5-1-2016

StressAware: App for Continuously Measuring and Monitoring Stress Levels in Real Time on the Amulet Wearable Device

George G. Boateng
Dartmouth College

StressAware: *App for Continuously Measuring and Monitoring Stress Levels in Real Time on the Amulet Wearable Device*

Dartmouth Computer Science Technical Report TR2016-802

Computer Science Honors Thesis

George Boateng '16

Thesis Outline

Abstract

Stress is the root cause of many diseases. Being able to monitor when and why a person is stressed could inform personal stress management as well as interventions when necessary. In this thesis, I present *StressAware,* an application on the Amulet wearable platform to measure the stress levels of individuals continuously and in real time. The app implements a stress detection model, continuously streams heart rate data from a commercial heart-rate monitor such as a Zephyr and Polar H7, classifies the stress level of an individual, logs the stress level and then displays it as a graph on the screen. I developed a stress detection model using a Linear Support Vector Machine. I trained my classifiers using data from 3 sources: PhysioNet, a public database with various physiological data, a field study, where subjects went about their normal daily activities and a lab study in a controlled environment, where subjects were exposed to various stressors. I used 73 data segments of stress data obtained from PhysioNet, 120 data segments from the field study, and 14 data segments from the lab study. I extracted 14 heart rate and heart rate variability features. With 10-fold cross validation for Radial Basis Function (RBF) SVM, I obtained an accuracy of 94.5% for the PhysioNet dataset and 100% for the field study dataset. And for the lab study, I obtained an accuracy of 64.29% with leave-one-out cross-validation. Testing the *StressAware* app revealed a projected battery life of up to 12 days before needing to recharge. Also, the usability feedback from subjects showed that the Amulet and Zephyr have a potential to be used by people for monitoring their stress levels. The results are promising, indicating that the app may be used for stress detection, and eventually for the development of stress-related intervention that could improve the health of individuals.

1 Introduction

The American Medical Association has noted that stress is the underlying cause of more than 60 percent of all human illness and disease (The Huffington Post, 2016). Stress can trigger onset or recurrence of addictive behaviors like unhealthy eating, smoking, or drug use. There is a need to measure stress, unobtrusively and continuously and in the field, because stress is so often tied to these challenging behaviors.

Knowing when, where, and why a person is under stress can help health professionals develop mechanisms to intervene "in the moment," in a way appropriate to the person and the condition, to help that person deal with the stress and avoid the unhealthy behavior or seek out healthy stress-reducing activities. Real-time stress measurement will also enable research at the Center for Technology and Behavioral Health, and other Dartmouth centers, where innovative new interventions are being developed to help people improve their health-related behaviors - primarily, people challenged with addiction or mental illness. Additionally, having more information about stress can help individuals manage their own stress levels.

In this work, I built an application for the Amulet, a low-power wrist-worn device that continuously monitors the stress level of individuals in real time using data from a commercial heart-rate chest strap such as the Zephyr. I developed a machine-learning model to detect stress using a Linear Support Vector Machine (SVM). I used data from PhysioNet, a public database with various physiological data, and two sets of studies approved by Dartmouth's Institutional Review Board - an in-lab study and a field study. For the lab study, I collected heart-rate data from subjects as they performed various stress-inducing activities. For the field study, volunteers wore the Amulet (running my StressAware app) and a Zephyr heart monitor for 8 hours. The app collected heart-rate data from subjects as well as their corresponding perceived stress levels as they went about their regular activities during waking hours. I then built an app for the Amulet platform that implements the developed stress-detection model, continuously streams heart-rate data from the Zephyr, classifies the stress level of an individual, logs the stress level, and then displays it as a graph on the screen.

In the remainder of this thesis, I describe the Amulet platform on which *StressAware* runs. I then provide a detailed description of the science of stress and the other work that has been done in stress research. I then give an overview of the *StressAware* app. Next, I describe the process of developing the *StressAware* machine-learning model. I then give a detailed description of the different components of the *StressAware* app. Next, I describe the energy efficiency results and usability feedback from the user study. I then describe various limitations of this work and propose ways for improvement. I finally tie together the results from the various parts of this thesis into my stated goal to develop an app the Amulet platform that continuously monitors the stress levels of individuals to aid personal stress management and intervention when necessary.

2 Background

In this section, I describe the necessary background to understand the work described in this thesis. First, I describe the Amulet platform on which the *StressAware* app runs and why it is suitable for running the app. I then describe the science of stress and its relation to stress measurement.

2.1 Amulet Wearable Device Platform

The Amulet platform is a hardware and software platform for developing energy- and resource-efficient applications on multi-application wearable devices. It includes an ultralow-power hardware architecture and a companion software framework, including a highly efficient event-driven programming model, low-power operating system, and developer tools for analyzing and profiling ultra-low-power applications at compile time.

The Amulet hardware is a two-processor system. Specifically, it has two micro-controllers: the MSP430 running applications, and the nRF51822 (aka M0) managing communication over Bluetooth. The MSP430 microcontroller has 2 KB of SRAM and 128 KB of integrated FRAM. The nRF51822 is used as a modem for communicating with peripheral BLE devices such as a heart-rate monitor.

The main board has two buttons, three capacitive touch sensors, a battery, a haptic buzzer, two LEDs embedded in the case, a secondary storage board that holds a microSD card reader, and a display screen. It has several sensors: microphone, light sensor, temperature sensors, accelerometer, and gyroscope.

The Amulet platform enables developers to write energy- and memory-efficient sensing applications that achieve long battery life on a secure, open-source, multi-application wearable device. The Amulet platform is hence useful for creating and running mHealth apps that need to continuously run for long hours to monitor the physiological and behavioral health of users.

2.2 Science of Stress

Stress is a physiological response to mental, emotional, or physical challenges we encounter (Sun et al., 2010). When a person is stressed, the hypothalamus signals to two systems in the body, namely, the Hypothalamic Pituitary-Adrenal (HPA) system and Sympathomedullary Pathway (SAM) system (McLeod, 2010).

The HPA is responsible for long-term stress response. The HPA deals with the adrenal cortex, which releases cortisol, a stress hormone whose function is to increase the amount of glucose available to a person in preparation for the stressful situation. Cortisol is present in saliva, urine and blood and measuring cortisol levels could be used as an indicator of a person's stress level (Ertin et al., 2011).

The SAM is responsible for short-term stress response. It deals with the autonomic nervous system (ANS), which is responsible for regulating the body's involuntary functions such as heart rate, respiratory rate, digestion etc. The autonomic nervous system has two parts: sympathetic nervous system (SNS) and parasympathetic nervous system (SNS). The SNS controls the "fight or flight" response and prepares the body for emergency or stressful situations. It increases heart rate and increases blood flow to the brain, heart and muscles. The PNS is active during rest and reduces heart rate. When a person is under stress, the SNS increases heart rate, sweating, respiratory rate etc. This response is reversed by the PNS when the stressful situation ends (Sun et al., 2010).

Having described that there is a cardiac response to stress, we can therefore infer whether a person is stressed by looking at their cardiac activity, which can be captured electrically with an electrocardiogram (ECG) (Figure 1).

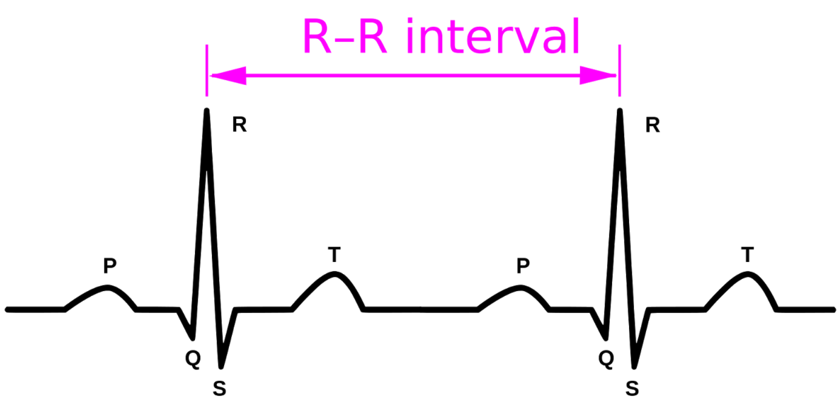

Figure 1: An Electrocardiogram

The heart rate (HR) is the number of R peaks in a minute. The RR interval (RRI) is the time interval between two R peaks. Heart rate variability (HRV), which is the variability in RR intervals, provides information about the relative activation between the SNS and PNS (Ertin et al., 2011). HRV may be used to distinguish the stress level of individuals.

3 Related Work

Several research projects aimed to collect continuous measures of stress both in and outside the laboratory setting. This research has shown that several physiological signals can be used, alone or in combination, to indicate a person's stress level.

A study conducted by Healey and Picard involved collecting and analyzing physiological data during real-world driving tasks to determine a driver's relative stress level (Healey et al., 2005). They recorded ECG, electromyogram (EMG), skin conductance, and respiration continuously while drivers followed a set route through open roads in the greater Boston area. They were able to distinguish between three stress levels – low stress, medium stress and high stress – using 5-minute intervals of data during rest, highway and city driving. They used linear discriminant analysis with 112 data segments and 22 features and had an accuracy of over 97% with leave-one-out cross-validation across different drivers and driving days. Healey and Picard's study shows that physiological signals can be used to determine stress levels. Specifically, they found that skin conductance and heart-rate metrics individually were closely correlated with drivers' stress and hence can be used to predict mental stress levels with high accuracy.

Another study, conducted by Ertin et al., used a custom suite of wireless sensors called Autosense to infer the stress of subjects (2011). AutoSense combines six sensors into a wearable chest band: two-lead Electrocardiogram for measurement of electrical activity of the heart; respiratory inductive plethysmograph (RIP) for measurement of relative lung volume and breathing rate at the rib cage; galvanic skin response (GSR) between the two ECG electrodes; skin temperature thermistor under the arm; ambient temperature sensor; and three-axis accelerometer to assess motion artifacts in the data and provide inferences about the subjects' physical activities (Ertin et al., 2011). They obtained data from participants via a lab study and a field study.

For the lab study, participants wore the Autosense sensor suite and underwent a rigorous stress protocol that consisted of public speaking, mental arithmetic, and a cold pressor test. The field study entailed participants wearing AutoSense during waking hours in their natural environment for two days. The data obtained from this sensor suite was then sent to an Android mobile phone, where 30 features were extracted to infer whether the person was stressed, speaking, had changed their posture, and the intensity of activity. The participants provided self reports of stress in both studies. They ran their study with 20+ subjects and obtained an accuracy of 90% on in-lab data and a median correlation of 0.72 with self-reported rating of stress.

4 Solution: StressAware

Given the need to continuously measure and monitor the stress levels of individuals, I developed the *StressAware* app for the Amulet wearable device. StressAware monitors a person's stress level on a scale of low, medium, and high, using data from a heart-rate monitor such as the Zephyr HxM, logs that information, and then displays the stress level over the past hours as a bar graph on the Amulet screen. The app uses an implementation of a machine-learning model that is trained offline. The information about stress levels can in the future be used by the user for management of their stress levels as well as prompt intervention when necessary.

5 Stress Detection Model - Machine Learning Offline

SVM is a classifier that performs classification by constructing a high-dimensional hyper-plane (Burges, 1998). SVM is recently popular for mining physiological data because of its ability to handle high dimensional data using minimal training features (Banaee, 2013). I focus on using SVM because it uses a subset of the training set - support vectors - for its prediction function as compared to other models like k-nearest neighbor (KNN), which will need to store all the data points in memory for prediction. It is hence memory efficient and ideal for low-memory platform like the Amulet. I trained two SVM models, one a Linear SVM and the other a Radial Basis Function (RBF) SVM, using the *scikit-learn* library (Pedregosa et al, 2011) to distinguish between low, medium, and high stress levels.

5.1 Data Extraction I: Physionet

I obtained the data used for training the machine-learning model from the MIT-BIH Multi-parameter Database in PhysioNet, a public database with various physiological data (PhysioNet, 2010). I used a PhysioNet dataset contributed by its creator, Jennifer Healey. The dataset was collected during a stress study that involved collecting and analyzing physiological data during real-world driving tasks to determine a driver's relative stress. It contains a collection of multiparameter recordings from healthy volunteers, taken while drivers followed a set route through open roads in the greater Boston area.

The data was collected from 17 participating drivers and consists of eight types of raw data – timestamp, ECG, EMG, foot galvanic skin response (GSR), hand GSR, instantaneous heart rate, marker, and respiration (Figure 2).

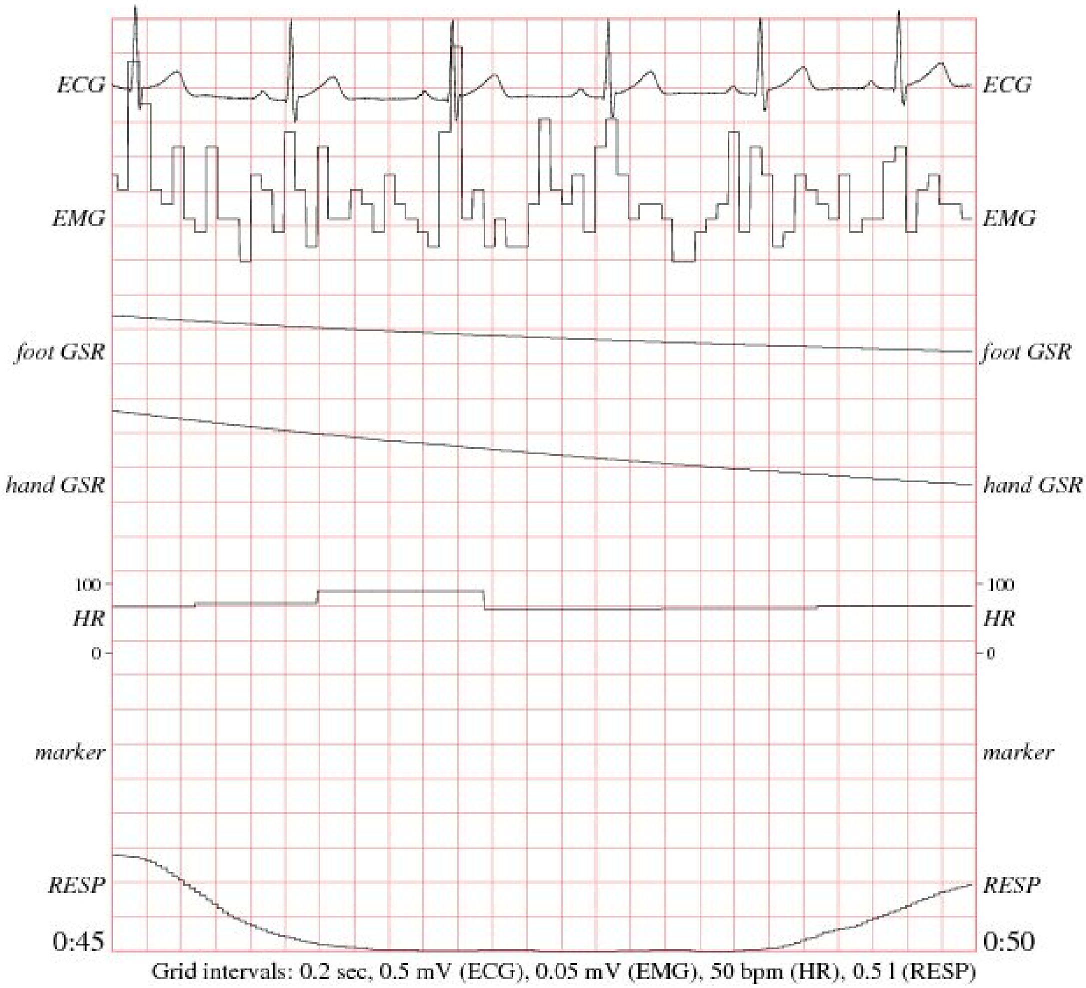

Figure 2: An example of driver bio-signal dataset obtained from PhysioNet (2010) website.

Healey et al. segmented the dataset into three stress levels based on the corresponding part of the drive the data was collected. Data from initial rest and final rest were annotated as low stress. Data from the drive through highways were annotated as moderate stress. Data from the drive through cities were annotated as high stress.

The dataset on PhysioNet did not clearly indicate which data points of the dataset corresponded to each of the three stress levels. As a result, I relied on the durations for each of the 7 segments of the drive dataset given in (Akbas, 2011) to assign the data points to various stress levels. Akbas (2011) found 10 of the 17 drives to have the clear segmentation, so I used the one of the 10 drivers' datasets specified by Akbas (Table 1). Specifically, I randomly chose and used Drive16 dataset.

Table 1: Time intervals of the 7 driving segments

Rec Name	Driving Period (min)							Total Rec Time (min)
	Initial Rest	City 1	HW1	City 2	HW2	City 3	Final Rest	
Drive05	15.13	1600	7.74	6.06	7.56	14.96	15.78	83.23
Drive06	15.05	14.49	7.32	6.53	7.64	12.29	15.05	78.37
Drive07	15.04	16.23	10.96	9.83	7.64	10.15	15.03	84.88
Drive08	15.00	12.31	7.23	9.51	7.64	13.43	15.07	80.19
Drive09	15.66	19.21	8.47	5.20	7.06	13.21	NA	68.82
Drive10	15.04	15.30	8.66	5.27	7.04	12.06	14.79	78.16
Drive11	15.02	15.81	7.43	7.15	6.96	11.72	14.99	79.08
Drive12	15.01	13.41	7.56	6.50	8.06	11.68	15.01	77.23
Drive15	15.00	12.54	7.24	5.99	6.82	12.12	15.00	74.71
Drive16	15.01	16.12	7.14	5.12	6.81	13.91	NA	64.11

I used the PhysioNet software to extract the HR and RRI from the ECG data. I extracted the 7 segments using the durations specified in Table 1. I then grouped the data into the 3 stress levels. Next, I split the data into 60-second time windows, which Hovsepian has shown is good for stress data analysis (Hovsepian, 2015). I had a train dataset containing 73 data points.

5.2 Data Extraction II: Field Study

I ran a field study from which I collected data from a total of 10 subjects. The participants wore the Amulet and Zephyr for one day, lasting between 8 and 12 hours. The Zephyr transmitted HR and RRI data to the Amulet throughout the day. The Amulet also recorded acceleration data.

The app on the Amulet logged 5 minutes or 1 minute of data every 10 mins. The app then prompted the subjects to answer 2 questions via the EMA component of the Amulet app. The app asked subjects to rate their stress levels and their activity levels between low, medium and high at the moment. There were 4 EMAs per hour and at least 32 EMAs per day. After I collected the data, I used data from 4 subjects to obtain a train dataset containing 120 data points.

5.3 Data Extraction II: Lab Study

I collected data from a total of 2 subjects in the lab study. The participants were subjected to mild stressors for about 80 mins that previous experiments have shown to induce stress (Linden, 1991; Poh et al., 2010; Sun et al., 2010; Plarre et al., 2011). Each subject in the protocol was exposed to 6 rest periods and 5 stressors: one public speaking stressor, two mental arithmetic stressors, one startling with a clap sound, and one cold pressor stressor. Table 2 shows the duration for each of the stressors and rest period.

Table 2: Duration of Stressors and Rest Periods

Session	**Duration (mins)**
Rest I	10

Public Speech	4
Rest II	5
Mental Arithmetic I (while seated)	4
Rest III	5
Mental Arithmetic II (while standing)	4
Rest IV	5
Startling with Clap Sound	4
Rest V	5
Cold Pressor	4
Rest VI	10

For the public speaking exercise, the participants were asked to spend the next 4 minutes preparing for an oral presentation of William Faulkner's 1950 Nobel Prize acceptance speech that was provided. They were then asked to read the speech out loud to completion.

For the mental arithmetic exercise, participants were asked to solve a series of mental arithmetic problems. When the problem is correctly solved or three incorrect answers given, the correct answer is given and the next question asked.

For the startling with a clap sound, the participants were asked to sit quietly with their eyes closed. They were then startled with a clap sound at some random time 4 times within a 4 min period.

For the cold pressor, the participants were asked to submerge their dominant hand in a bucket of chilled ice water for as long as they could tolerate.

During the lab study, each subject wore a Zephyr chest strap, a commercial device that measures HR and RRI. The Zephyr transmitted HR and RRI data to the Amulet throughout the duration of the study. The subjects were periodically asked to rate their stress level between low, medium and high for that session, which I later used for the stress annotation. I had a train dataset containing 14 data points from the lab study.

5.4 Feature Extraction

I extracted various HR and HRV features that previous studies have shown to be relevant for stress detection. The features are as follows (Table 3):

- HR features: mean_hr, std_hr, median_hr, percentile_20_hr, percentile_80_hr (Munla et al, 2015; Hovsepian et al. 2015)
- HRV features (Time based): mean_rri, std_rri, rMSSD, NN50, pNN50, median_rri, max_rri, min_rri, percentile_80_rri (Munla et al, 2015; Sun et al, 2010; Plarre et al, 2013)

Table 3: Features Extracted from 60-second time windows of training dataset

Features	**Description**
mean_hr	Mean of heart rate
std_hr	Standard deviation of heart rate
median_hr	Median of heart rate
percentile_80_hr	80th percentile of heart rate
percentile_20_hr	20th percentile of heart rate

mean_rri	Mean of RRI
std_rri	Standard deviation of RRI
rMSSD	Root mean square of the difference between successive RRI
NN50	Number of successive differences in RRI that are greater than *50* ms milliseconds
pNN50	Percentage of total RRI that successively differ by more than 50ms milliseconds
median_rri	Median of RRI
max_rri	Maximum of RRI
min_rri	Minimum of RRI
percentile_80_rri	80th percentile of RRI

5.5 Training/Classification

I trained two SVM models: Linear and RBF SVM. The models classified the data into 3 stress levels - low, medium and high. I ran various experiments to test the two classifiers.

5.6 Testing and Results I: Physionet Dataset

I experimented with different sets of the HR and HRV features. I also experimented with normalizing the data set. I ran these experiments using 10-fold cross validation and evaluated the resulting accuracy.

To evaluate the effect of normalization, I ran 10-fold cross validation with all the 14 features normalized, and then without the features normalized. I normalized the features by making them zero mean and unit variance since various models like SVM assume the data is normalized. These models produce less accurate results without normalization of the features (Pedregosa et al, 2011). When the feature vector is normalized, the accuracy of Linear SVM improves from 53.42% to 63.01%. However, RBF SVM's accuracy rather decreases from 90.41% to 72.6% (Table 4). Overall, RBF did much better than Linear SVM. This result shows that normalizing the feature vector is only necessary for Linear SVM and should be avoided for RBF SVM.

Table 4: Accuracy for Normalized and Non Normalized Features

Normalized		**Not Normalized**	
Linear SVM	**RBF SVM**	**Linear SVM**	**RBF SVM**
63.01%	72.6%	53.42%	90.41%

To evaluate the importance of subsets of the feature set, I ran 10-fold cross validation with 4 sets of features: all features, only HR features, only RRI features, and features that represent an aggregate of several HR and RRI values. The RRI feature set and HR feature set had the least accuracy of 71.23% and 76.71% respectively for RBF SVM (Table 5). The "all features" set did better with 90.41% for RBF. The features that represent an aggregate did best with 94.52% for RBF. Again, RBF did much better than Linear SVM overall. The result shows that it is best to use features that aggregate several HR and RRI values rather than features that are directly chosen from the HR and RRI values such as maximum, minimum, median and percentiles.

Table 5: Prediction Accuracy of Different Feature Sets

Features Description	Feature Set	Linear SVM (Normalized)	RBF SVM
HR features	mean_hr,median_hr, percentile_20_hr, percentile_80_hr, std_hr	64.38%	76.71%
RRI features	min_rri, max_rri,median_rri,percentile_80_rri, mean_rri, std_rri,rMSSD,NN50, pNN50	61.64%	71.23%
Features that aggregate several HR and RRI values	mean_hr, std_hr, mean_rri, std_rri,rMSSD, NN50, pNN50	63.01%	94.52%
All Features	mean_hr, median_hr, percentile_20_hr, percentile_80_hr, std_hr, min_rri, max_rri, median_rri, percentile_80_rri, mean_rri, std_rri, rMSSD, NN50, pNN50	63.01%	90.41%

The best performing model from these two experiments was RBF, which had an accuracy of 94.52% with the following features: mean_hr, std_hr, mean_rri, std_rri, rMSSD, NN50, pNN50.

5.7 Testing and Results II: Field Study

I ran 10-fold cross-validation on the data from the field study using Linear SVM and RBF SVM. I also tested the accuracy of the two models with and without acceleration data.

To test the effect of including accelerometer data, I computed an additional feature: mean of acceleration. I included this feature in the following feature set - mean_hr, std_hr, mean_rri, std_rri,rMSSD, NN50, and pNN50 - which had produced the highest accuracy in previous experiments. I used data from 4 subjects (120 data points) and then ran 10-fold cross-validation. RBF had an accuracy of 89.17% without acceleration, which increased to 100% when acceleration is included (Table 6). Since acceleration captures a person's activity level, it is useful in distinguishing between an increase in heart rate stemming from increased activity and that stemming from stress. Including acceleration helps in accurately classifying the stress level of a person.

Table 6: Accuracy of Field Study with Mean Acceleration Feature Included

Acceleration Present	**Linear SVM**	**RBF SVM**
No	54.17%	89.17%
Yes	52.5%	100%

5.8 Testing and Results III: Lab Study

I ran leave-one-out cross-validation on the data from the lab study using Linear SVM and RBF SVM. I was unable to run 10-fold cross validation because I had 14 data points. Linear SVM had an accuracy of 50% and RBF had an accuracy of 64.29%. RBF once again performed better than Linear SVM. Because of the limited amount of data, I could not infer much from the results.

6 StressAware - App on Amulet

The StressAware app consist of five components: Ecological Momentary Assessment (EMA), Data Collector, Stress Detector, Stress Level Graph, and Data Logger.

6.1 EMA

The Ecological Momentary Assessment (EMA) component of the app is responsible for intermittently asking the user about her/his stress level (Figure 3). The EMA results are used as ground truth of a person's stress level. This component of the app is only used for data collection during the user study.

Figure 3: EMA with 2 questions about stress and activity levels

6.2 Data Collector

The data collector is responsible for getting 5 minutes or 1 minute worth of HR and RRI data from a heart rate monitor (Figure 4). It also collects acceleration data from the Amulet. The data is used by the stress detection model and also logged by the data logger.

Figure 4: Screenshot of HR and RRI data being collected by the Amulet

6.3 Stress Detector

The stress detector determines the stress level of the user. I compute the feature vector from the 60-second HR and RRI data obtained by the data collector. Then, I scale the feature vector using the scaling factors from the trained model. The stress classifier then uses the feature vector for the prediction.

The stress classifier is an implementation of the prediction equation of a Linear SVM. The equation is:

$$y = wx + b$$

where y is vector that holds the result of the evaluation for the 3 stress classes, x is the computed feature vector, w is the coefficient matrix and b is the intercept vector. The values for w and b are obtained from the linear model that was trained offline. Since this is a multilabel classification, I implemented the "one-vs-the-rest" approach for multi-label classification since the *scikit learn* Linear SVM function used this approach. In this approach, three classifiers are trained for each of the classes and the result of solving the equation is a vector containing a value for each of the three classes. The class with the maximum value is the predicted class.

6.4 Stress Level Graph

The stress level graph displays the stress level of the user over the past 2 hours as a bar graph (Figure 5). This information could provide better insight to users about their stress pattern on a particular day.

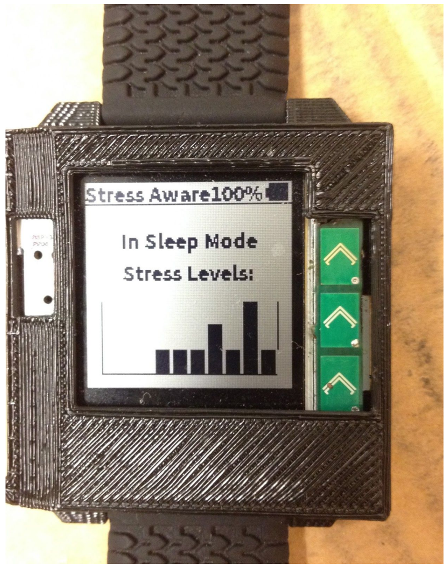

Figure 5: Graph of the last 7 stress levels of the user

6.5 Data Logger

The data logger logs the HR, RRI and acceleration data to a file on the SD card. It also logs the stress level as indicated by the user. I use the logged data to develop the stress detection model. This mode is also only used during the user study.

7 Usability Feedback

After the field study, I created a survey to evaluate the usability of the Amulet in monitoring the stress level of individuals in the wild. All 10 participants in the study filled the survey. There were 3 males and 7 females.

Sixty percent (60%) of subjects mentioned that they were motivated to participate in the study out of interest in stress monitoring. This result shows a general interest of people in stress monitoring.

Half (50%) of the participants mentioned that the Amulet was comfortable to wear. Some people mentioned that the Amulet is a bit bulky. Some were irritated by the frequent EMAs making the wearing experience uncomfortable.

Eighty percent (80%) of participants mentioned it was easy to answer the questions on the Amulet. They mentioned that the questions were simple and clear, and the Amulet was easy to navigate in answering the questions. Some thought the 3 options made it easier to answer, whereas others thought there should be more than 3 options to rate stress level. This result demonstrates the potential of using wrist-worn devices such as the Amulet for EMAs in comparison to mobile phone-based EMAs.

Twenty percent (20%) mentioned that the stress graph was useful. Some people thought it was not useful because they were not stressed on that day. Others did not pay attention to the graph. One suggestion was to make graph show for much longer time periods.

Sixty percent (60%) of participants mentioned that the Zephyr was comfortable to wear. All but one of the females found it comfortable whereas all the males found it uncomfortable. This result was not surprising since males are not used to wearing straps around their torso.

Overall, participants thought the study was a good experience. People enjoyed being able to see their heart rate and stress level, and as they made the connection to the activity they were involved with at the moment. The only concerns were about the bulkiness of the Amulet, the frequent EMAs and the frequent disconnection of the

Amulet from the Zephyr. There were suggestions that the devices should automatically measure stress rather than ask, showing a real interest of people in stress measurement and monitoring. In fact, 70% of participants mentioned that they will wear the Amulet and Zephyr if it automatically measured and monitored their stress level for several hours and days. The responses from the survey show that the Amulet and Zephyr have a potential to be used by people for monitoring their stress levels.

8 Energy Efficiency of *StressAware*

I tested the energy efficiency of the *StressAware* app. I ran *StressAware* for 8 hours as it computed stress levels every 5 minute. I logged the battery voltage level over the 8 hour period. The graph of the log shows battery percentage as the y-axis and time (seconds) as the x-axis. The battery level dropped linearly from 100% to 98% over the 8-hour period indicating a 2% loss in battery life (Figure 6).

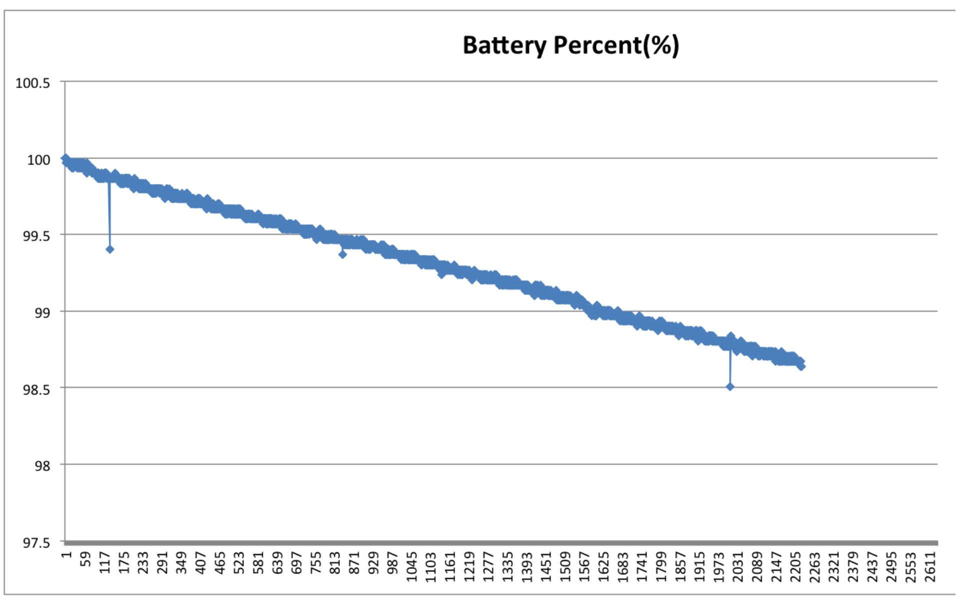

Figure 6: Graph of Battery Percentage over 8 hours running *StressAware*

However, after 8 hours, the battery level indicated a 92% battery level. This means that *StressAware* consumes 8% of battery life per 8 hours: 1% per hour. *StressAware* is expected to run for about 8 hours a day during which period a person is awake. This result means that the app is projected to run for about 2 weeks (12 days) before needing to be recharged. *StressAwar*e is in effect energy efficient.

9 Limitations and Future Work

There were a number of limitations of my experiments that affected my results. Within those limitations lie the opportunity for work in the future. There are two main areas for improvement over this work in the future: improvement of accuracy of the stress detection model and including analytics for understanding causes of stress.

I used only time-based features and no frequency-based features for training the classifiers. Some frequency-based features are total energy in low frequency (LF) of RRI, total energy in high frequency (HF) of RRI, LF/HF, and (LF + MF)/HF. Deriving these features entail performing Fast Fourier Transforms (FFT), which are computationally intensive so I decided to only focus on time-based features for this thesis. However, frequency-based features also capture the nuances in the response to stress by the sympathetic and parasympathetic nervous system. Hence, computing these features could produce more accurate models.

Also, the Amulet platform's API responsible for providing the *StressAware* app with RRI values reports only one RRI value per second. However, the Bluetooth protocol used by the Zephyr mentions that the number of RRIs sent by the Zephyr could be more than one value. In effect, in cases where multiple values are sent by the Zephyr, my app get's only one of those values. The absence of other consecutive RRIs could affect the accuracy of features like NN50, pNN50 and rMSSD that use differences of consecutive RRIs in computing their values. In the future, the creators of the Amulet platform can add the functionality to RRI API to provide all RRI values received via Bluetooth.

The Zephyr is a commercial heart-rate monitor and does not report medical-grade HR and RRI values. In effect, the values used for computing the features would not be as accurate as those obtained from using a medical grade device that produces HR and RRIs. In the future, using a device that produces more accurate HR and RRIs could result in a more accurate stress detection model. Also, obtaining data like breathing rate from a better grade of Zephyr such as the Zephyr Bioharness could be used to improve the accuracy of the stress detection model.

Additionally, I implemented a Linear SVM model in the *StressAware* app rather than an RBF SVM model even though RBF SVM had the highest accuracy. RBF SVM is more computationally intensive and requires more memory to store all the support vectors, which could be as large as the number of data points in the training dataset. Linear SVM on the other hand is not computationally intensive and only requires storing the coefficient matrix, which has size [no of classes, no of features]. In the future, various techniques could be used to address the memory and computational intensity of the RBF decision function such as exploring various approximations of the decision function and also reading the support vectors from the SD card at run time.

Also, after developing the stress detection model from running 10-fold cross-validation, I did not test the detection model with new data from subjects in the wild. In the future, I will run the stress detection model while obtaining ground truth using EMA from subjects. Using a confusion matrix could provide a better insight into the data points that are being misclassified. This information can be used to improve the model.

Finally, I did not focus on collecting data that could be used to infer the causes of a person's stress. In the future, I could keep track of data such as location, noise level, sleep duration, etc., which could be used to diagnose the cause of stress. I also did not store stress data for days. Keeping track of this information and making it readily accessible to the user could help users understand their stress patterns.

10 Conclusion

In this work, I presented *StressAware,* an application on the Amulet wearable platform to measure the stress levels of individuals continuously and in real time. The app implements a stress detection model, continuously streams heart rate data from a commercial heart-rate monitor such as a Zephyr and Polar H7, classifies the stress level of an individual, logs the stress level, and then displays it as a graph on the screen.

The machine-learning results show an accuracy of 94.5% for the PhysioNet dataset, 100% for the field study dataset, and 64.29% for the lab study with RBF SVM. Testing the *StressAware* app revealed a projected battery life of up to 12 days before needing to recharge. Also, the usability feedback from subjects revealed an interest in stress monitoring and showed that the Amulet and Zephyr have a potential to be used by people for monitoring their stress levels.

The machine-learning results, energy efficiency results, and usability results are promising, and show that *StressAware* has the potential to be used for stress measurement and monitoring. The usage of the app could eventually inform the development of stress-related intervention and personal stress management that could improve the health of individuals.

11 Acknowledgement

First, I acknowledge Professor David Kotz, who is my thesis advisor and a co-PI on the Amulet project. I'm grateful for the opportunity to work in the Amulet research group and also for his guidance every step of the way through this thesis. Second, I acknowledge Professor Ryan Halter who is also a co-PI on the Amulet project for his guidance and advice throughout this thesis. I also thank Dr. Sarah Lord who helped me write up the IRB protocol to run my stress detection user study. I also acknowledge the members of the Amulet team that built and continue to work on the Amulet platform: Ron Peterson, Tianlong Yun, Josiah Hester, Travis Peters. I'm grateful for their help with resolving various issues I encountered while using the Amulet platform. I also acknowledge Gunnar Pope who ran the lab study from which I got my lab study data. Finally, I acknowledge friends who volunteered to participate in my user study and enabled me to obtain data for training my classifiers.

References

Akbas, A. (2011). Evaluation of the physiological data indicating the dynamic stress level of drivers. *Scientific Research and Essays*, *6*(2), 430-439.

Banaee, H., Ahmed, M. U., & Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors*, *13*(12), 17472-17500.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, *2*(2), 121-167.

Ertin, E., Stohs, N., Kumar, S., Raij, A., al'Absi, M., & Shah, S. (2011, November). AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems* (pp. 274-287). ACM.

Hovsepian, K., al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M., & Kumar, S. (2015, September). cStress: towards a gold standard for continuous stress assessment in the mobile environment. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 493-504). ACM.

J. Healey and R. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 2, pp. 156– 166, 2005.

Linden, W. (1991). What do arithmetic stress tests measure? Protocol variations and cardiovascular responses. *Psychophysiology*, *28*(1), 91-102.

McLeod, S. A. (2010). What is the Stress Response. Retrieved from www.simplypsychology.org/stress-biology.html

Munla, N., Khalil, M., Shahin, A., & Mourad, A. (2015, September). Driver stress level detection using HRV analysis. In *Advances in Biomedical Engineering (ICABME), 2015 International Conference on* (pp. 61-64). IEEE.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830.

PhysioNet (2010). Stress Recognition in Automobile Drives (drivedb): physiobank / signal archives / multi-parameter database. Available: http://physionet.org/cgi-bin/ATM

Plarre, K., Raij, A., Hossain, S. M., Ali, A. A., Nakajima, M., al'Absi, M., ... & Siewiorek, D. (2011, April). Continuous inference of psychological stress from sensory measurements collected in the natural environment. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on* (pp. 97-108). IEEE.

Poh, M. Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *Biomedical Engineering, IEEE Transactions on*, *57*(5), 1243-1252.

Sun, F. T., Kuo, C., Cheng, H. T., Buthpitiya, S., Collins, P., & Griss, M. (2010). Activity-aware mental stress detection using physiological sensors. In*Mobile computing, applications, and services* (pp. 211-230). Springer Berlin Heidelberg.

The Huffington Post. "How Stress Affects the Body (INFOGRAPHIC)." TheHuffingtonPost.com, n.d. Web. 16 May 2016.