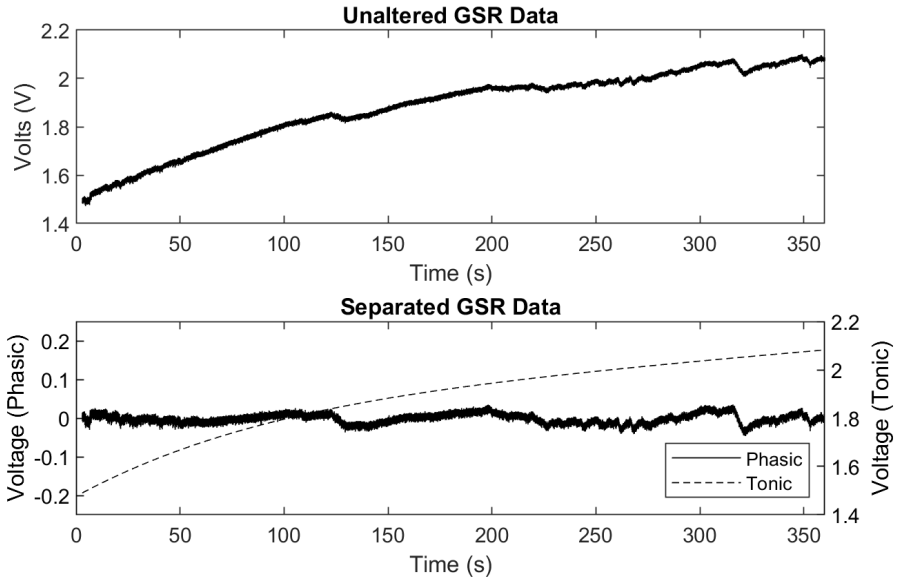

Figure 5.3: Separation of Tonic and Phasic GSR components (Bottom) from raw GSR sensor data (Top) via 5^{th} order trendline removal.

5.2.4 GSR Features

The tonic (near DC) and phasic (approximately 0.2 Hz to 2 Hz) components of the galvanic skin response were extracted from the raw GSR data by subtracting a 5^{th} order trendline from the data vector. The MATLAB 'detrend' function weights each power automatically and allows the GSR components to be reliably separated as seen in Figure 5.3. Mean, median, maximum, minimum, standard deviation, and dominant frequency are extracted from both measures. Contact-loss events were easily removed from the phasic data by eliminating GSR events with rates of change of 0.1 V/sec or greater. Contact-loss events did not appear to affect the tonic skin response data.

Many data collections demonstrated early drift behavior, where early timestamps revealed a positive rate of change which tapered off as the collection continued. This may have been due to the inflexible FR4 material used for the sensor electrodes. The

electrodermal activity intended to be measured by the sensor may have collected to form a conductive layer between the skin and the electrode, increasing conductance quickly over a short time before settling to a more consistent level for the remainder of the data collection.

5.2.5 Temperature Features

The first step in processing the temperature data was to separate the skin temperature and ambient temperature measurements. In order to restrict the binary data structure to 16 bytes per sample, the two temperature sensors were forced to share the same two bytes reserved for temperature measurements. Each time the 512 byte buffer was flushed to the SD card, the temperature sensor being queried was switched between the ambient location and the skin location. As a result, each four seconds of data taken at 200 Hz contained 400 samples of skin temperature data and 400 samples of ambient temperature data, separated by 16 samples at a time.

To extract both sets of temperature data (skin and ambient), the raw data was duplicated to have two identical vectors. For the skin temperature data, 16 samples were linearly interpolated after each 16 samples with a zero sample offset, overwriting the ambient data in that vector. In the case of the ambient data, 16 samples were linearly interpolated every 16 samples after a 16 sample offset, effectively overwriting the skin temperature data in that vector. A third data vector was then generated by subtracting the ambient data from the skin data to determine the relative difference between skin and ambient temperature. Mean, median, maximum, minimum, and standard deviation values were calculated from the skin temperature data and the relative difference temperature. The dominant frequency of each of these vectors was

also calculated. Ambient temperature data was not considered a physiological parameter, and therefore no features were extracted specifically from those measurements.

5.3 Conclusion

In this chapter, specific features of interest in the raw data were identified. The feature extraction and data processing methods were described from the initial conversion from binary data through the final calculations for each of the various sensors. Overall, a total of 82 features were calculated for use in stress prediction efforts.

Several datasets were omitted from analysis due to extensive data artifacts caused by a depleted battery and subsequent power loss, or in some cases by EMG activity picked up by the ground lead of the ECG sensor. Additional data collection was performed using a fully charged power pack to replace the corrupted data, and was completed without issue. The ECG ground lead was relocated away from major muscle clusters to prevent ground wander due to muscle activity. In the 80 datasets considered, the data was artifact and corruption free outside of these instances. Data quality observed from the WIHMS garment shows promise in the system as an accessible utility for wearable sensing data collection.

GSR sensors occasionally lost contact with the subject's skin, but in the cases where this did not occur the data returned by these sensors characterized a quantitative representation of skin conductivity at the collection site. Contact-loss events were easily isolated by examining the first derivative of the sensor data; large differentials (more than 0.1V/sec) indicated that the sensor in question lost contact with the subject. These events were limited in occurrence and can be resolved by stitching

elastic around the torso region of the garment to encourage full sensor contact. The temperature sensors returned appropriate values each time they were queried, and the rate of change between each value was within reasonable expectation based on [95]. The difference between skin and ambient temperature readings was observed to fluctuate slightly over the course of the data collection, though it was unclear whether this behavior was due to acute instantaneous stress response or a draft in the collection environment.

In the next chapter, features identified and extracted in this chapter will be used in an attempt to predict the self-estimated stress values associated with each data collection. A correlation analysis will identify linear and monotonic relationships between features and stress levels, then machine learning methods will be employed to create a subject-specific model to predict stress levels of each data collection.

Chapter 6

Stress Prediction

Machine Learning (ML) methods were applied to examine the relationship between specific features of interest extracted from the data and the associated self evaluation values of stress. An ML approach was pursued over traditional statistical methods because these methods are quickly scalable and can detect subtle trends in high dimension datasets. Three analysis methods were employed to understand these relationships: Correlation, Regression, and Classification. First, basic linear correlation coefficients were calculated between all features and each of the four self perceived stress evaluation tests. Correlation amongst each the extracted features was also calculated to explore whether certain features could be omitted from processing to reduce the dimensionality of the prediction problem by eliminating features with low feature-stress correlation.

Machine Learning Regression models were then trained on the reduced data to attempt to predict the associated self evaluation stress values of each of the four self tests. For each self test method, 19 different algorithms were trained, and the most accurate model for each self evaluation was given further attention to attempt to

improve performance by tuning hyperparameters. Finally, Machine Learning Classification models were also trained to attempt to predict the categorical outcome of three of the four self evaluation tests. Unlike in the Regression case, 25 models were trained as predictors for each of three self test outcomes. The most accurate model was then tuned in an attempt to increase classifier performance on the test data.

6.1 Correlation Analysis

The first effort in stress prediction was to determine whether or not a linear relationship existed between the data features and the self evaluated stress levels. This was accomplished by determining the Pearson Correlation Coefficient, r, between each feature and each of the four self evaluation methods. The Pearson Correlation Coefficient is a statistical measure of linear correlation between two variables. Values fall between -1 and 1, where 1 indicates a perfectly positive linear relationship between the two variables and -1 indicates a perfectly negative linear relationship. A correlation value of 0 would indicate the complete absence of a linear relationship between the two variables. The correlation coefficient of two variables A and B is calculated by

$$r_{A,B} = \frac{cov(A,B)}{\sigma_A * \sigma_B} \tag{6.1}$$

where $cov(A,B)$ is the covariance between variables A and B, and σ denotes standard deviation.

One of the core assumptions of Pearson correlation analysis is a normal distribution of the variables considered. In order to determine whether the data features were normally distributed, a Lilliefors test was used on each feature independently.

Model	Feature	Correlation (r)	Significance (p)
Cohen PSS	Mean Exhale Length	-0.332	.0028
Cohen PSS	Mean Inhale Length	-0.301	.0071
Symptoms	Mean Exhale Length	-0.351	.0015
Symptoms	Mean Inhale Length	-0.324	.0036

Table 6.1: Pearson rank coefficients for data features and self-estimated stress

This test is used to test the null hypothesis that data originates from a normally distributed population. Mean and variance are calculated for each feature, after which the discrepancy between the empirical distribution function and the cumulative distribution function described by those values is determined. If the discrepancy is large enough, the data is not believed to resemble a normal distribution.

In the case of the 82 calculated features from the WIHMS dataset, ten were determined to be normally distributed. These were maximum and average heart rate, LFHF ratio, minimum inhalation and exhalation lengths, respiration rate, mean and median skin temperature, mean rate of change of skin temperature, and both mean and median skin temperature difference from ambient temperature.

Across the four self-estimation methods examined for Pearson linear correlation, only four of the normally distributed features were determined to have significant relationships with stress at a level of $p < .05$. The p value is the evidence against the null hypothesis that the feature and a self-estimated stress value are unrelated. The smaller the p value, the more likely the feature in question is related to the stress test value being considered. Pearson correlations can be seen in Table 6.1. It should be noted that even when investigating the linear correlation across all 80 data collections, there is not a single significant relationship between the features examined and the Ardell Wellness or SCRI test scores.

After these less than encouraging results, the Spearman Rank Coefficient was also calculated. This parameter quantifies the existence of a monotonic relationship rather than one which is purely linear. In other words, the Spearman rank determines the extent to which every output increases or decreases compared to the previous output, given that the every input increases or decreases compared to the input before it. This measure of rank correlation relaxes assumptions on the data that are necessary for a Pearson correlation, specifically the absence of outliers, the linearity between inputs (data collection features) and outputs (self-estimated stress levels), and the normal distribution of the data. The formula for the Spearman Rank is given as

$$r_{A,B} = \frac{cov(rg_A, rg_B)}{\sigma_A * \sigma_B} \tag{6.2}$$

where rg_A and rg_B are the sorted rank of each entry of A and B. The Spearman correlations returned more promising results with 24 significant relationships among the four tests, and at least two relationships for each test. These correlations can be seen in Table 6.2.

From these results it is clear that many more features are monotonically related to self-estimated stress levels as compared to the number of features with a linear relationship. Moreover, it appears that monotonic relationships are predominantly present in respiration features rather than features extracted from other physiological responses. From these calculations it is also evident that of all 82 features considered, only the maximum inhalation length is monotonically related to all four self-estimated stress level values.

Pearson and Spearman correlations were also evaluated on subsections of the dataset grouped by the time of day each data collection was created. 43 'Morning'

Model	Feature	Correlation (r)	Significance (p)
Cohen PSS	STD of RR-Interval	0.309	.0056
Cohen PSS	Min Exhale Length	-0.434	.0001
Cohen PSS	Mean Exhale Length	-0.391	.0004
Cohen PSS	STD of Exhale Length	-0.352	.0015
Cohen PSS	Max Inhale Length	-0.411	.0002
Cohen PSS	Min Inhale Length	-0.364	.0010
Cohen PSS	Mean Inhale Length	-0.352	.0014
Cohen PSS	STD of Inhale Length	-0.349	.0016
Cohen PSS	Max Inhale Amplitude	-0.374	.0007
Cohen PSS	Respiration Frequency	0.357	.0012
Cohen PSS	STD of Tonic GSR	0.408	.0002
Ardell	Min Exhale Length	0.385	.0005
Ardell	Mean Exhale Length	0.380	.0005
Ardell	STD of Exhale Length	0.361	.011
Ardell	Max Inhale Length	0.336	.0025
Ardell	Mean Inhale Length	0.365	.0009
Ardell	STD of Inhale Length	0.361	.0011
Ardell	Respiration Frequency	-0.357	.0013
Ardell	Mean of Tonic GSR	0.324	.0036
Ardell	Median Temperature	0.306	.0061
Symptoms	Max Inhale Length	-0.351	.0015
Symptoms	Max Inhale Amplitude	-0.324	.0036
SCRI	Max Inhale Length	0.335	.0025
SCRI	STD of Tonic GSR	-0.503	.0001

Table 6.2: Spearman rank coefficients for data features and self-estimated stress. The only feature monotonically related to all four self-tests is the maximum inhalation length.

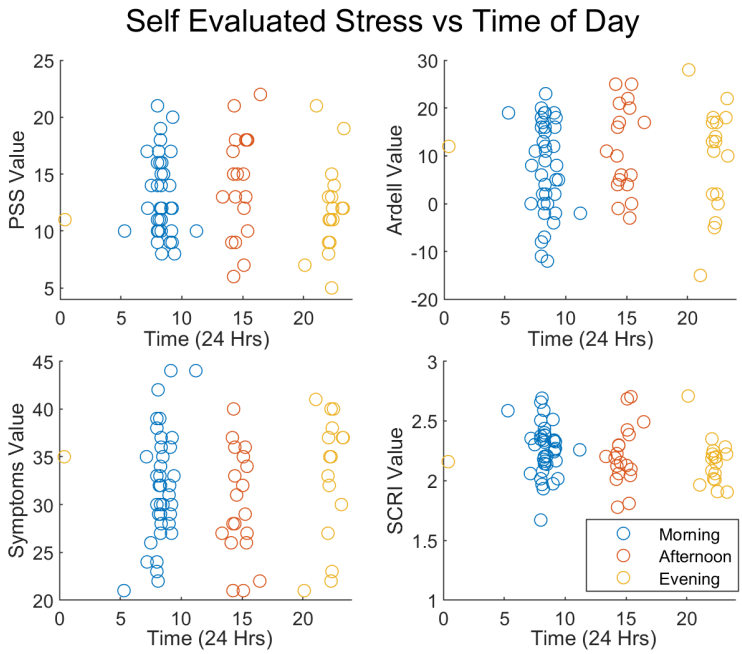

Figure 6.1: Distribution of self-evaluated stress according to time of day

data collections were taken between 4:00 and 12:00. 19 'Afternoon' data collections were created between 12:01 and 18:00, and 18 'Evening' data collections were created between 18:01 and 3:59 the following morning. This division was performed based on the knowledge that some physiological parameters, specifically those related to the heart rate, vary over the course of a normal day. The distributions of self-evaluated stress values over these time periods can be seen in Figure 6.1. The similar distributions of stress values suggest that the time of day a self-evaluation of stress is performed does not significantly impact the outcome of the test. To determine whether or not this assumption is valid, basic statistical parameters were calculated from each of the time-based subsections.

Mean, median, range, maximum and minimum, and standard deviation were extracted from the morning, afternoon, and evening data subsets across all four self-

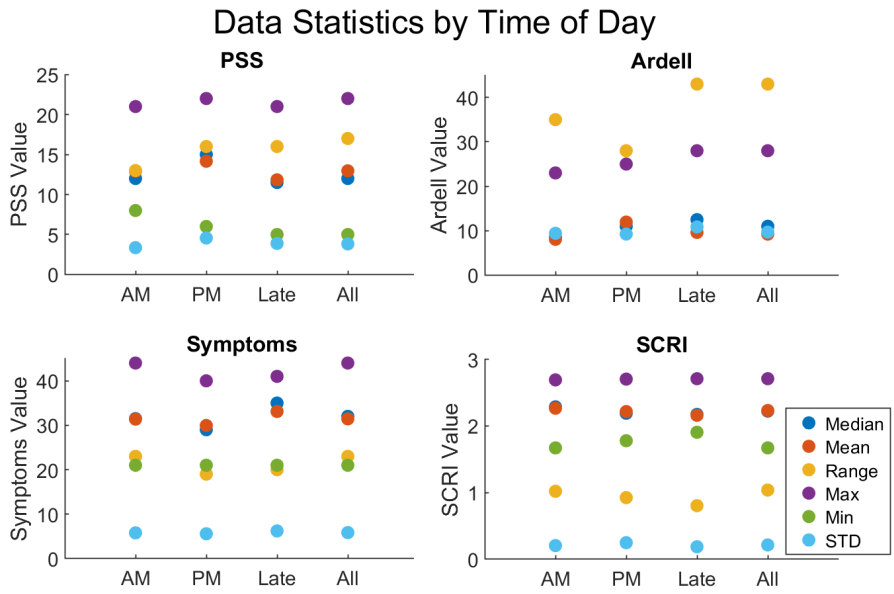

Figure 6.2: Mean, Median, Range, Maximum, Minimum, and Standard Deviation values for all four stress tests based on the time of day the data collection was taken.

evaluation methods of stress. These were then compared to the same parameters extracted from the entire dataset of 80 collections. The results can be seen plotted in Figure 6.2.

Overall, it appears the distribution of stress scores was not significantly affected by the time of day at which a data collection was created. With the exception of the range parameter for the Ardell Wellness test, all of the measurements varied by 10% or less when the time of day was taken into account. Pearson and Spearman correlations were calculated between all data features and the time at which the data collection was created. With a significance level of $p < .05$, ten of the calculated parameters were found to have significant correlation with time of day. Eight of these features were exclusively tied to the PPG-based 'Rate of Change' calculations, all of which had Pearson and Spearman correlation coefficients between .57 and .60. The features in question were the positive rate of change maximum, median, and mean

values of the both LED's and the negative rate of change maximum and standard deviation values of the both LED's. The calculated SPO2 parameters also showed significant linear correlation with time, and returned a Pearson coefficient of 0.40, but the significance level of these same features of the Spearman correlation was greater than .05.

It should be noted that the COVID-19 pandemic seems to have influenced the dataset in a noticeable fashion. When Vermont entered a pandemic lockdown in March of 2020, the University of Vermont transitioned to online and remote coursework. Despite the disruption caused by this shift, the daily schedule and overall routine of the data collection subject was relatively unchanged. It was not until the end of the semester when classes no longer met on a fixed schedule that the subject was affected by the lockdown. The date of the last normally scheduled academic classes for the Spring 2020 semester was May 7, coinciding with the 30th data collection in the dataset.

If the Pearson correlation analysis described above is performed only on the first 30 data collections there are 24 feature-stress Pearson correlations and 29 Spearman correlations across the four self-tests, all of which have correlation coefficients between 0.3 and 0.5. Of the 24 Pearson correlations, five are associated with the PSS, nine with Ardell Wellness, seven with Symptoms, and three with SCRI. Of the Spearman correlations, three are associated with the PSS, 20 with Ardell Wellness, two with Symptoms, and four with SCRI. When the same analysis is performed only on data collected after May 7, there are 14 significant Pearson correlations and 39 significant Spearman correlations. Four Pearson correlations are associated with PSS, three with Ardell Wellness, four with Symptoms, and three with SCRI. 12 Spearman correla-

Significant Relationships: 4/6/2020 - 5/7/2020												
	Pearson						Spearman					
	Mean	Med	Max	Min	STD	Freq	Mean	Med	Max	Min	STD	Freq
RespAmp	1				1					1		
InLen											2	
ExLen		4	2					3,4	2		2	
RedROC(+)			3							2		
RedROC(-)									2			
skinT		4						2		2,4		
skinROC		4	3		1			2	3	2,4	1,2,4	
tempROC		3	3	3				2				
SPO(1)		2			2		2	2			2	
SPO(2)		2	2		2		2	2	2		2	
rMSSD	1,2						1					
LFHF	1,2											
RespRate	3											

Significant Relationships: 5/7/2020 - 6/24/2020												
	Pearson						Spearman					
	Mean	Med	Max	Min	STD	Freq	Mean	Med	Max	Min	STD	Freq
rrInt					1						1	
exLen	1,3,4						*All*			*All*	1,2,3	
inLen	1,3,4						1,2,4		*All*	1,2,3	1,2,3	
resp						3	1,3		1,2,4			1,2,3
GSRton					*All*						*All*	
RedROC(+)								4				
IrROC(+)		2										
rMSSD	2											

Table 6.3: Difference in significant relationships between mental stress self-tests taken between 4/6/2020 through 5/7/2020 and those taken between 5/7/2020 and 6/24/2020. 1:PSS, 2:Ardell, 3:Symptoms, 4:SCRI. See Glossary for feature definitions. Note only four features correlate with all four tests after 5/7, and no features correlate with all tests prior to 5/7.

tions are associated with PSS, ten with Ardell Wellness, ten with Symptoms, and seven with SCRI. These relationships can be seen in Table 6.3. Abbreviated feature names present in the table are defined in the Glossary. In this set of data collections four features have a Spearman correlation with all four mental stress self-tests: mean exhalation length, minimum exhalation length, maximum inhalation length, and standard deviation of tonic GSR.

In the period between 5/7 and 6/24 fewer features reveal significant correlations with stress than in the previous time span. Specifically, the respiration amplitude, skin temperature, blood oxygen saturation, LFHF, and respiration rate features no longer reveal correlations to stress values. It is not entirely clear why 5/7 marks such a clear dividing line in the correlation between these features and self-estimated stress.

6.2 Machine Learning Approach

All features passed into machine learning algorithms were normalized by calculating the Z-score of the feature vector. This measure is a common and simple method of data normalization, and indicates how many standard deviations the sample is from the mean of that particular feature. The formula for calculating the Z-score is

$$Z = \frac{\chi - \mu}{\sigma} \tag{6.3}$$

where χ is the individual value being transformed, μ is the mean of the data vector, and σ is the vector's standard deviation value.

70 data collections were used as training data, with 10 data collections excluded to

be used exclusively to test model performance. The 10 held-out data collections were selected by sorting all 80 datasets from least to greatest self-evaluation score for each of the four test methods and removing one collection at each 10th percentile of the score distribution. As a result, models for each self-evaluation method were trained and tested on four different permutations of the 80 data collections. 7-fold cross validation was used with the training data collections to prevent overfitting of the machine learning model to the data, as well as to identify and tune hyperparameters which aptly characterized the entire dataset. Model performance was evaluated based on the root mean square error (RMSE) of the algorithm in cases of regression models and on accuracy, specificity, and sensitivity in the case of classification models.

The MATLAB applications *RegressionLearner* and $Classification Learner$ were utilized to create and evaluate the machine learning models described in this chapter. After generation and training in the application, the best performing regression model and classification model for each evaluation method were exported to the workspace and saved in order to be applied to the held-out test data.

6.3 Machine Learning Algorithms

In this chapter, several supervised machine learning algorithm types are mentioned with little associated detail. This section serves to elaborate on each of the algorithms mentioned in the sections below:

1. *Decision Tree:* Decision Tree models can be applied to both regression and classification predictive model problems and are not constrained by the data type of the dependent variables. In this approach, the entire dataset is first grouped

into one population, called the Root Node. The population is subsequently split (or 'branched') into smaller homogeneous groups based on features within the data. These groups may either be divided again in which case they are called decision nodes, or remain as they are and be denoted as terminal nodes or leaves.

These types of models are advantageous in that they are non-parametric and require little data cleaning, but they can easily overfit the model to the training data. This tendency can be mitigated by placing constraints on the model such as a limit on leaf size or by restricting the number of features considered for each branch. In this work, the 'medium' and 'course' designations of tree models refer to these constraints. The medium tree restricts the total number of splits to 20 or fewer, while a course tree restricts the maximum number of splits to four or fewer. A 'bagged' tree implements Breiman's random-forest algorithm by training multiple trees on random subsets of data with replacement to generate multiple tree models from the same data.

2. *Gaussian Process Regression:* Gaussian Process Regression models are, as the name suggests, intended for regression problems. These models are nonparametric Bayesian approaches to machine learning which are able to provide uncertainty measurements on their predictions. Gaussian processes establish a probability based on Bayes' rule that a given observation will fall into a certain value based on previous observations. 'Prior distribution' of previous observations is denoted $p(w)$ based on parameter w. The 'posterior distribution' is then calculated based on both the current dataset and the prior distribution using

Bayes' Rule:

$$p(w|y, X) = \frac{p(y|X, w) * p(w)}{p(y|X)} \tag{6.4}$$

Finally, the 'predictive distribution' can be calculated to determine the probability that current observations are associated with a specific value. This is performed by weighing all possible predictions by their posterior distributions:

$$p(f^*|x^*, y, X) = \int_w p(f^*|x^*, w) * p(w|y, X)dw \tag{6.5}$$

where f^* is the prediction label and x^* is the current observation.

3. *Support Vector Machine:* Support Vector Machines (SVM) are classification models in which data points are plotted in an n-dimensional space and a plane is plotted in that space which separates each class. This division is placed in the n-dimensional space such that the distance from each point to the line is maximized. The plane is known as a hyperplane. Support vectors are the data points on each side of the hyperplane closest to the division.

   Some data will not allow for a hyperplane to cleanly separate all classes, in which case a 'soft' hyperplane may be applied. Modifying terms such as 'gaussian', 'linear', and 'cubic' refer to the kernel function utilized in the classifier to fit a soft hyperplane to the data. Kernels are functions describing the hyperplane in the n-dimensional data space which transform the hyperplane into a 'hypersurface.' Hypersurface based SVM models allow for nonlinear classification of data.

4. *KNN:* The K-Nearest-Neighbors algorithm can be used in both regression and

classification problems. In this approach training data is used to store available classes multi-dimensionally based on their features, and new data is sorted based on the distance of its features from those of the existing data. This approach is computationally expensive due to the comparison of feature distance, but is simple to implement.

The designations employed in this work such as 'cosine' and 'weighted' refer to the distance metric used to sort new data and the cost of further distance to the performance evaluation. Altering the distance metric examined when sorting data may allocate new data to a different class despite utilizing the same training dataset. The cosine distance metric is calculated as follows:

$$d = 1 - \frac{x * y'}{\sqrt{(x * x')(y * y')}} \tag{6.6}$$

where x and y denote features in the training data and x' and y' denote features from the test data. In a weighted KNN, all neighbors are assigned a weight attribute based on their distance to the current sample. Closer neighbors contribute more heavily to the decision outcome.

Material in this section was sourced from [96].

6.4 Machine Learning: Regression

The MATLAB *RegressionLearner* application was used to train 19 different machine learning algorithms to predict self-reported stress values from the data features. These same models were also trained without normalizing the input features,

Self Test	Model	RMSE	PCA?
Cohen PSS	Medium Tree	3.07	No
Ardell Wellness	Gaussian Process Regression	8.72	Yes
Symptoms	Gaussian SVM	5.23	No
SCRI	Bagged Trees	0.19	No

Table 6.4: Best performing regression model outcomes of machine learning model training using RegressionLearner.

but this approach invariably returned a higher RMSE across all 76 models (19 per each of the four methods). Principal component analysis (PCA) was also investigated as an approach to reduce the dimensionality of the machine learning models. Two approaches were employed: the first only utilized features which were found to have a significant Spearman correlation to the self-test outcomes, while the second utilized *RegressionLearner's* automated PCA feature selector. In the correlation based approach all models performed significantly worse than when considering all data features. The use of automated PCA to train predictive models did not reduce the RMSE of the highest performing model for PSS, Symptoms, or SCRI, but did reduce RMSE of the Ardell Wellness model from 9.39 to 8.72. The results of the *RegressionLearner* training can be seen in Table 6.4.

Regression models for PSS, Symptoms of Stress, and SCRI performed relatively well when applied to the held-out data with mean error rates of 26.45%, 17.13%, and 7.56% respectively. Median values for all three models are very close in value to the mean as well: deviating by 1% or less in all three cases. The regression model for the Ardell Wellness test returned inferior performance by comparison. While the mean accuracy error was only 1.88%, the median error was greater than 50% and the standard deviation of percent error was almost 300%. The minimum percent error witnessed in the ten held-out data predictions was incorrect by an incredible 485%.

	Mean	Median	Max	Min	Range	STD
Cohen PSS	26.45%	24.95%	60%	1.35%	58.65%	17.46%
Ardell Wellness	1.88%	50.44%	636.51%	484.87%	1120.39%	293.00%
Symptoms	17.13%	16.13%	40.51%	1.21%	39.3%	12.37%
SCRI	7.56%	6.81%	16.25%	0.47%	15.78%	5.94%

Table 6.5: Percent error of predictive model results on hold out data

The results of the trained models on the held-out data are shown in Table 6.5. The percentages in this table relate to the percent error of the trained model predictions compared to the true stress value associated with each of the ten test cases. Percent error was calculated as

$$Error = (\frac{|Predicted| - |True|}{|True|}) * 100 \tag{6.7}$$

6.5 Machine Learning: Classification

As noted in Section 2.3, three of the four self evaluation methods examined in this work are categorized not only by a continuous scale, but also by categorical classification. In the case of the Cohen PSS, scores between 0 and 12 are classified "Low Stress," 13 to 24 is considered "Moderate Stress," and 25 to 40 is considered "High Stress." In the case of the Ardell Wellness test, six categories exist corresponding to score ranges -75 to -51, -50 to -25, -24 to 0, 1 to 24, 25 to 50, and 51 to 70. These categories indicate a range of coping abilty from nonexistent stress coping skills to superior stress coping skills. The Stress Coping Resources Inventory distributes scores into four categories. Scores less than 1.4 suggest less than average stress coping ability. Scores between 1.5 and 2.4 are considered average stress coping. Values between

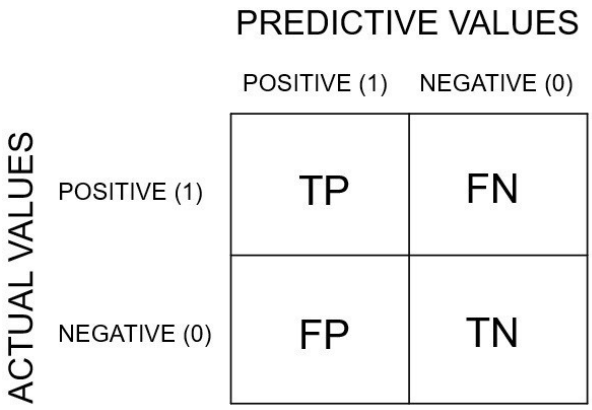

Figure 6.3: A sample Confusion Matrix indicating True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) regions of sample classification. Reproduced from [97].

2.5 and 3.4 are attributed to above average stress coping, and scores above 3.5 are considered indicative of superior stress coping. The Symptoms of Stress Frequency self evaluation does not feature categorical classification, and is not considered in this section.

Each data collection was assigned a categorical value corresponding to each the continuous scores obtained on the Cohen PSS, the Ardell Wellness, and the SCRI tests. These categorical values ranged from 1 to 3, 1 to 6, and 1 to 4 respectively. MATLAB's *ClassificationLearner* application was used to train 19 different machine learning classification models and identify the algorithm with the highest performance.

Performance of each model was determined by three parameters: Accuracy, Specificity, and Sensitivity. These parameters were determined from the confusion matrix, which visually represents all model predictions as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positives occur when the model successfully categorizes a sample as a member of the 'positive' class, while

Test	Model	Accuracy	Specificity	Sensitivity
PSS	Logistic Regression	74%	69%	76%
Ardell	Cubic SVM	79%	94%	6%
SCRI	Medium KNN	82%	N/A	81%

Table 6.6: Results of machine learning classification model training on WIHMS dataset

True Negatives occur when the model successfully categorizes a sample as a member of the 'negative' class. False positives and negatives occur when the model incorrectly assigns a sample to either the positive or negative class. Figure 6.3 shows a sample confusion matrix with each of these regions indicated.

The Accuracy metric is a measure of how many predictions the model has correctly classified as a member of the positive class. Accuracy is simply calculated by

$$Accuracy = \frac{TP}{TP + TN + FP + FN} \tag{6.8}$$

Specificity, also known as the True Negative Rate, is a metric for the number of True Negatives which are identified correctly. The formula for Specificity is

$$Specificity = \frac{TN}{TN + FP} \tag{6.9}$$

Sensitivity, the third performance metric, is also called the True Positive Rate. It is a metric to identify the proportion of correctly identified positive samples compared to the incorrectly identified negative samples. The Sensitivity formula is

$$Sensitivity = \frac{TP}{TP + FN} \tag{6.10}$$

The most successful classification models for the PSS, Ardell Wellness, and SCRI

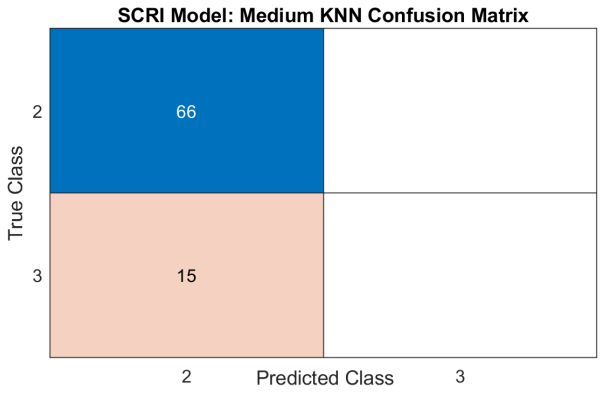

Figure 6.4: Confusion matrix of Medium KNN model predictions for the SCRI stress values. The matrix reveals 66 True Positives (Top Left), zero False Positives (Top Right), 15 False Negatives (Bottom Left), and zero True Negatives (Bottom Right).

tests were Logistic Regression, Cubic SVM, and Medium KNN respectively. The calculated Accuracy, Specificity, and Sensitivity parameters for each of these models can be seen in Table 6.6, but the results were initially less than encouraging. The highest Accuracy of any of the models was 82% in the case of the SCRI. However, further inspection reveals that this accuracy was accomplished by predicting *all samples* belonged to Class Two and *none* were assigned to Class Three. The confusion matrix of the Medium KNN classifier can be seen in Figure 6.4.

It was noted in this evaluation that many classes were not represented in the dataset, and some classes were drastically over-represented. Of the 80 data collections, there were 47 'Low Stress' PSS results and 32 'Moderate Stress' results. The 'High Stress' class was not represented. The Ardell results showed 18 Class Three results (scores -25 to -1), 59 Class Four results (scores 0 to 25), and three Class Five results (26 to 50). There were no instances of Classes One, Two, or Six. In the case of the SCRI test, there were 65 'Average Stresscoper' results and 15 'Above Average Stresscoper' results, leaving the 'Below Average Stresscoper' and 'Superior Stresscoper' classes

unrepresented.

For the under-represented classes, a Synthetic Minority Oversampling Technique (SMOTE) algorithm was employed to produce a roughly even number of samples across all represented classes of each of the three self-evaluation tests. In SMOTE, synthetic samples are created by interpolation in the dimensional space between existing samples of a minority class to increase the representation of that class as compared to a distinct majority class. The algorithm relies on a specified number of nearest-neighbors in the minority class to create new samples, and can generate any number of desired synthetic samples. As the number of synthetic samples increases relative to the existing samples however, each new sample is less and less diverse than the synthetic sample generated before it [98].

SMOTE was applied to both Classes One and Two of the PSS, Classes Three and Five of the Ardell Wellness test, and to Class Three of the SCRI. PSS Class Three, Ardell Wellness Classes One, Two, and Six, and SCRI Classes One and Four remained unrepresented. The extent of synthetic minority oversampling on the size of the dataset can be seen in Table 6.7.

With the larger synthetically oversized dataset, *ClassificationLearner* was again employed but with 8-fold cross validation (for greater redundancy to avoid overfitting) and with again with ten collections held-out. This resulted in a training data collection size of 185 for the PSS, 174 for the Ardell Wellness test, and 176 for the SCRI. The larger overall size of the dataset enabled a greater number of data collections to be used for model training and showed improvement based on performance calculations.

After the application of SMOTE to the dataset, classification performance *drasti-*

Self-Test	Class	Pre-SMOTE	Post-SMOTE	% Increase
PSS	1	48	96	100%
	2	33	99	200%
	Total	81	195	142%
Ardell	3	18	54	200%
	4	60	60	0%
	5	3	60	1900%
	Total	81	174	115%
SCRI	2	66	66	0%
	3	15	60	300%
	Total	81	186	130%

Table 6.7: Number of representations of each class before and after SMOTE algorithm implementation

Test	Model	Accuracy	Specificity	Sensitivity
PSS	KNN	100%	100%	100%
Ardell	SVM	97%	100%	92%
SCRI	SVM	92%	100%	85%

Table 6.8: Machine learning model performance on test data after SMOTE application to training data

cally improved across the three self-evaluation methods as can be seen in Table 6.8. The best performing classification type of algorithm also changed for the PSS and SCRI. The highest performance model for the PSS was initially a Logistic Regression algorithm with 74% Accuracy, 69% Specificity, and 76% Sensitivity, but after SMOTE was a K-Nearest-Neighbor algorithm with 100% Accuracy, Specificity, and Sensitivity. This algorithm correctly predicted the self-evaluated stress level of every single data collection in the test set. In the training set, there were only 6 False Positive predictions and zero False Negative predictions. The SCRI algorithm transitioned from a medium k-Nearest-Neighbor algorithm with 82% Accuracy and 81% Sensitivity (despite abysmal true performance) to a Support Vector Machine with 92% Accuracy, 100% Specificity, and 85% Sensitivity when applied to the held-out data. On the training dataset, there was only one False Positive and three False Negatives.

The greatest performance improvement was that of the Ardell Wellness classifier. Both before and after SMOTE the Ardell Wellness was classified by a Support Vector Machine, but performance increased from 79% to 92% Accuracy, 94% to 100% Specificity, and 6% to 92% Sensitivity when applied to held-out data. On the training data, three collections belonging to Class Four were mistakenly attributed to Class Three, and five collections belonging to Class Three were mistakenly attributed to Class Four. All other predictions on the training data were correct.

6.6 Conclusion

There are other machine learning methods that may have returned more successful results without the application of the SMOTE algorithm, but the limited size (n = 80) of the WIHMS Stress dataset excluded these methods from practical application. Specifically, deep learning techniques can be trained to automatically detect features of interest from a training dataset, and then perform classification in real time based on these features [99]. A significant and reoccurring consideration during this work was regarding the selection of features extracted from the data and ensuring that they accurately characterized the overall data collection. Implementing automatic feature selection could produce more accurate predictive models. Deep learning methods are typically applied to image processing applications but could theoretically be applied to raw WIHMS datasets, where each set of raw data from a collection occupies a 129,981 x 9 matrix after 200 Hz resampling. Unfortunately, neural networks are typically trained on databases ranging in size from thousands to millions of images. MATLAB specifically offers access to 18 pre-trained Deep Neural Networks, many of which are trained on the Image-Net database, which contains over 14 million images as of June 2020 [100]. Even to obtain a comparatively small set of 10,000 collections for training a deep learning neural network would take nine years of data collection from a single subject generating three collections per day. The sliding window technique used to extract features from each data collection could be used to consider every window as an individual stress identification for the purposes of expanding a training database, but might confound the neural network in cases where windows reflect events relating to acute stress reactions or movement artifacts. This limitation was

successfully mitigated in the WIHMS effort by applying the SMOTE algorithm to expand the number of under-represented classes in the classification approach. On a larger scale, this approach may not be viable as the algorithm does not consider the possibility that neighboring samples may belong to other classes. Applied with the intent of creating a dataset of appropriate size for a neural network, class divisions may become unclear and hinder classification efforts.

More limiting than the size of the dataset is the inherent subjectivity of the self reporting method of stress estimation. This is actually a significant takeaway from the work presented in this thesis, and was part of the motivation for the work to begin with. Subjects in a negative state of mind are more likely to overemphasize negative responses, while subjects in a positive state of mind are more likely to answer more positively regardless of current stress level due to confirmation bias. The consequence of this bias is that self-perceived stress may not be reported in a consistent method, which complicates models of perceived mental stress.

The specific features of interest extracted from the dataset may have also complicated modeling attempts. Chapter 2.4 identified features from previous work that have been demonstrated to indicate subject stress. Many of these features were found to correlate poorly with self-perceived stress values in this work. Not only does this result promote the importance of subject-specific models of stress estimation, but also suggests that other, less frequently cited or more recently discovered physiological features may hold insight to mental stress levels. Newly innovated technologies, such as wearable hormonal and chemical monitoring, may hold provide the missing link between mental stress and wearable sensors.

On a more positive note, the application of the SMOTE algorithm allowed near-

perfect classification of self-evaluated stress levels for the PSS, Ardell Wellness, and SCRI stress tests. These three subject-specific models performed with accuracies of 100%, 97%, and 92% respectively when tested on the held-out data. Specificity was 100% with all three models and the lowest sensitivity level of the three models was 84% in the case of the SCRI classifier.

Overall, the results of the work in this chapter validate the WIHMS platform for the purposes of monitoring and estimation of self-perceived mental stress levels. In only ten and a half minutes of data collection, the garment is able to capture physiological signals which predict mental stress levels with high accuracy when combined with subject specific models! This outcome shows promise in the ability to monitor subject or patient stress levels during clinical monitoring events.

In the next chapter, remaining areas of improvement are identified for the WIHMS platform and the three specific contributions outlined in Section 1.3 are revisited. Some parting words are also offered from a point of reflection upon this work.

Chapter 7

Conclusions and Future Work

The work described in this thesis aimed to prototype a novel modular and open-source wearable garment, and apply the device to the develop predictive algorithms for mental stress detection. WIHMS, as developed at the time of writing, is capable of measuring five digital sensors and four analog sensors at a rate of 200 Hz. The three contributions of this work proposed in Section 1.3 have been accomplished as follows:

7.1 Contributions

1. *An Open Source and Modular Health Monitoring Hardware Solution*

   Chapter 4 described the development of the Wearable Integrated Health Monitoring System (WIHMS). Design resources for all custom components of the system, as well as source code for the data collection and sensor interaction, are located in a publicly accessible GitHub repository [60]. These resources are offered under the MIT Open Source License and may be replicated or distributed

under the conditions of that license.

2. *A Public Dataset of Self-Estimated Stress Level and Measured Biosignals*
   80 data collections are available for public use at [60]. As noted in Section 5.1, this data was collected using the hardware developed in this work, and is presented publicly for further analysis. This dataset will be expanded upon in the coming months as restrictions related to the COVID-19 pandemic lift in order to examine the effects on stress variability as daily life slowly returns to its pre-pandemic state.

3. *Machine Learning Models for Prediction of Self-Perceived Mental Stress Levels*
   As described in Chapter 6, multiple models were generated to predict the self-estimated stress level associated with each data collection. These models varied in accuracy from 75% to 100%, where models based on Decision Tree regression algorithms and Support Vector Machine classification algorithms performed better than other approaches examined. Moreover, the results of this investigation demonstrate that from the data examined, the SCRI method of stress estimation is more easily predicted than other methods examined.

7.2 Future Work

The WIHMS platform successfully fulfills its intended row as an open source, modular, inexpensive wearable sensing device. However, like any ongoing hardware effort, several areas of improvement have been identified for future iterations of the system.

7.2.1 Integrated Data Processing

The core of the WIHMS platform is inarguably the combined unit consisting of the data aggregation/consolidation board and the Arduino Mega 2560. To date, the separation of these individual components has not proven to be an issue, but does increase cost and complexity of the system. If the Atmel ATmega2560 chip at the core of the Arduino were to be incorporated into the design of the data board, the design would require two fewer 8P8C connectors and cables. I2C and SPI lines currently running between the data board and the Arduino could be eliminated, reducing the failure risk of each bus inherently assumed by increasing the number of hardware interconnections.

7.2.2 Sensor Alterations

The first iteration of the sensing array proved functional, but can absolutely be improved upon. The Bosch BNO055 absolute orientation sensor performed as intended, but includes onboard fusion and auto-calibration algorithms which slow the output data rate. The unit also includes a 3-axis magnetometer for performing compass and heading measurements, which is unnecessary in this application and increases the overall cost of the system. The STMicroelectronics L3DG20H IMU module offers an accelerometer and gyroscope of the same precision and board size as the BNO055, but with lower power consumption, more options for output data rate, and a retail price several dollars cheaper than the current unit.

The custom GSR sensor would also benefit from a simple redesign of the sensor pads which contact the skin. It was noted in [36] that rigid sensors often have difficulty

maintaining contact between the subject and the sensing surface, and the device used in this work was no exception. Replacing both rigid pads of the custom sensor with a single 3.5 mm stereo jack would allow the sensor to be drastically reduced in size, and would allow the use of commercially available 2-lead snap electrode leads. Through this modification, the same type of 3M Red Dot electrodes utilized for ECG collection could also be used for reliable GSR collection without contact-loss events.

A final, more long term revision goal for the sensor platform is the transition to textile electrodes to eliminate the need for adhesive electrode pads altogether. Washable, reusable textile electrodes were intended to be sewn into the very fabric of the WIHMS garment from the inception of the project. In prior work, silver plated nylon thread was found to be the most reliable woven textile ECG electrode material in terms of signal quality, but quickly suffered from corrosion and signal degradation after several uses [101]. Other work has investigated the application of a PEDOT: PSS conductive polymer to conventional textiles as noted in Section 3.2 to yield a washable, reusable electrode which maintained its conductivity for up to 50 wash cycles. A PEDOT: PSS application was a favored approach to ECG monitoring in the WIHMS platform, but at the time of garment construction the price of PEDOT: PSS was considered outside of the budget of the project. As of June 1, 2020, a 25 gram quantity of PEDOT: PSS can be purchased for $161.

7.2.3 Motion Artifact Ruggedization

For the purposes of hardware validation and consistency between the 80 data collections that make up the dataset of *Contribution 2*, the WIHMS garment was used to collect data while the subject intentionally remained motionless. In order to be

considered a truly noninvasive sensing device, the WIHMS garment must be capable of being worn during daily activity. In the current sensor configuration, the lack of a reference IMU to baseline against the gyroscope used for respiration results in a measurement that is meaningless any time the subject is in motion. The ECG sensor is also susceptible to noise from muscular activity in the pectoral muscles because of the placement of the positive and negative leads. An EMG sensor could be introduced to subtract a quantitative baseline from the ECG data, or the positive and negative leads could be moved to a region isolated from muscle activity. The GSR sensor also suffers contact-loss events exacerbated by movement. The sensor modifications outlined above in Section 7.2.2 would aleviate this concern.

Software solutions also exist for motion artifact ruggedization and removal. In recent years, adaptive filtering techniques have been developed specifically for use in ECG and PPG signals in wearable systems [103, 104]. These techniques can be readily applied to WIHMS sensor data, but are not by themselves a substitute for hardware ruggedization and informed sensor placement. A combination of hardware and software solutions to motion artifact ruggedization and remediation should advance the WIHMS platform towards unrestrained use during activity.

7.2.4 Wireless Data Transfer

In the modern age of networked devices and expanding Internet of Things (IoT) architecture, the manual data transfer method employed by the WIHMS platform is antiquated at best. Each time data is to be exported from the system, the μSD card must be manually ejected from the SD carrier on the data aggregation board, placed in a μSD to SD adapter, and manually copied to the host PC. Many solutions exist

to incorporate Bluetooth Low Energy (BLE) or WiFi capability to Arduino-based designs. The host PC could easily be configured to act as a basestation to receive a wireless data transmission from WIHMS at the conclusion of the data collection or at the press of a button. Wireless data transfer over either BLE of WiFi is a commonly desired feature in hobbyist designs, and as a result is supported widely on public forums [102].

7.2.5 IRB Approval and Extended Dataset

The efforts in this thesis produced subject-specific models of mental stress as a direct consequence of the n = 1 sample size. By obtaining Institutional Review Board (IRB) approval, the WIHMS garment could be deployed to collect physiological data from a larger cohort of subjects. IRB approval was not sought during this research effort in favor of accelerating the development of the WIHMS hardware and its deployment for data collection by means of self-experimentation. A larger study might reveal a universal model for mental stress prediction, rather than the subject-specific models developed during earlier chapters.

7.3 Final Words

The work presented in this thesis serves as a novel bridge between several focus areas of science and medicine. By utilizing open source software and hardware design in a wearable sensing application, it is hoped that other researchers may begin to embrace the open source mindset in their research. To our knowledge, this work also resulted in the first mid/long term data collection effort linking multiple physiological

parameters and self perceived stress levels. Mental stress affects every human being, potentially in a negative way. By making efforts to track and understand how mental stress manifests itself, researchers may begin to understand universal methods of stress mitigation based on physiological manifestations of stress.

It is suspected that the WIHMS dataset creation, like many aspects of daily life in 2020, was affected by the COVID-19 pandemic. When the official stay-at-home orders were implemented, the variance of stress levels in all four methods of self-reported stress estimation fell sharply. It appears that quarantine and isolation contributed to a more consistent level of mental stress. The lack of novelty in the daily routine during the pandemic lockdown restricted the influence of outside influences on self-perceived stress. Prior to the stay-at-home order, changes in daily routine such as upcoming deadlines or traffic congestion during a commute could contribute to stress levels but were no longer a factor during lockdown. These unprecedented global events may have confounded the training of machine learning models by altering stress manifestations midway through the dataset. However, while these events may have altered the intended narrative of this thesis, they do offer an interesting insight on the effect of isolation of self-perceived stress.

As the world gradually returns to normal over the coming months, a limited number of further data collections will be added to the dataset to provide a more rounded and complete insight into the self-perceived mental stress levels before, during, and after a worldwide health crisis. While the WIHMS Stress dataset did not fully serve to prove our intended hypothesis, it may still serve to paint a picture of daily life during this time in not only American, but Global history.

The future holds much uncertainty. In the coming weeks, months, and years, we

will all face more trials and tribulations, but the future is still there, waiting for us to shape it how we will.

"In three words, I can sum up everything I have learned about life: It Goes On."

- Robert Frost

Bibliography

[1] W.R. Hawkins and J.F. Zieglschmid, SP-368 Biomedical Results of Apollo, chapter 2, NASA, 1995.

[2] S. Rawat, S. Vats, and P. Kumar, "Evaluating and Exploring the MYO Armband," in 2016 International Conference System Modeling & Advancement in Research Trends (SMART). IEEE, 2016, pp.115-120.

[3] J. Parak and I. Korhonen, "Evaluation of Wearable Consumer Heart Rate Monitors Based on Photopletysmography," in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2014, pp. 3670-3673.

[4] S. Patel, H. Park, P. Bonato, et al. A Review of Wearable Sensors and Systems with Application in Rehabilitation. Journal of NeuroEngineering Rehabilitation 9, 21 (2012). https://doi.org/10.1186/1743-0003-9-21

[5] "Stress Synonyms, Stress Antonyms," Merriam-Webster. [Online]. Available: https://www.merriam-webster.com/thesaurus/stress. [Accessed: 31-May-2020].

[6] "Chronic Stress Puts Your Health at Risk," Mayo Clinic, 19-Mar-2019. [Online]. Available: https://www.mayoclinic.org/healthy-lifestyle/stress-management/in-depth/stress/art-20046037. [Accessed: 01-Mar-2020].

[7] L. Murphy, D. Dubois, and J. Hurrell, "Accident Reduction Through Stress Management," Journal of Safety Research, vol. 18, no. 1, p. 46, 1987.

[8] K. Turgeman-Lupo and M. Biron, "Make it to Work (and Back Home) Safely: the Effect of Psychological Work Stressors on Employee Behaviour While Commuting by Car," European Journal of Work and Organizational Psychology, vol. 26, no. 2, pp. 161-170, Dec. 2016.

[9] R. E. Bowman, K. D. Beck, and V. N. Luine, "Chronic Stress Effects on Memory: Sex Differences in Performance and Monoaminergic Activity," Hormones and Behavior, vol. 43, no. 1, pp. 48-59, 2003.

[10] T. W. Colligan and E. M. Higgins, "Workplace Stress," Journal of Workplace Behavioral Health, vol. 21, no. 2, pp. 89-97, 2006.

[11] "Workplace Stress," The American Institute of Stress, 20-Feb-2020. [Online]. Available: https://www.stress.org/workplace-stress. [Accessed: 01-Mar-2020].

[12] S. Cohen, T. Kamarck, and R. Mermelstein, "A Global Measure of Perceived Stress," Journal of Health and Social Behavior, vol. 24, no. 4, p. 385, 1983.

[13] "Stress Assessments," New York State United Teachers, 2013. [Online]. Available: https://www.nysut.org/ /media/files/nysut/resources/2013/april/social-services/socialservices_stressassessments.pdf?la=e. [Accessed: 01-Feb-2020].

[14] G. Holtmann, R. Kriebel, and M. V. Singer, "Mental Stress and Gastric Acid Secretion," Digestive Diseases and Sciences, vol. 35, no. 8, pp. 998-1007, 1990.

[15] Y. Takemura, S. Kikuchi, and Y. Inaba, "Does Psychological Stress Improve Physical Performance?," The Tohoku Journal of Experimental Medicine, vol. 187, no. 2, pp. 111-120, 1999.

[16] B. Marr, "How Successful People Deal With Stress," Entrepreneur, 30-Sep-2014. [Online]. Available: https://www.entrepreneur.com/article/237861. [Accessed: 10-Jun-2020].

[17] V. L. Burrows, "The Medicalization of Stress: Hans Selye and the Transformation of the Postwar Medical Marketplace," Ph.D Dissertation, College of History, City University of New York, 2015. Accessed on: June 15, 2020. [Online]. Available: academicworks.cuny.edu/cgi/viewcontent.cgi?article=1891&context=gc_etds.

[18] R. Viner, "Putting Stress in Life: Hans Selye and the Making of Stress Theory," Social Studies of Science, vol. 29, no. 3, p. 391-410, 1999.

[19] A. Y. Shalev, R. Yehuda, and A. C. McFarlane, International handbook of human response to trauma. New York, NY: Kluwer Academic/Plenum, 2000.

[20] R. S. Lazarus and S. Folkman, Stress, appraisal, and coping. New York, NY: Springer, 1984.

[21] N.Stamp, "The Revolutionary Origins of Self-Care," locallove, 10-Dec-2019. [Online]. Available: https://locallove.ca/issues/the-revolutionary-origins-of-self-care/. [Accessed: 09-Jun-2020].

[22] A. Lorde, A burst of Light: essays. Ithaca, NY: Firebrand Books, 1988.

[23] "2015 Stress in America," American Psychological Association. [Online]. Available: https://www.apa.org/news/press/releases/stress/2015/snapshot. [Accessed: 17-Jun-2020].

[24] T. H. Holmes and R. H. Rahe, (1967). The social readjustment rating scale. Journal of Psychosomatic Research, 11(2), 213-218. doi:10.1016/0022-3999(67)90010-4

[25] P. F. Lovibond and S. H. Lovibond, "The Structure of Negative Emotional States: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories," March 1995. Behaviour Research and Therapy, vol. 33, no. 3, p. 335-343. doi:10.1016/0005-7967(94)00075-U

[26] D. B. Ardell and J. G. Langdon, Wellness: the Body, Mind, and Spirit. Dubuque, IA: Kendall/Hunt Pub. Co., 1989.

[27] K. B. Matheny and C. J. McCarthy, Write your own prescription for stress. Oakland, CA: New Harbinger Publications, 2000.

[28] "QRS complex" Wikipedia, 13-May-2020. [Online]. Available: https://en.wikipedia.org/wiki/QRS_complex. [Accessed: 18-Jun-2020].

[29] S. N. H. Malek, K. Chellappan, and R. Jaafar, "Short Review of Electrocardiogram (ECG) Technique Versus Optical Techniques for Monitoring Vascular Health," IFMBE Proceedings International Conference for Innovation in Biomedical Engineering and Life Sciences, pp. 222-225, 2015.

[30] F. Shaffer and J.P. Ginsberg, "An Overview of Heartrate Variability Metrics and Norms," Frontiers in Public Health, vol. 5, pp. 258, 2017.

[31] H. G. Kim, E. J. Cheon, D. S. Bai, Y. H. Lee, and B. H. Koo, "Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature," Psychiatry Investigation, vol. 15, no. 3, pp. 235-245, 2018.

[32] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," 2011 Computing

[33] R. Sahoo and S. Sethi, "Functional Analysis of Mental Stress Based on Physiological Data of GSR Sensor," Advances in Intelligent Systems and Computing Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1, pp. 109-117, 2015.

[34] J. Synnott, D. Dietzel, and M. Ioannou, "A review of the polygraph: history, methodology and current status," Crime Psychology Review, vol. 1, no. 1, pp. 59-83, 2015.

[35] "Publication Recommendations for Electrodermal Measurements," Psychophysiology, vol. 49, no. 8, pp. 1017-1034, 2012.

[36] H. F. Posada-Quintero and K. H. Chon, "Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review," Sensors, vol. 20, no. 2, p. 479, 2020.

[37] H. Kataoka, H. Kano, H. Yoshida, A. Saijo, M. Yasuda, and M. Osumi, "Development of a Skin Temperature Measuring System for Non-Contact Stress Evaluation," Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond, 1998.

[38] K. A. Herborn, J. L. Graves, P. Jerem, N. P. Evans, R. Nager, D. J. Mccafferty, and D. E. Mckeegan, "Skin temperature reveals the intensity of acute stress," Physiology & Behavior, vol. 152, pp. 225-230, 2015.

[39] "The oldest watch in the world," Your Watch Hub, 12-Dec-2017. [Online]. Available: https://www.yourwatchhub.com/watches/oldest-watch-in-the-world/. [Accessed: 13-Jun-2020].

[40] Howard, Alexander, "Hearing Aids: Smaller and Smarter." New York Times, November 26, 1998.

[41] Mills, Mara. "Hearing Aids and the History of Electronics Miniaturization." IEEE Annals of the History of Computing 33.2 (2011): 24-44.

[42] S. Prasanth, "Fitness Tracking and Recognition in Smartwatches," Muse Wearables, 21-Jan-2020. [Online]. Available: https://us.musewearables.com/blogs/news/fitness-tracking-and-recognition-in-smartwatches. [Accessed: 16-Jun-2020].

[43] "Polar's 40 Years Of Incredible Firsts," Polar Blog, 31-Mar-2020. [Online]. Available: https://www.polar.com/blog/40-years-of-incredible-firsts-polar-history/. [Accessed: 09-Jun-2020].

[44] "The story of Fitbit: How a wooden box was bought by Google for $2.1bn," Wareable, 04-Apr-2020. [Online]. Available: https://www.wareable.com/fitbit/story-of-fitbit-7936. [Accessed: 01-Jun-2020].

[45] B. Stephenson, "7 Apps for Checking Your Heart Rate Using Just Your Smartphone," Lifewire, 21-Apr-2020. [Online]. Available: https://www.lifewire.com/best-heart-rate-apps-4580412. [Accessed: 01-Jun-2020].

[46] "Wearable Healthcare Technology and Devices," MC10. [Online]. Available: https://www.mc10inc.com/. [Accessed: 16-Jun-2020].

[47] Seema Rawat, Somya Vats, and Praveen Kumar, "Evaluating and exploring the MYO ARMBAND," in 2016 International Conference System Modeling & Advancement in Research Trends (SMART). IEEE, 2016, pp.115-120.

[48] DPA, "Controversy grows over Google's Glass project," BusinessLine, 12-Mar-2018. [Online]. Available: https://www.thehindubusinessline.com/info-tech/controversy-grows-over-googles-glass-project/article20595712.ece. [Accessed: 10-Jun-2020].

[49] "Wearable Technology," Myontec. [Online]. Available: https://www.myontec.com/. [Accessed: 16-Jun-2020].

[50] "Athos Coach Collection," Athos. [Online]. Available: https://shop.liveathos.com/collections/coaching-collection. [Accessed: 15-Jun-2020].

[51] M. Patterson, D. Mcgrath, and B. Caulfield, "Using a Tri-axial Accelerometer to Detect Technique Breakdown Due to Fatigue in Distance Runners: A Preliminary Perspective," 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011.

[52] A. Ankhili, X. Tao, C. Cochrane, D. Coulon, and V. Koncar, "Washable and Reliable Textile Electrodes Embedded into Underwear Fabric for Electrocardiography (ECG) Monitoring," Materials, vol. 11, no. 2, p. 256, 2018.

[53] "Mocap Resources: Optical Motion Capture Systems," Optical Motion Capture Systems. [Online]. Available: https://metamotion.com/motion-capture/optical-motion-capture-1.htm. [Accessed: 1-Jun-2020].

[54] L.J. Abadie, C.W. Lloyd, and M.J. Shelhamer, "The Human Body in Space," NASA, 30-Mar-2016. [Online]. Available: https://www.nasa.gov/hrp/bodyinspace. [Accessed: 16-Jun-2020].

[55] M. D. Rienzo, E. Vaini, and P. Lombardi, "Development of a Smart Garment for the Assessment of Cardiac Mechanical Performance and Other Vital Signs

During Sleep in Microgravity," Sensors and Actuators A: Physical, vol. 274, pp. 19-27, 2018.

[56] Hexoskin, "Astroskin Vital Signs Monitoring Platform," Carre Technologies inc (Hexoskin). [Online]. Available: https://www.hexoskin.com/pages/astroskin-vital-signs-monitoring-platform-for-advanced-research. [Accessed: 17-Feb-2020].

[57] A. Kumar, E. Lewin, P. Cowings, and W. Toscano. *Evaluation of the Accuracy of Astroskin as a Behavioral Health Self-Monitoring System for Spaceflight.* 2015. Poster session presented at 9th Annual STEM Summer Internship Symposium, Salinas, CA

[58] "Wearable Device Revenue Worldwide 2016-2022," Statista, 19-Feb-2020. [Online]. Available: https://www.statista.com/statistics/610447/wearable-device-revenue-worldwide/. [Accessed: 28-Jul-2020].

[59] "The MIT License," The MIT License | Open Source Initiative. [Online]. Available: https://opensource.org/licenses/MIT. [Accessed: 31-May-2020].

[60] B. Hewgill, "bhewgs/WIHMS-Development," GitHub, 15-May-2020. [Online]. Available: https://github.com/bhewgs/WIHMS-Development. [Accessed: 31-May-2020].

[61] P.H. Charlton, T. Bonnici, L. Tarassenko, D.A. Clifton, R. Beale, and P.J. Watkinson, "An Assessment of Algorithms to Estimate Respiratory Rate From the Electrocardiogram and Photoplethysmogram," Physiological Measurement, vol. 37, no. 4, pp.610, 2016.

[62] P.A.M.R. Martins, "Cuffless Blood Pressure Estimation," Repositorio Universidade Nova, Dec-2019. [Online]. Available: https://run.unl.pt/handle/10362/90752.

[63] L.C. Dascualu, C. Babis, O. Chivu, G Iacobescu, A.M. Alecusan, A. Semenescu, "Measurements of Galvanic Skin Response on Subjects Affected by Stress," Acta Universitatis Cibiniensis. Technical Series, vol. 72, no. 1, pp. 25-29, 2019.

[64] N. Bourdillon, L. Schmitt, S. Yazdani, J.-M. Vesin, and G. P. Millet, "Minimal Window Duration for Accurate HRV Recording in Athletes," Frontiers, 26-Jul-2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2017.00456/full. [Accessed: 02-May-2020].

[65] False, "MAX30102 High-Sensitivity Pulse Oximeter and Heart-Rate Sensor for Wearable Health - Maxim Integrated," MAX30102 High-Sensitivity Pulse Oximeter and Heart-Rate Sensor for Wearable Health - Maxim Integrated. [Online]. Available: https://www.maximintegrated.com/en/products/interface/sensor-interface/MAX30102.html?utm_source=googleads. [Accessed: 15-Apr-2020].

[66] S. K. Longmore, G. Y. Lui, G. Naik, P. P. Breen, B. Jalaludin, and G. D. Gargiulo, "A Comparison of Reflective Photoplethysmography for Detection of Heart Rate, Blood Oxygen Saturation, and Respiration Rate at Various Anatomical Locations," Sensors, vol. 19, no. 8, p. 1874, 2019.

[67] "AD8232," AD8232 Datasheet and Product Info | Analog Devices. [Online]. Available: https://www.analog.com/en/products/ad8232.html. [Accessed: 22-Mar-2020].

[68] "The Standard 12 Lead ECG," ECG Learning Center - An introduction to clinical electrocardiography. [Online]. Available: https://ecg.utah.edu/lesson/1. [Accessed: 31-May-2020].

[69] "3M Red Dot ECG Electrodes," 3M in the United States. [Online]. Available: https://www.3m.com/3M/en_US/company-us/all-3m-products/All-3M-Products/Health-Care/Patient-Monitoring/Electrodes/?N=5002385+8707795+8711017+8711096+8711111+3294857497. [Accessed: 09-Jan-2020].

[70] "BNO055," Bosch Sensortec, 28-May-2020. [Online]. Available: https://www.bosch-sensortec.com/products/smart-sensors/bno055.html. [Accessed: 22-Feb-2020].

[71] C. Massaroni, A. Nicolo, D. L. Presti, M. Sacchetti, S. Silvestri, and E. Schena, "Contact-Based Methods for Measuring Respiratory Rate," Sensors, vol. 19, no. 4, p. 908, 2019.

[72] "NTCLE100E3474GB0," DigiKey. [Online]. Available: https://www.digikey.com/product-detail/en/vishay-bc-components/NTCLE100E3474GB0/BC2429-ND/2230757. [Accessed: 31-May-2020].

[73] NXP Semiconductors. UM10204 I2C-bus specification and user manual. [Online] - Rev. 4 - 4th April 2020. Available: http://www.nxp.com/documents/user_manual/UM10204.pdf. [Accessed 5-May-2020]

[74] A. Pandit, "Serial Communication Protocols," Serial Communication Protocols: Basics, Transmission Modes, Synchronous, Asynchronous Serial Protocols, 29-Apr-2019. [Online]. Available: https://circuitdigest.com/tutorial/serial-communication-protocols. [Accessed: 30-Apr-2020].

[75] M. A. Meguid, Mohamed Abdel Meguid's answer to "What are the basic differences between the I2C and SPI interfaces?," Quora, 04-Nov-2018. [Online]. Available: https://www.quora.com/What-are-the-basic-differences-between-the-I2C-and-SPI-interfaces. [Accessed: 08-Jan-2020].

[76] "RS485, RS422 and RS232," https://www.omega.com/en-us/, 23-Dec-2019. [Online]. Available: https://www.omega.com/en-us/resources/rs422-rs485-rs232. [Accessed: 31-May-2020].

[77] L. Ada, "I2C addresses!," Adafruit Learning System. [Online]. Available: https://learn.adafruit.com/i2c-addresses/the-list. [Accessed: 02-Mar-2020].

[78] "PCA9615DP: 2-channel multipoint Fast-mode Plus differential I2C-bus buffer with hot-swap logic," NXP. [Online]. Available: https://www.nxp.com/products/interfaces/ic-spi-serial-interface-devices/ic-bus-repeaters-hubs-extenders/2-channel-multipoint-fast-mode-plus-differential-i2c-bus-buffer-with-hot-swap-logic:PCA9615DP. [Accessed: 04-May-2020].

[79] P. Dhaker, "Introduction to SPI Interface," Introduction to SPI Interface | Analog Devices, Sep-2018. [Online]. Available: https://www.analog.com/en/analog-dialogue/articles/introduction-to-spi-interface.html. [Accessed: 26-Feb-2020].

[80] "IEEE 802.3," IEEE 802.3 ETHERNET. [Online]. Available: http://www.ieee802.org/3/. [Accessed: 31-May-2020].

[81] "Description of the FAT32 File System," support.microsoft.com, 02-Sep-2018. [Online]. Available: https://support.microsoft.com/en-us/help/154997/description-of-the-fat32-file-system. [Accessed: 31-May-2020].

[82] "Vital Signs," Cleveland Clinic, 23-Jan-2019. [Online]. Available: https://my.clevelandclinic.org/health/articles/10881-vital-signs. [Accessed: 18-May-2020].

[83] "Remote Patient Monitoring Solutions, Body Worn Sensors and Connected Health Platform." Hexoskin. Available: https://cdn.shopify.com/s/files/1/0284/7802/files/ Hexoskin_-_Pager_-_Connected_Health.pdf?932. [Accessed: 10-Feb-2020].

[84] Hexoskin and Carre Technologies inc, "Astroskin Now Available - Pushes the Boundaries of Health Research on Earth and in Space!," Carre Technologies inc (Hexoskin), 19-Sep-2018. [Online]. Available: https://www.hexoskin.com/blogs/news/astroskin. [Accessed: 13-Feb-2020].

[85] "The Most Accurate Sleep and Activity Tracker," Oura Ring. [Online]. Available: https://ouraring.com/. [Accessed: 25-May-2020].

[86] S. Mahdiani, V. Jeyhani, M. Peltokangas, and A. Vehkaoja, "Is 50 Hz High Enough ECG Sampling Frequency for Accurate HRV Analysis?," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015.

[87] S. Beres, L. Holczer, and L. Hejjel, "On the Minimal Adequate Sampling Frequency of the Photoplethysmogram for Pulse Rate Monitoring and Heart Rate Variability Analysis in Mobile and Wearable Technology," Measurement Science Review, vol. 19, no. 5, pp. 232-240, 2019.

[88] L. G. Tereshchenko and M. E. Josephson, "Frequency Content and Characteristics of Ventricular Conduction," Journal of Electrocardiology, vol. 48, no. 6, pp. 933-937, 2015.

[89] S. Lahmiri, "Comparative Study of ECG Signal Denoising by Wavelet Thresholding in Empirical and Variational Mode Decomposition Domains," Healthcare Technology Letters, vol. 1, no. 3, pp. 104-109, 2014.

[90] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," IEEE Transactions on Biomedical Engineering, vol. BME-32, no. 3, pp. 230-236, 1985.

[91] K.K. Kim, J.S Kim, Y.G. Lim, and K.S. Park, "The Effect of Missing RR-Interval Data on Heart Rate Variability Analysis in the Frequency Domain," Physiological Measurement, vol. 30, no. 10,pp. 1039, 2009.

[92] A.E. Awodeyi, S.R. Alty, and M. Ghavami, "Median Filter Approach for Removal of Baseline Wander in Photoplethysmography Signals," in 2013 European Modelling Symposium. IEEE, 2013,pp. 261-264.

[93] M.A. Motin, C.K. Karmakar,and M. Palaniswami, "Ensemble Empirical Mode Decomposition with Principal Component Analysis: A Novel Approach for Extracting Respiratory Rate and Heart Rate from Photoplethysmographic Signal," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 3, pp. 766-774, 2017.

[94] "Hypoxemia (low blood oxygen)," Mayo Clinic, 01-Dec-2018. [Online]. Available: https://www.mayoclinic.org/symptoms/hypoxemia/basics/definition/sym-20050930. [Accessed: 05-Jun-2020].

[95] A.A. Romanovsky, "Skin Temperature: Its Role in Thermoregulation," Acta Physiologica, vol. 210, no. 3, pp. 498-507, 2014.

[96] S. Ray, "Commonly Used Machine Learning Algorithms: Data Science," Analytics Vidhya, 15-Apr-2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/. [Accessed: 02-May-2020].

[97] Mudgalvivek, "Machine Learning: Confusion Matrix (Error Matrix)," Medium, 24-Jun-2020. [Online]. Available: https://medium.com/@mudgalvivek2911/machine-learning-confusion-matrix-error-matrix-c518bca18de7. [Accessed: 31-Jul-2020].

[98] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[99] J. Brownlee, "What is Deep Learning?," Machine Learning Mastery, 19-Dec-2019. [Online]. Available: https://machinelearningmastery.com/what-is-deep-learning/. [Accessed: 04-Jun-2020].

[100] "Deep Network Designer," Pretrained Deep Neural Networks - MATLAB; Simulink. [Online]. Available: https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html. [Accessed: 04-Jun-2020].

[101] X. An and G. Stylios, "A Hybrid Textile Electrode for Electrocardiogram (ECG) Measurement and Motion Tracking," Materials, vol. 11, no. 10, p. 1887, 2018.

[102] "Search Results for 'BLE 2560'," Adafruit Customer Service Forums Search. [Online]. Available: https://forums.adafruit.com/search.php?keywords=ble+2560. [Accessed: 01-Jun-2020].

[103] C. Zou, Y. Qin, C. Sun, W. Li, and W. Chen, "Motion Artifact Removal Based on Periodical Property for ECG Monitoring with Wearable Systems," Pervasive and Mobile Computing, vol. 40, pp. 267-278, 2017.

[104] D. Pollreisz and N. Taherinejad, "Detection and Removal of Motion Artifacts in PPG Signals," Mobile Networks and Applications, 2019.

Appendix A
Glossary

1. *RespAmp:* Respiration Amplitude. The offset in raw ADC counts from the beginning of an exhalation cycle to the start of the inhalation cycle. Measured using the BNO055.

2. *InLen:* Inhalation Length. The time, in seconds, between each local minimum and its subsequent maximum. Measured using the BNO055.

3. *ExLen:* Exhalation Length. The time, in seconds, between each local maximum and its subsequent minimum. Measured using the BNO055.

4. *RedROC(+):* Positive Rate of Change from the Red LED of the MAX30102. This is the positive first derivative of the red PPG vector.

5. *RedROC(-):* Negative Rate of Change from the Red LED of the MAX30102. This is the negative first derivative of the red PPG vector.

6. *skinT:* Skin Temperature. This value is the calculated skin surface temperature from the thermistor located on the subject's left flank.

7. *skinROC:* Skin Temperature Rate of Change. SkinROC is the first order difference between each subsequent point in a noise-filtered vector of skin temperature data.

8. *tempROC:* Rate of change of the difference between skin and ambient temperatures. This value is the first order difference between subsequent samples of the skin/ambient temperature offset.

9. *SPO(1):* One of two calculations of blood oxygen saturation determined from the red and infrared MAX30102 measurements. The equation for this parameter is located in the text as Equation 5.2.

10. *SPO(2):* One of two calculations of blood oxygen saturation determined from the red and infrared MAX30102 measurements. The equation for this parameter is located in the text as Equation 5.3.

11. *rMSSD:* Root Mean Square of Successive Differences. This is a time domain measurement used to characterize heart rate variability, calculated from the ECG vector. The equation for this value is

$$rMSSD = \sqrt{\frac{\Sigma(RR_i - RR_{i+1})}{N-1}} \tag{A.1}$$

    where RR_i is a sample in the vector of RR-intervals and N is the length of the vector.

12. *LFHF:* Low Frequency to High Frequency Ratio. Described in Section 2.4.1, the LFHF value is the power ratio of spectral content in Low RR-interval frequencies (0.05-0.15 Hz) and High RR-interval frequencies (0.15-0.40 Hz).

13. *RespRate:* Respiration Rate, extracted from the BNO055 gyroscope vector. The RespRate value is calculated as the mean length in seconds between each subsequent breath characterized by a new inhalation cycle.

Appendix B
Bill of Materials

Component	Vendor	Part Number	Assembly	Each	Total
Bicycling Shirt	Amazon	B07C7VNQG1	Base	$19.99	$19.99
Arduino Mega 2560	Amazon	B00D9NA4CY	Processing	$14.99	$14.99
MAX30102 Board	Amazon	B07ZQNC8XP	PPG Sensor	$7.50	$7.50
Gravity ECG	DFRobot	SEN0213	ECG Sensor	$19.90	$19.90
Bosch BNO055	Amazon	B07W4HHV9F	Respiration	$15.40	$15.40
SSD1306	Amazon	B07X25T786	Display	$4.90	$4.90
470kΩ Resistor	Digikey	S473KCACT	Temperature	$0.07	$0.14
470kΩ Thermistor	Digikey	BC2429	Temperature	$0.87	$1.74
Moisture PCB	JLCPCB	Y4-2781365A	GSR Sensor	$3.94	$3.94
5.1kΩ Resistor	Digikey	311-5.10KHRCT	GSR Sensor	$0.02	$0.05
S8050 Transistor	Digikey	MMSS8050	GSR Sensor	$0.17	$0.17
Interface PCB	JLCPCB	Y3-2781365A	Interface	$9.80	$9.80
TCA9517	Digikey	TCA9517DR	Interface	$0.83	$4.13
8P8C Connector	Digikey	609-1046	Interface	$0.52	$3.12
4P4C Connector	Digikey	AE10381	Interface	$0.47	$1.88
1 μF Capacitor	Digikey	1276-1182-1	Interface	$0.02	$0.22
.1 μF Capacitor	Digikey	399-1100-1	Interface	$0.01	$0.19
10 μF Capacitor	Digikey	490-12736-1	Interface	$0.30	$0.90
64 GB μSD Card	Amazon	B06XX29S9Q	Interface	$11.99	$11.99
μSD Breakout	Amazon	B07PFDFPPC	Interface	$5.00	$5.00
DS3231 Clock	Digikey	1528-1598	Interface	$13.00	$13.00
Coin Cell Battery	Digikey	P033-ND	Interface	$1.00	$1.00
0Ω Resistor	Digikey	541-0.0SBCT	Interface	$0.05	$0.20
Micro USB	Digikey	732-5960-1	Interface	$2.07	$2.07
3.3 V TVS Diode	Digikey	F7701CT	Interface	$0.49	$0.49
5 V TVS Diode	Digikey	SMF5A-TPMSCT	Interface	$0.41	$0.41
2.2kΩ Resistor	Digikey	311-2.20KHRCT	Interface	$0.01	$0.04
3.3kΩ Resistor	Digikey	A129694CT	Interface	$0.01	$0.05
5 V Regulator	Digikey	S7V7F5	Interface	$4.95	$4.95
3.3 V Regulator	Digikey	S7V8F3	Interface	$5.95	$5.95
4P4C Cable	Amazon	B01F3CSWC2	Cabling	$2.16	$6.48
8P8C Cable	Amazon	B00E5I7VJG	Cabling	$2.20	$8.79
Total	-	-	-	-	**$169.38**

Table B.1: Bill of materials for construction of the WIHMS garment

Appendix C
Supplemental Figures

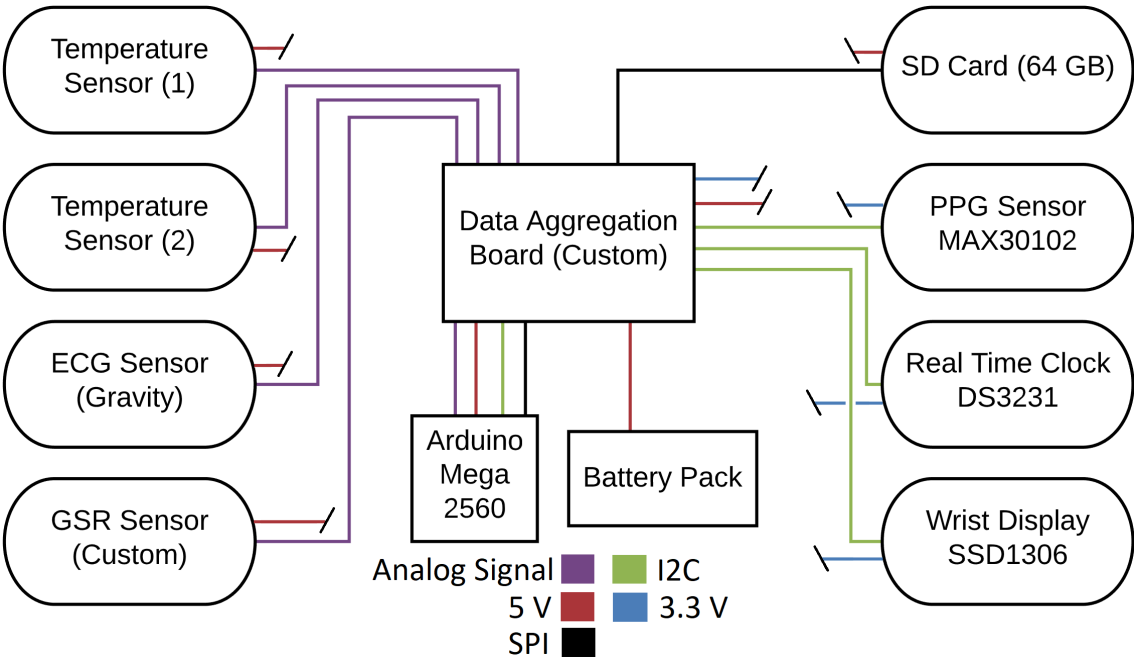

Figure C.1: System block diagram layout

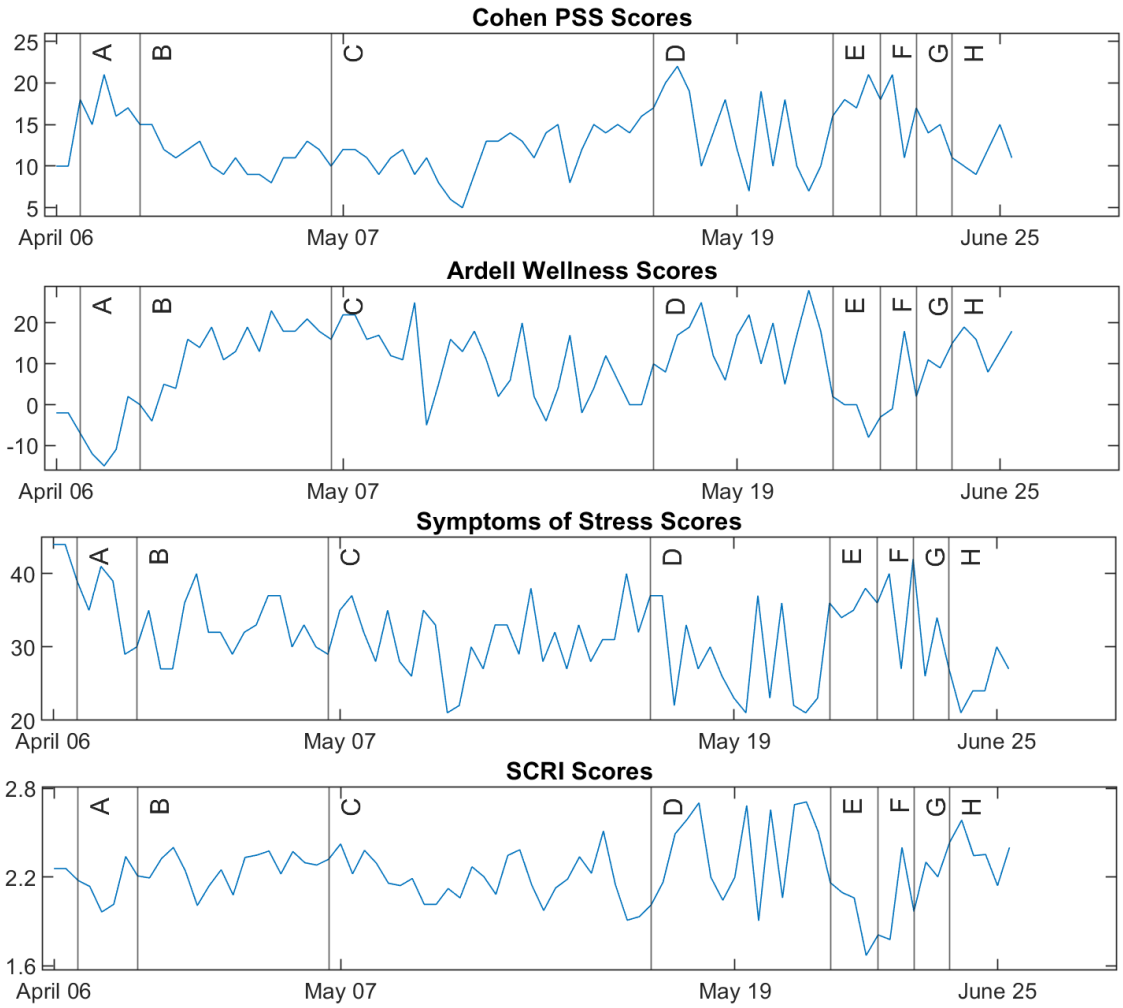

Figure C.2: Stress scores over time with accompanying events. A: April 10, 100,000 global COVID deaths. B: April 20, Oil prices go negative. C: May 7, Last day of classes at UVM. D: May 16, Large personal investment losses. E: May 26, George Floyd protests begin. F: May 30, SpaceX Crew Dragon launch. G: June 6, Nationwide police brutality protests. H: June 12, Submission of remote conference presentation.

Appendix D
Source Code

This appendix includes the source code for the operation of the measurement system discussed throughout this thesis.

D.1 Embedded Processing

Arduino code used for data collection can be found at

https://github.com/bhewgs/WIHMS-Development/tree/master/Embedded

D.2 Data Analysis Code

MATLAB scripts and functions used for data processing and analysis can be found at

https://github.com/bhewgs/WIHMS-Development/tree/master/Processing