Open Access Master's Theses

2018

EXPLORATORY ANALYSIS OF THE ACCURACY OF BIOSENSOR IN WEARABLE DEVICE

Leichen Dai
University of Rhode Island, lcdai@my.uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/theses

EXPLORATORY ANALYSIS OF THE ACCURACY OF

BIOSENSOR IN WEARABLE DEVICE

BY

LEICHEN DAI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

ELECTRICAL ENGINEERING

UNIVERSITY OF RHODE ISLAND

2018

MASTER OF SCIENCE THESIS
OF
LEICHEN DAI

APPROVED:

    Thesis Committee:

  Major Professor   Yan Sun

                         Bin Li

                         Lisa DiPippo

                         Nasser H. Zawia
         DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2018

ABSTRACT

The new wave of wireless technologies, fitness trackers, and body sensors have had a great impact on personal biometric tracking and monitoring. These technologies make a great contribution to personal health care, and can even be used in clinical settings. Among all of these devices, smartwatches are one of the most popular, and are becoming increasingly common among the general public. Commercially available smartwatches incorporate sophisticated algorithms and multi-sensor technologies, which are capable of providing users with real-time biometrics. Some of these sensors include a photoplethysmography (PPG) sensor that detects the wearer's heart rate, Galvanic skin response sensors which can provide skin surface information, and an accelerometer which can be used to provide activity and movement information. When considering clinical applications, researchers find the smartwatch's PPG sensor to be of most interest, as heart rate is one of the most important vitals that are monitored for clinical purposes. Heart rate can be used to detect and prevent serious diseases, such as cardiovascular diseases and seizures. However, the accuracy of PPG sensors still needs thorough investigation. Although the ability of wearable PPG sensors to reliably measure heart rate in regular movement (i.e. walking or jogging) has been demonstrated in previous research, there doesn't exist premier research that focuses on the accuracy of a PPG sensor in daily activities, such as brushing one's teeth, cooking, or vacuuming. These activities are of interest because they involve short periods of high frequency vibrations or intense wrist movements, which could affect the smartwatch's heart rate calculation. To validate the relative accuracy of a smartwatch's PPG sensor in these activities, a Microsoft Band (MB) and a Huawei

Android smartwatch (HW) were used to conduct a series of experiments from which the heart rate signals were gathered and evaluated. Six participants were recruited to collect data from these two smartwatches, which involved completing a set of three daily activities under a specific protocol. The participants completed these sets of activities twice, giving us enough data to compare the collected heart rate between the two watches. Each activity was further divided into different stages, including the Rest Stage, Dominant Hand Active Stage (D-Active Stage), and Non Dominant Hand Active Stage (N-Active Stage). The heart rate differences between each watch during the same activity and the same stage of all activities were evaluated. We also investigated how relative heart rate accuracy was affected by skin tone, and if we could tell which hand the watch was being worn, being the user's dominant or non dominant hand.

During the experiment, each subject wore a MB and a HW on the wrist of their dominant hand. Care was taken to follow proper wear guidelines as suggested for each device in order to collect the most reliable data possible. Each participant did a series of timed activities including cutting vegetables, electric tooth brushing, and walking along a given route. The participant was asked to follow timed instructions from the experiment instructor. The heart rate measurements of the two devices were stored in separate CSV files in their Bluetooth-connected smartphones to be processed for further analysis. After a close examination of the experiment's results, the vegetable cutting activity showed the largest heart rate differences among two devices, and the Dominant Hand Active Stage of cutting vegetables had the largest heart rate difference. Among all three test cases, electric tooth brushing shows the smallest heart

rate difference in both the rest and active stages, which indicated that the influence of high frequency vibration is smaller than the magnitude of movement. Statistical results show that the user's relative heart rate accuracy will be affected by daily activities even when a smartwatch is being worn on their non dominant hand. However, the influence is much smaller than if the watch is worn on the wrist of the user's dominant hand. Furthermore, the skin tone of the participant also shows some effect on the relative accuracy of optical heart rate sensor as well.

Based on the findings of these experiments, we discovered that a further exploration of the heart rate anomaly detection algorithm is required. This algorithm was used to identify the anomaly in the smartwatch's heart rate measurement while the user was completing an activity. The heart rate from the MB was compared with a pulse oximeter in order to tune the parameters of the anomaly algorithm. Data received from a separate test stage showed that the anomaly detection algorithm with tuned parameters can detect most of the heart rate anomalies identified by an examination of the heart rate signals.

ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Yan Sun, for the opportunity to complete this thesis. Her advice, support, and guidance helped me through the entire research process. Not only did she provide valuable academic guidance but also was a great mentor in my life. She told me how to balance work and family, and gave me tremendous support during the time we worked together. All of her advice and support has been greatly appreciated. I would also like to thank my committee members, Dr. Bin Li and Dr. Lisa DiPippo, for providing me with advice and reviewing my research. Lastly, I would like to thank Dr. Kunal Mankodiya for his guidance, advice, and motivation through all of our experiments and meetings, and in helping me to complete my research.

Numerous other individuals also provided assistance in my research. Everyone from the Wearable Biosensing Lab, especially Andrew Peltier, had assisted me in setting up lab equipment and helped to collect data in every experiment. I would like to thank Dr. Frederick Vetter, the Electrical Engineering Graduate director, for his guidance and help in filling out all the required forms correctly, and Dr. Haibo He, the Dean of Electrical, Computer and Biomedical Engineering Department, for his guidance and help in completing forms as well as his guidance for machine learning related research.

None of this would have been possible without the continued support and encouragement from my parents, Zhongmin Dai and Mingyan Xu and my husband, Daxian Yun, and my sweet boys, Franklin and Benjamin. You are all the power source that carried me through all of the ups and downs, and give me the courage to finish

this thesis. Last and most important, I want to give my most heartfelt thanks to my grandma Guiyin Zhang. Thank you for raising me up and gave me everything you have without any doubt.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Over the past five years, wearable biosensor technologies have undergone rapid development and have shown advancements in the sport, fitness, and health industries. Wearable biosensors were initially developed for personal fitness and performance monitoring, which can provide basic real-time quantitative feedback of biometrics. The latest generation of devices, such as activity trackers, can provide immediate feedback on multiple biometrics related to the quality of the consumers' physical activity, health, and exercise (Lyons, Lewis, Mayrsohn, & Rowland, 2014). With the sensing capabilities of wearable biosensors improved, there has been an increased interest in their application in medical settings, specifically for disease management and preventive health behaviors monitoring. Additionally, clinicians are increasingly interested in capturing patient-reported outcomes, including the patient's current status, symptoms, and adverse events such as falls and heart attacks. Furthermore, the usefulness of wearable biosensor technologies have been studied in both outpatients and hospitalized patients as a means of enhancing routine monitoring, or as part of an early warning system to detect clinical deterioration (Pelizzo,G. Guddo,A. Aurora P, Annalisa D. S, 2018).

These wearable devices have achieved tremendous success, as the market has grown from 113.2 million shipments in 2017 to 222.3 million in 2021 with a compound annual growth rate (CAGR) of 18.4% according to latest IDC report (IDC 2017). Because of this, the pursuit for practical and accurate approaches to assess

personal health biometrics and physical activities continues to lay emphasis on wearable biosensor technologies. Optical blood flow sensing using photoplethysmography (PPG) techniques to measure heart rate proves to hold the most importance, as heart rate is one of the most critical biometrics that both personal and clinical consumers are interested in. PPG is a non-invasive method for the detection of heart rate, and is connected with the optical properties of vascular tissue using a probe, usually being LEDs. PPG sensors use LED lights to shine directly into the skin and interact with changes in blood volume to configure a heart rate. Heart rate is determined based on the theory that blood flow through the artery is inversely related to the amount of light refracted (Maeda, Y., Sekine, M., & Tamura, T. 2011).

OBJECTIVE AND SCOPE

In addition to functioning as a time tracking device, smartwatches act like a mini computer, and have numerous functions beyond showing time. A smartwatch is a wrist-worn ''general-purpose, networked computer with an array of sensors''. Smartwatches have the potential to transform health care by supporting and evaluating health in everyday living because they are familiar to most people, and enable near-real time, continuous monitoring of physical activity and physiological measures without interference of consumers. Also, smartwatches support tailored messaging and reminders, and enable communication between patients, family members, and health care providers. Overall, smartwatches with PPG-based heart rate detection allows for unobtrusive and objective monitoring of physical exertion as well as systematic exercise prescriptions. There has been a steady focus on improving overall PPG performance, and the use of PPG technology for heart rate monitoring has shown

2

acceptable validity. However the variability is large among different devices, indicating that the accuracy of these trackers is dependent on the specific device used and the type and intensity of the activity (Lee, J., Matsumura, K., 2013). Therefore, PPG-based activity trackers remain under scrutiny due to the number of extrinsic factors that may interrupt proper heart rate detection, such as ambient light, sweat, anatomical placement, movement, and skin contact force (Parak, J., & Korhonen, I. 2014 and Teng, X. F., & Zhang, Y. T. 2004). There is a present lack of scientifically-based validation studies on the accuracy of multi-sensor PPG activity trackers in daily activities other than exercises such as cooking, vacuuming, use of electric appliance, and others. The aforementioned commercial devices are no exception. Thus, future studies to assess the validity of heart rate measurements in daily activities from commercially available devices would be highly warranted.

In this study, a Microsoft Band and a Huawei Android smartwatch, both of which are commercially available, were used to conduct a series of experiments focusing on daily activities. The accuracy of the watch's PPG based heart rate sensor was thoroughly evaluated based on various experimental results. Furthermore, based on the findings of two the smartwatches' heart rate relative accuracy experiments, an anomaly detection algorithm was developed to detect wearable sensor heart rate disturbances in daily activities.

CHAPTER 2

REVIEW OF LITERATURE

Evolving from simple pedometers, consumer-oriented electronic devices, such as smartphones with the help of apps and other services, are now able to capture a variety of parameters directly relevant to human health. With recent developments in micro-technology, data processing and storage, wireless communication and networking infrastructure, and battery capacity, wearable devices have made it possible for individuals to produce ever-larger streams of data across the lifespan, throughout the course of health and illness, and in a geospatial context (Mercer et al., 2016). (Chuah S H, 2016) developed a theoretical model to show the adoption of smartwatch acceptance. This research proved that the usefulness and visibility of a smartwatch are the most important factors that drive adoption intention. Smartwatches represent a type of fashinology, and the magnitude of the antecedents is influenced by an individual's perception of viewing a smartwatch as a technology or as a fashion accessory.

(V. P. Cornet, 2018) conducted a systematic review of smartphone-based passive sensing for health and wellbeing. There were thirty-five papers reviewed, among which most of them used the Android operating system and an array of smartphone sensors. The studies show that smartphone-based passive sensing for health and wellbeing demonstrated promise, and invited continued research and investment. Existing studies suffer from weaknesses in research design, lack of feedback and clinical integration, and inadequate attention to privacy issues. (B. Reeder, A. David, 2016) provided another systematic review of smartwatch uses for health and wellness,

which included seventeen studies published between 2014 and 2016. Their studies involved participants with illnesses such as Parkinson's disease, epilepsy, and diabetes, all of which were given consumer-grade smartwatches that collected their biometrics. The conclusion of this study was that consumer grade smartwatches have penetrated the health research space rapidly since 2014. However, the smartwatches' technical functions, acceptability, and effectiveness in supporting health must be validated in larger field studies that enroll more participants living with the conditions these studies target.

(Lewis K., 2017) focused on validating the heart rate data of the PPG sensor in two devices, the Basis Peak and Fitbit Charge HR, by conducting an experiment in which each participant had to complete a set of five minute exercises. These exercises included low intensity cycling, high intensity cycling, walking, jogging, running, arm raises with self-selected resistance, lunges with self-selected resistance, and planking. The heart rate accuracy of each device was compared to a criterion device and Pearson-Correlation. Mean absolute difference and Bland-Altman method were applied to each comparison as well. The conclusion of this study was that both devices perform with better accuracy during periods of rest and low physical exertion. Additionally, there were device-specific discrepancies in performance across various exercises. Meanwhile, another researcher (Stahl, S. E. 2016) provided a similar experiment with more devices. In this experiment, participants had to wear six different devices (Scosche Rhythm, Mio Alpha, Fitbit Charge HR, Basis Peak, Microsoft Band and TomTom Runner Cardio wireless HR monitors) while walking and running at 3.2, 4.8, 6.4, 8.0 and 9.6km/h respectively, with the heart rate data

being compared to a criterion measurement. The results shows that all six devices show good validity, with the criterion device and wearable device having the potential to overcome the limitations of the traditional chest strap. (Fukushima, H., 2012) provided a heart rate estimation by using a wrist-type photoplethysmography (PPG) sensor while their subject was running. An algorithm that estimated heart rate from the PPG sensor was proposed in the study. The algorithm utilized the built in accelerometer to gain knowledge of the subject's body motion and arm position to improve the heart rate accuracy. Two components were used in their method. One of which was rejecting artifacts with the power spectrum's difference between PPG and acceleration obtained by frequency analysis. The other was the reliability of heart rate estimation defined by the acceleration. Results had shown that the heart rate from a PPG sensor had a higher degree of usability compared to existing methods using ECG. Similar for all three studies, these studies failed to provide analysis of the facts that influence the accuracy of heart rate. Also, the studies only involved exercise movement. The analysis of daily activities was never involved in any of these works.

In addition to previous studies, (Kroll R.R., Boyd JG, 2016 and Kroll, R.R.; McKenzie, E.D., 2017) and (Pelizzo,G., Guddo,A. Aurora P, 2018) were able to test the accuracy of wearable devices on hospitalized patients and validate the optical sensor in wearable devices by comparing them to hospital grade devices. Kroll R.R. conducted a 24 hour heart rate monitoring with personal fitness trackers on 50 stable patients in the ICU. He found that the personal fitness trackers' derived heart rates were slightly lower than those derived from hospital grade cECG monitoring, and that they perform even worse in patients that are not in sinus rhythm. Pelizzo G. was able

6

to collect heart rate data from a Fitbit Charge HR on 30 patients. The patients were admitted to the Pediatric Surgery Unit for minor elective laparoscopic or open surgical procedures, which were performed under general or local anesthesia. The heart rate from Fitbit Charge HR was accurate, and matched with cECG and SpO2R monitoring during pediatric surgical procedures. However, the accuracy of optical heart rate sensor was evaluated in a stable environment, and there were no movements involved in these experiments. Since movement is usually considered as a major factor that can influence the accuracy of a wearable device, their studies failed to evaluate the heart rate accuracy of patients in active motion.

Besides these studies, (Lee, J., Matsumura, K., 2013) finished a study which evaluated the influences of different LED lights used in PPG sensors. A comparison of the HR measured by electrocardiography (ECG) with HR measured by 530 (Green), 645 (red), and 470 nm (blue) wavelength light PPG during baseline and performing hand waving was conducted. There were 12 participants enrolled in the experiment and the HR measured by ECG was used as a reference. The results showed that the limit of agreement in Bland-Altman plots between the HR measured by ECG and HR measured by 530 nm light PPG (±0.61 bpm) was smaller than 645 and 470 nm light PPG (±3.20 bpm and ±2.23 bpm, respectively). The ΔSNR (the difference between baseline and task values) of 530 and 470nm light PPG was significantly smaller than ΔSNR for red light PPG 645 (red). They conclude that 530nm light PPG could be a more suitable method than 645 and 470nm light PPG for monitoring HR in normal daily life.

In order to further evaluate the abnormal values in the heart rate of the smartwatch's optical heart rate sensor, an anomaly detection algorithm is necessary for any clinical usage of smartwatches. Abnormal heart rate values may be the result of many reasons in wearable devices, such as hardware faults, corrupted sensors, energy depletion, calibration, electromagnetic interference, signal fading, disrupted connectivity, sweat from the wearer, a detached sensor, etc. All of these factors can lead to a faulty diagnosis, and need to be excluded to reduce false alarms and unnecessary intervention of healthcare professionals. (O. Salem, Y. Liu, 2013) proposed a new framework for anomaly detection in medical wireless sensor networks which is based on the Mahalanobis distance for spatial analysis, and a kernel density estimator for the identification of abnormal temporal patterns.

One problem with this technique is its high dependency on the predefined threshold of MD. An appropriate threshold is quite difficult to figure out, and a single threshold may also not be suitable for outlier detection in multidimensional data. According to a statement in the work, the proposed framework can update the statistical parameters and obtain more a precise evaluation of the normal state of the patient. According to the experiment, the proposed approach can achieve good detection accuracy with a low false alarm rate (lower than 5.5%) on both real systems and synthetic medical datasets. Another piece of research from (O. Salem, Y. Liu, 2014) proposed a lightweight online anomaly detection framework which uses a smartphone as a base station. Haar wavelet decomposition, non-seasonal Holt-Winters forecasting, and the Hampel filter for spatial analysis were deployed in this framework. The framework was tested on real physiological datasets and proved that

both efficiency and reliability were improved by this framework. (S. H. Liou, Y. H. Wu, Y. S. Syu, 2012) introduced an anomaly detection algorithm which took the advantage of the regularity of ECG to detect ECG anomaly. The proposed method could explore the intrinsic signal structure and represent the ECG segments on a low dimensional space. The normal ECG segments will constitute a manifold, and the anomaly could be detected automatically. However, this method is focused on the regularity of ECG signals rather than the heart rate measured in a given period. Furthermore, (M. Haescher, D. J. C. Matthies, 2015) conducted a study using a smartwatch as a wearable device to detect anomaly activities of three different scenarios. These scenarios are the detection of sleep apnea, the detection of epileptic seizures, and the detection of accidents such as falling or car crashes. This study presents how to use a smartwatch as a base device to detect abnormal activities rather than to identify abnormal measurement of heart rate date from smartwatches.

ANALYSIS OF LITERATURE

An important observation to note is how the heart rate accuracy is evaluated throughout the various publications. In previous research, the accuracy of an optical heart rate sensor was validated either during designed exercises or by analyzing a hospital patient in stable and calm conditions. There is no publication research that examined how a wearable device's optical sensor performs in an individual's daily life, and how the accuracy of heart rate calculations can be affected by daily activities. Table 1 shows a summary of current research and their lack of data on the heart rate accuracy of wearable devices in daily activities.

To further investigate how wearable devices can be used in health monitoring or early discovery of certain diseases, the performance of wearable devices in daily activities needs to be further evaluated.

For further exploration, we also reviewed various anomaly detection algorithms used in the biosensors of wearable devices. In this thesis, we use CUSUM to explore the possibility of detecting the heart rate abnormalities measured in daily activities.

Table 1. Literature comparison of wearable device biosensor accuracy studies

Research focus	Pros/Cons
Systematic review of wearable devices biosensor usefulness	Only provide general information of wearable device research works.
	Provide research guidance for future research works
	Lack of details about specific device or activities
Accuracy of heart rate sensor of wearable device on various exercise activities (walk, jog, run, lift etc.)	Only focus on specific exercise activities, fail to provide heart rate accuracy analysis on general activities.
	Exercise activities only include small portion of everyone's daily activities. For some older population, exercise is not fit the use case of wearable devices
Accuracy of heart rate sensor of wearable device in hospital patient usage	Good pioneer research work on real patients
	In hospital patients are in quiet and stable state where wearable device are designed to have good performance
	Fail to provide heart rate data of patients with movements
Need to find: Accuracy of heart rate sensor of wearable device in daily activities Algorithm to detect heart rate anomaly of wearable device	Objective of this thesis to evaluate the accuracy of heart rate sensor of wearable device in daily activities and discover various facts that can have impact on heart rate sensor validation
	Further exploration on anomaly detection of wearable device heart rate data

CHAPTER 3

MATERIALS AND METHODS

3.1 Materials

A Microsoft Band and a Huawei Android smartwatch were used in this experiment. One of the reasons we choose the Microsoft Band is the sheer amount of tech that's built into it. It contains nearly every biosensor that a smartwatch could potentially contain. A complete list of sensors in the Microsoft Band include: Optical heart rate sensor, 3-axis accelerometer, 3-axis Gyrometer, GPS, Ambient light sensor, UV sensor, Skin temperature sensor, Capacitive sensor and Galvanic skin response sensor. However, not all sensors are used by the software yet, neither on the band nor in the accompanying phone app. For the purpose of this experiment, heart rate data from the optical heart rate sensor, movement data from the 3-axis accelerometer sensor, and skin response data from the Galvanic skin response sensor are streamed to a smartphone application via a Bluetooth connection. The reporting frequency of each type of data is different, and further resampling and interpolation are needed to get matching data for all sensors. The optical heart rate sensor uses a green LED light which measures the amount of light refracted in the blood vessels utilizing the PPG techniques mentioned in literature review. It's located on the back of the clasp of the band, and the band can be worn with the face either on the inside or on top of the wrist. On the other hand, the Huawei Android smartwatch only has basic physical tracking sensors, such as a 6-Axis motion sensor (Gyroscope and Accelerometer), an

Optical Heart Rate Sensor (PPG), and a Barometer despite being advertised as a fashionable watch. The optical heart rate sensor uses a green LED light to measure heart rate and all data can be streamed to a smartphone application via Bluetooth connection.

3.2 Experiment procedure and participants

This prospective experiment recruited six healthy adults between the ages of 22 and 50. Each subject was educated with the procedures first, and then underwent the same procedure through direct verbal communication. Participants gave verbal informed consent to three different activities which are cutting vegetables, tooth brushing with an electric toothbrush, and walking. These activities were chosen to represent the most common everyday activities, each with a differing amounts of movement. Cutting vegetables is a normal daily activity, but the movement is quite intense compare to tooth brushing and walking. For tooth brushing, we had the subjects use an electric toothbrush because of its high frequency vibration. The vibration involved can represent a set of daily activities that use electric appliances, such as vacuuming or shaving with an electric razor. Walking is the most common and moderate daily activity, and it can represent almost every activity that no intense movement involved. An instructor timed each activity and gave corresponding instructions to the participant throughout the whole experiment. Biosensor data was streamed to a phone application and stored in separated files automatically once the experiment started.

Each participant wore both smartwatches on the wrist of their dominant hand to get the same level of intensity of movement for each activity. Care was taken to follow the proper user guidelines as suggested by the manufacturer for each device. This is necessary to help make sure that the smartwatch was tight enough to steadily hold the optical heart rate sensor onto the subject's wrist. For the vegetable cutting and tooth brushing activities, data was collected twice from both the dominant hand and non dominant hand. These activities started with a one minute rest period, which is followed by a one minute active period with the given activity performed with the subject's dominant hand. This is followed by a half minute rest period followed by a one minute active period, this time with the same activity performed with the subject's non dominant hand. An extra half minute rest period was used to finish up the test cases. For the walk activity, the participant was asked to rest for one minute, then walk along a given route at a normal walking speed. Another one minute rest period was used to finish up the walk test case.

Figure 1. Procedure for cutting and tooth brushing test cases. There is no timer for walking test case

3.3 Smartwatch procedures

The Microsoft band and Huawei Android smartwatch were attached to the same wrist, this being the wrist of the subject's dominant hand in accordance to the manufacturer

instructions. We implemented this to study the difference of the biosensor data accuracy between the subject's dominant hand and non dominant hand. Data acquisition from each device was time-synced, with both being synced to the NTP time server. Heart rate data, accelerometer data, light data and Galvanic skin response (GSR) data were collected from the Microsoft Band using a data extraction software program specifically customized to sync with the Microsoft Band via Bluetooth transmission. For the Huawei Android smartwatch data acquisition, dedicated software was also designed to transmit the data to another smartphone via Bluetooth communication. Only heart rate data was captured, and upon completion of the testing protocol, the exercise metrics was imported into a Microsoft Excel spreadsheet. The mobile application settings for both devices were adjusted appropriately for each subject. Each device was confirmed to have full battery charge and proper functionality prior to testing.

3.4 Pre-process and metrics analysis

Time synced sensor data from each device was concurrently and continuously acquired by two separate smartphones and their corresponding applications. The reporting rate for each smartwatch differed from one another. For the Microsoft band, the heart rate reporting rate is one data point per second. The 3-axis accelerometer data was captured eight times per second, and the Galvanic skin response data was captured five times per second. For Huawei smartwatch, the heart rate was was only reported when the value changed. This data was stored in the smartwatch's corresponding smartphone, as spreadsheets and further process was required to align the data and resample them to same reporting rate.

To identify the heart rate difference between the active and rest stages, we divided the data of each test case into three segments, which are rest segment, dominant active segment, and non dominant segment. The rest segments are all the time slot that exist between activities. The dominant active segment is the time slot consisting of activity with the participant's dominant hand, while the non dominant segment is the time slot containing activity with their non dominant hand.

To evaluate the heart rate difference between two devices, four levels of metric analysis were implemented to give quantity analysis of various metrics.

1. The Student's T test compares the two averages and tells whether they are different from each other. The T test also indicates how significant the differences are. The larger the t score, the more of a difference there is between groups. Therefore, a large t-score indicates that the groups are more different from each other, while a small t-score indicates that the groups are similar. We explored the t-scores of the heart rate measurements from the Microsoft Band and the Huawei Smartwatch to examine the similarity of relative accuracy between the two optical sensors.

2. Root Mean Square Error (RMSE) is the standard deviation of the residuals, which are measures of how far from the regression line data points are. RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit.

3. The mean absolute differential between the heart rate of both the MB and the HW represented the average difference score regardless of direction of the difference.

4. The Bland-Altman method was used to further assess the agreement between the two devices for heart rate measurements and whether the difference varied in a

systematic or ambiguous way over the rage of measurements. The Bland-Altman calculates the mean difference between two methods of measurement (the 'bias'), and 95% limits of agreement as the mean difference. It is expected that the 95% limits include 95% of differences between the two measurement methods (Bland & Altman, 1986).

All four levels of analysis were implemented on the heart rate data from the three different activity stages. They were also used on the task-specific HR data and accelerometer data, which can be used to indicate the intensity of movement.

3.5 Results and discussion

After collecting the heart rate data from both smartwatches, data analysis was required to truly understand the differences between the two. First, an examination of the experiment results for each subject was conducted. This examination included the heart rate waveforms for all three test cases and the absolute difference of heart rate for each test case for both devices. The student's T-Test, Mean Absolute Difference (MAD), and Root Mean Square Root (RMSE) that were mentioned in the previous chapter were applied to the heart rate measurement of each test case.

Among all six subjects, we picked the results of subject one as an example to reveal the heart rate differences of the two devices in each of the three different test cases. In Figure 2, it's obvious that cutting vegetables had the most significant heart rate difference during the action periods, which was 33bps at time of 90 second. However, both the Mean Absolute Difference and RMSE of the walking test case are larger than the other test cases, which are 11.78bps and 14.94 respectively. The T-Score values agree with the MAD and RMSE, which is 11.98 for walking test case.

This was only a little bit larger than the cutting test case, and more than two times larger than brushing test case. Meanwhile, the electric tooth brushing test case has the smallest heart rate difference with the MAD and RMSE being 3.76bps and 4.79 respectively. This matches our expectation that the electric tooth brushing study is the one with smallest movement among these three activities.

From Figure 4, the heart rate measurement of HW is significantly higher than the measurement of MB. The mean bias is 2.9 ± 15 bpm over the heart rate measurement of MB.

Table 2. Heart rate date of Subject 1 of each test case.

Parameter	Cutting Vegetables	Electric Tooth brushing	Walking
Mean Absolute Diff	9.54	3.76	11.78
Max Diff	33	14.96	26.21
RMSE	13.86	4.79	14.94
T-Score	10.6	4.86	11.98

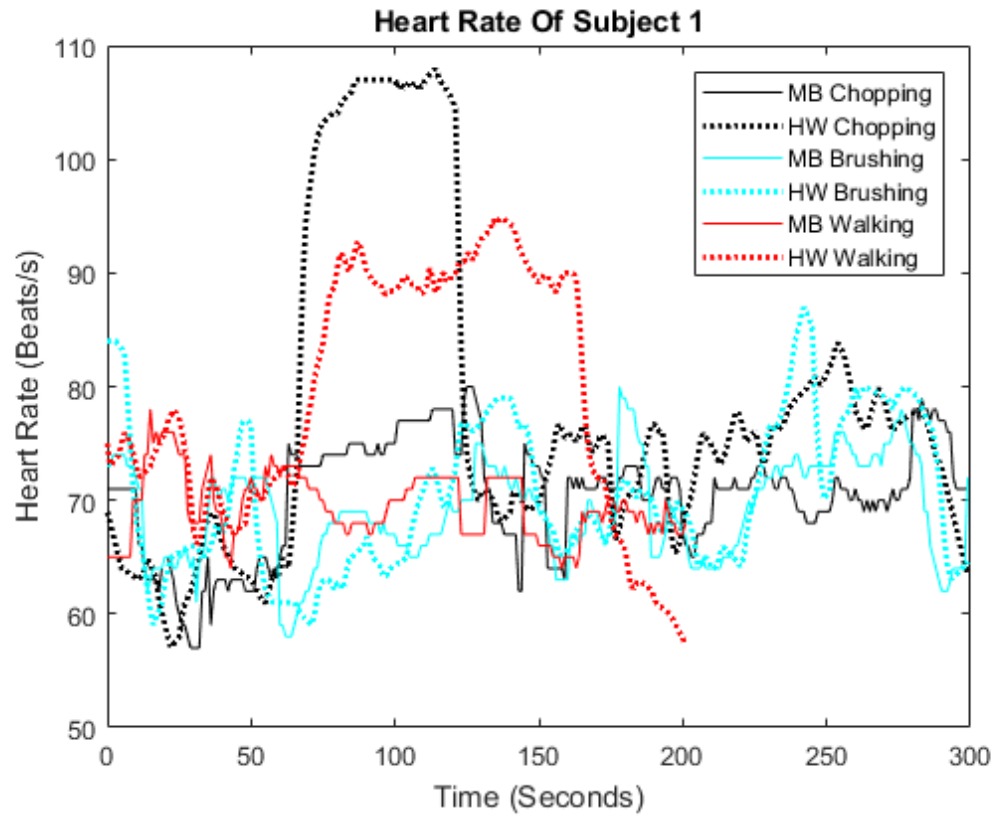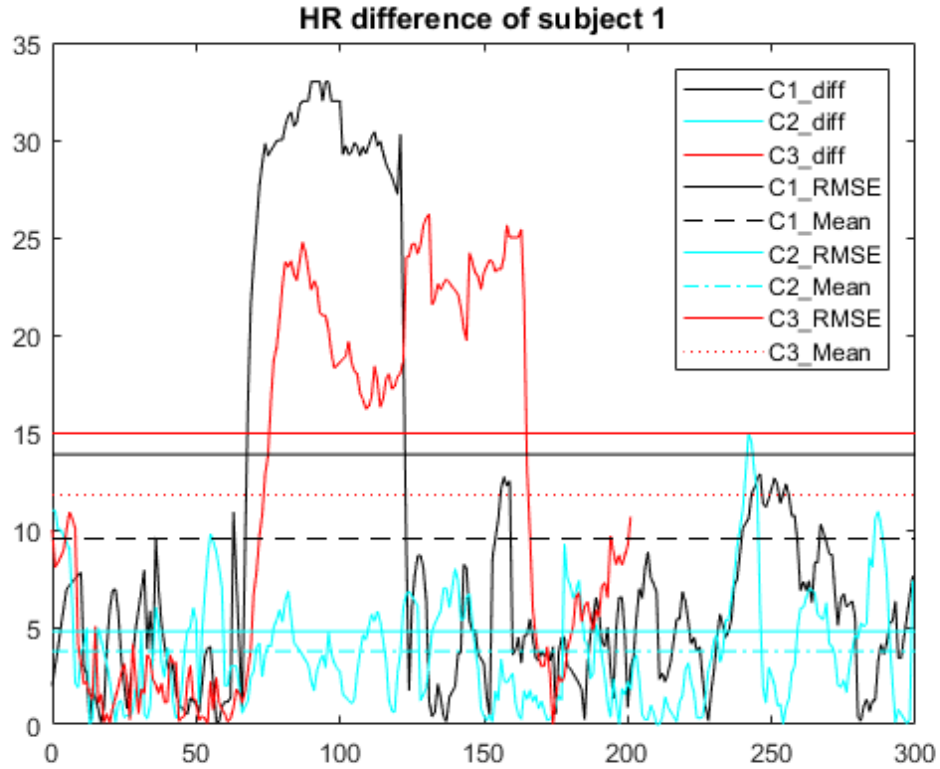Figure 2. Heart rate measurements of subject one in all test cases.

Figure 3. Absolute difference of heart rate measurement of Subject one of Microsoft Band and Huawei Android smartwatch. Both MAD and RMSE are in this figure.
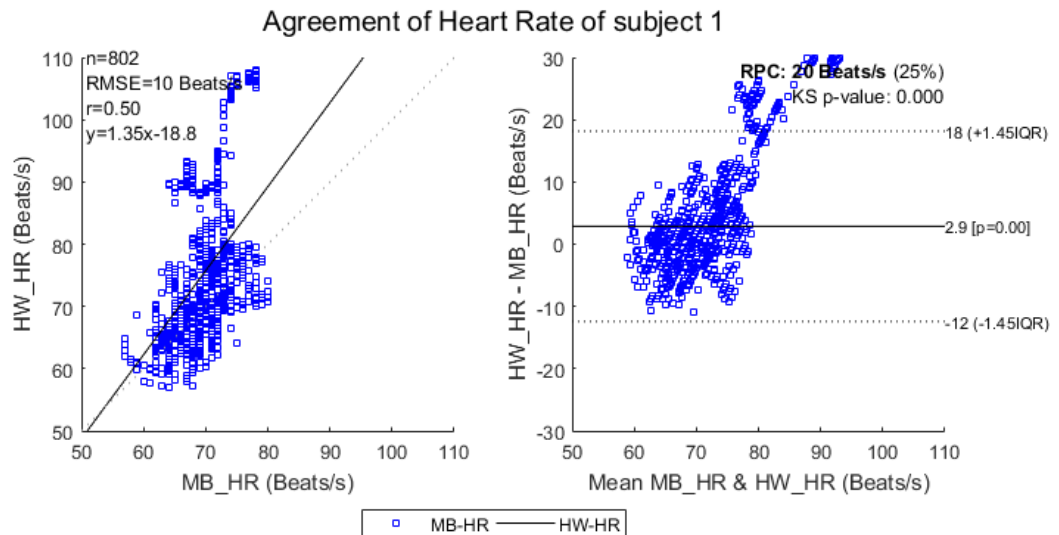

Figure 4. Results for heart rate of MB and HW of Subject one. Correlation between two test devices and Bland-Altman Plots indicating mean bias scores and 95% limits of agreement.

3.4.1 Heart rate difference of different stages in same activity

When examining time synced Microsoft Band and Huawei Android watch heart rate data in different action stages during same activities, the heart rate difference of the Action stages are larger than those of the Rest stages. Take the heart rate data of subject four as an example. For the cutting vegetables activity, the Mean absolute difference of the Active stage of the subject's dominant hand is 4.94 times higher (23.03bps VS 4.66bps) than the Rest stage, and the RMSE is 2.32 times higher (28.36bps VS 12.21bps) than Rest stage with the T-Score of D-Active stage being 2.5 times higher (9.74 vs 3.89) than Rest stage. The non dominant hand Active stage also shows heart rate differences when compared to the Rest stage, as the Mean absolute difference is 1.31 times higher. However, the RMSE and T-Score of the N-Active stage is smaller than those of the Rest stage. Meanwhile, the electric tooth brushing test shows similar heart rate differences for both its Active stages and its Rest stage. The Mean absolute difference is only 1.29 times higher, and RMSE is the same, with the T-Score being 2.2 times higher. The Rest and Active stages' heart rate differences of the walking test is higher than the other two tests, with Mean absolute difference being 4.69 times higher, and the RMSE being 3.15 times higher than the Rest stage. As for the Bland-Altman plot, the Rest stage and D-Active stage of the cutting vegetables activity is 0.91± 5bps, and 20 ± 35 bps while that of N-Active stage is 2.2 ± 6 bps. The Bland-Altman mean bias values shows that the heart rate measurement of the Rest and N-Active stages are higher than the D-Active stage. Overall, from all three test cases, we've found that heart rate differences between each device in the Active stage is larger than the Rest stage, which meets our expectation.

Table 3. Results of heart rate data of subject 4 in each stage

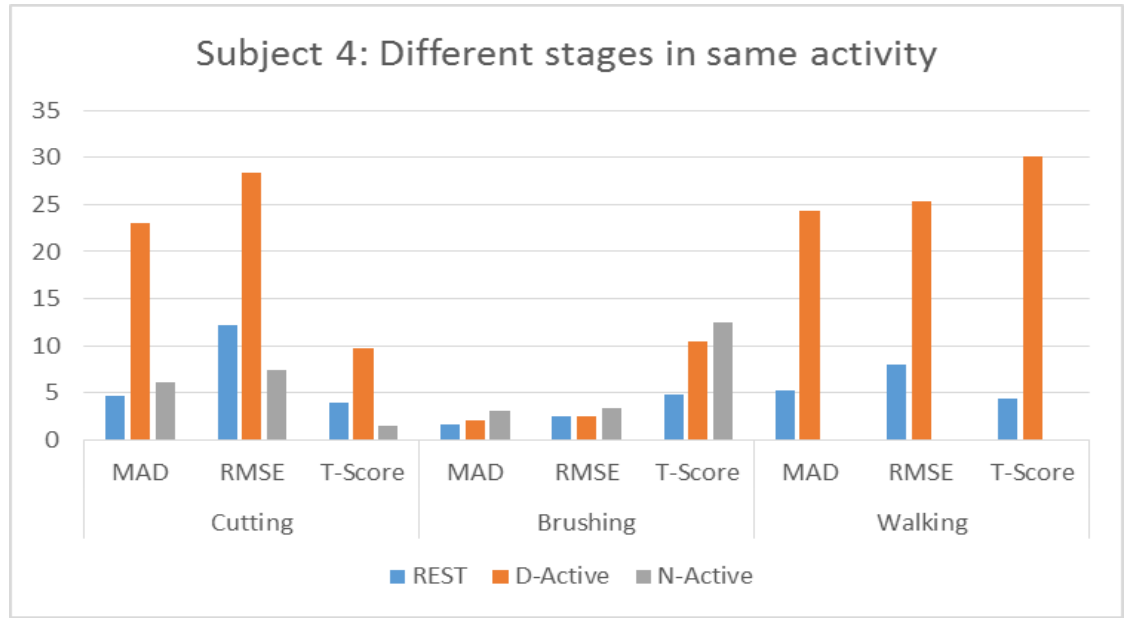|	Cutting			Brushing			Walking		
	MAD	RMSE	T-Score	MAD	RMSE	T-Score	MAD	RMSE	T-Score
REST	4.66	12.21	3.89	1.62	2.48	4.75	5.19	8.05	4.32
D-Active	23.03	28.36	9.74	2.1	2.48	10.45	24.34	25.4	30.1
N-Active	6.1	7.48	1.45	3	3.43	12.5			

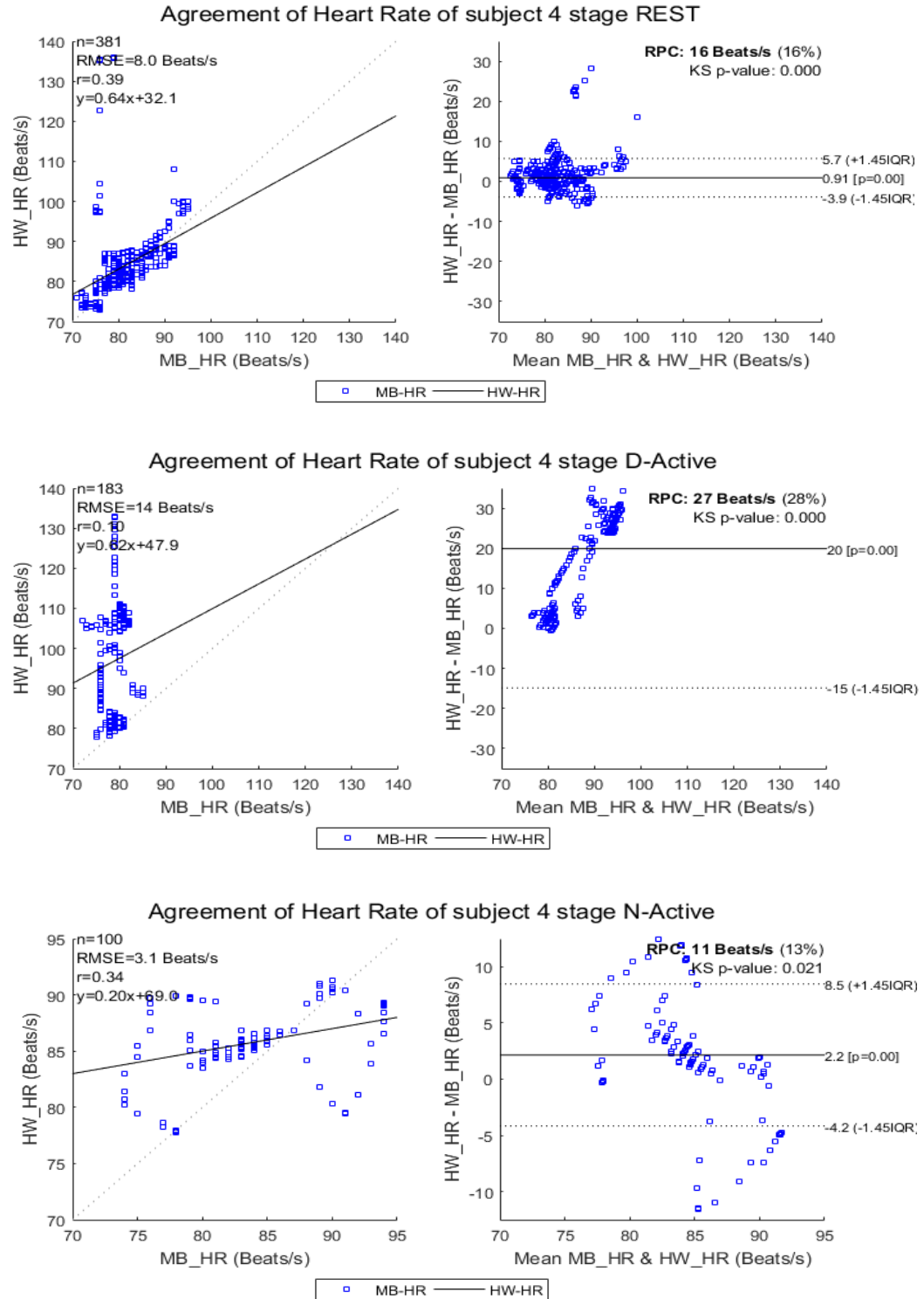Figure 5. MAD, RMSE and T-Score of heart rate difference in different stages of each test case.

Figure 6. Results for Rest stage, D-Active stage and N-Active stage heart rate differences of two devices. Correlation between two stages of each test case and case average of Bland-Altman Plots indicating mean bias scores and 95% limits of agreement.

3.4.2 Heart rate difference of same stage in different activities

When examining time synced Microsoft Band and Huawei Android smartwatch heart rate data of the same stages in different activities, we calculated three metric measurements, as mentioned in the previous chapter, and explored the results for each subject. In this chapter, we'll use the data from subject four as an example. The heart rate difference of all three Rest stages in the three test cases are almost all the same. The heart rate difference is less than 1bps for both MAD and RMSE. However, the T-Scores showed that even though the MAD and RMSE are almost the same, the actual signals are different from each other, as the cutting test case has a T-Score as high as 30.98, and the T-Score of the walking test case has a T-Score as low as 1.43. As for the Bland-Altman plot, the mean bias is -3.8 ± 7 bps, which indicates that the heart rate difference of the Rest stage in all three test cases has a very high agreement. However, when comparing the Active stages, the heart rate differences show some interest findings. From Figure 7 we can infer that the vegetable cutting test has the largest heart rate differences among all three tests. The Mean absolute difference of the tests are 58.82bps, which is 8.5 times higher than electric toothbrush test, and 4.26 times higher than walking test. This indicates that intensity and magnitude are the main factors of optical sensor accuracy, rather than vibration frequency. Also, the T-Score of the cutting vegetables case is much higher than the other two cases. For the Bland-Altman plot, the agreement of heart rate differences in different active stages are very limited. The heart rate differences in the Rest stages can be regarded as the basic device difference between Microsoft Band and Huawei Android watch. In that case, for the cutting vegetables test case, the heart rate differences of the active stage

is 8 times higher than the rest stage, while the electrical tooth brushing test case is almost same as the heart rate difference of the rest and active stages.

Table 4. Results of heart rate data of subject 3 for same stage of different activities

	Rest			D-Active			N-Active		
	MAD	RMSE	T-Score	MAD	RMSE	T-Score	MAD	RMSE	T-Score
Cutting	7.26	7.75	30.98	58.82	59.13	67.81	2.9	3.58	4.29
Brushing	6.24	7.92	12.08	6.82	7.87	9.37	16.16	28.95	4.02
Walking	5.45	8.35	1.43	13.81	14.93	21.78			
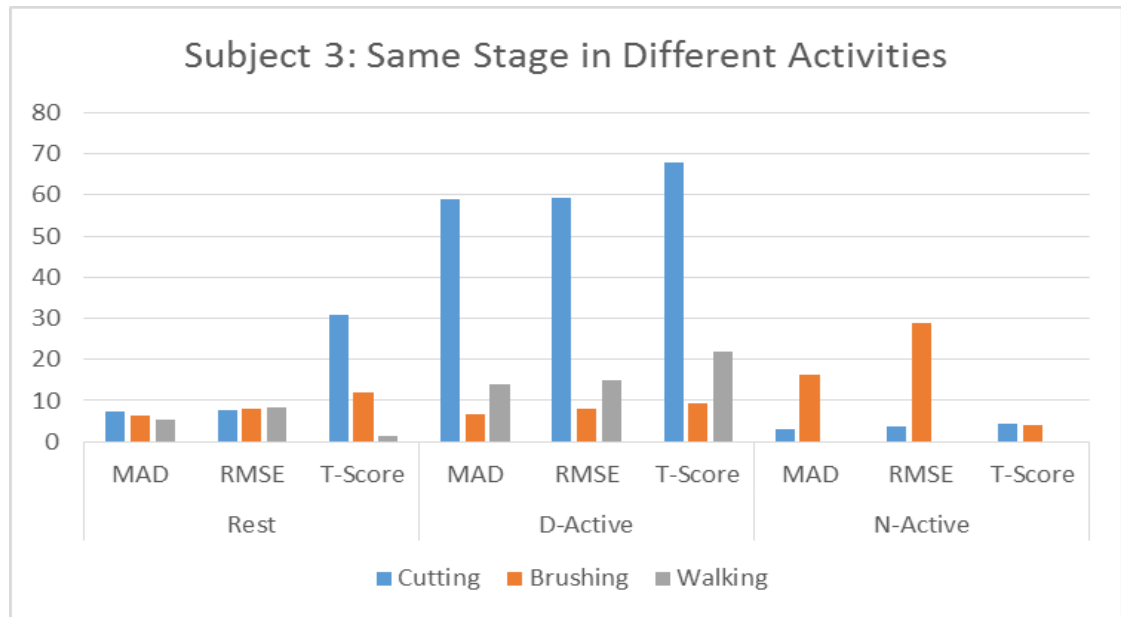
Figure 7. Mean absolute difference and RMSE and T-Score for same stage in different activities

Figure 8. Results for heart rate differences of two devices in Rest stage between case 1 and case 2. Correlation between two test cases of Rest stage and Bland-Altman Plots indicating mean bias scores and 95% limits of agreement.

Figure 9. Results for heart rate differences of two devices in D-Active stage between case 1 and case 2. Correlation between two test cases of D-Active stage and Bland-Altman Plots indicating mean bias scores and 95% limits of agreement.

3.4.3 Dominate hand activity VS Non-dominate hand active

When considering the relative heart rate accuracy of these two devices, the differences between dominant hand activities and non dominant hand activities is another factor that draws our interest. As one can imagine, it is normal that some people like to wear their watch on the wrist of their dominant hand, while others prefer to wear their watch on the wrist of their non dominant hand, but will this have any effect on the relative accuracy of optical heart rate sensors? From Figure 10, we choose the test results from subject 2 as an example. We found that in the electric tooth brushing test, dominant hand activity has a larger heart rate difference than non dominant hand activities over all three measurements. MAD, RMSE and T-Scores of the dominant hand Active stage are 2.36, 2.13 and 1.75 times higher that of the non dominant hand Active stage. As for the vegetable cutting test, the dominant hand activities and non dominant hand activities show similar results with the electric tooth brushing test case on MAD and RMSE measurements, with the T-Score of the dominant hand active stage being smaller than that of the non dominant hand active stage. It is obvious that drastic movement will affect optical sensor accuracy, but high frequency vibration with small magnitude has much less of an influence on the accuracy of an optical sensor, which matches our previous findings. However, wearing a smartwatch on one's non dominant wrist can reduce the influence of daily activities on its optical sensor's accuracy. The heart rate difference of N-Active stage is still two times higher than the heart rate difference of Rest stage.

By comparing Figure 11 and Figure 12, it's obvious that heart rate difference between the D-Active stage and the N-Active stage in the tooth brushing test case has

better agreement than the pair in the vegetable cutting test case, with the mean bias

being -8 ± 14 bps in the tooth brushing test case versus -20 ± 23 bps in the cutting
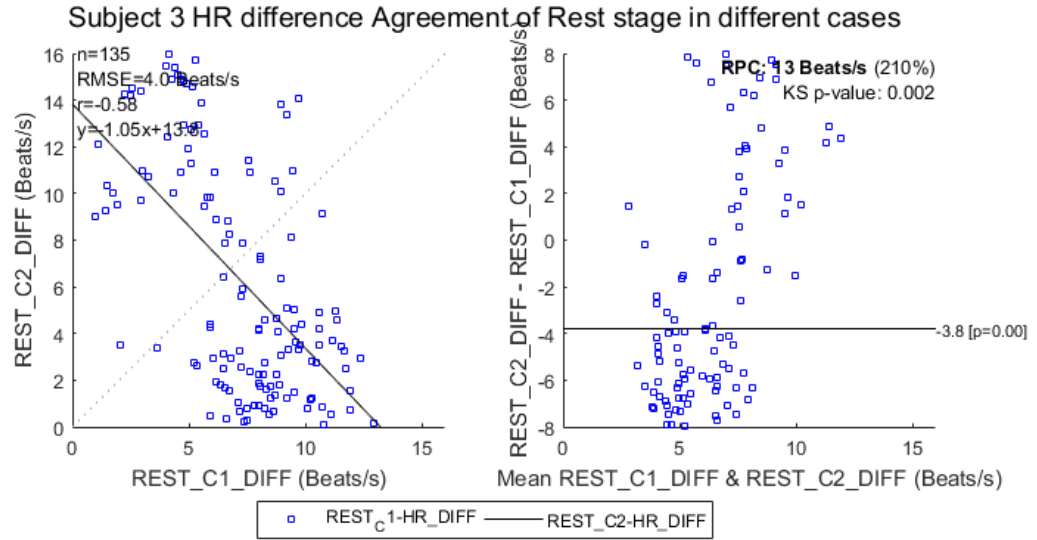
vegetables test case.
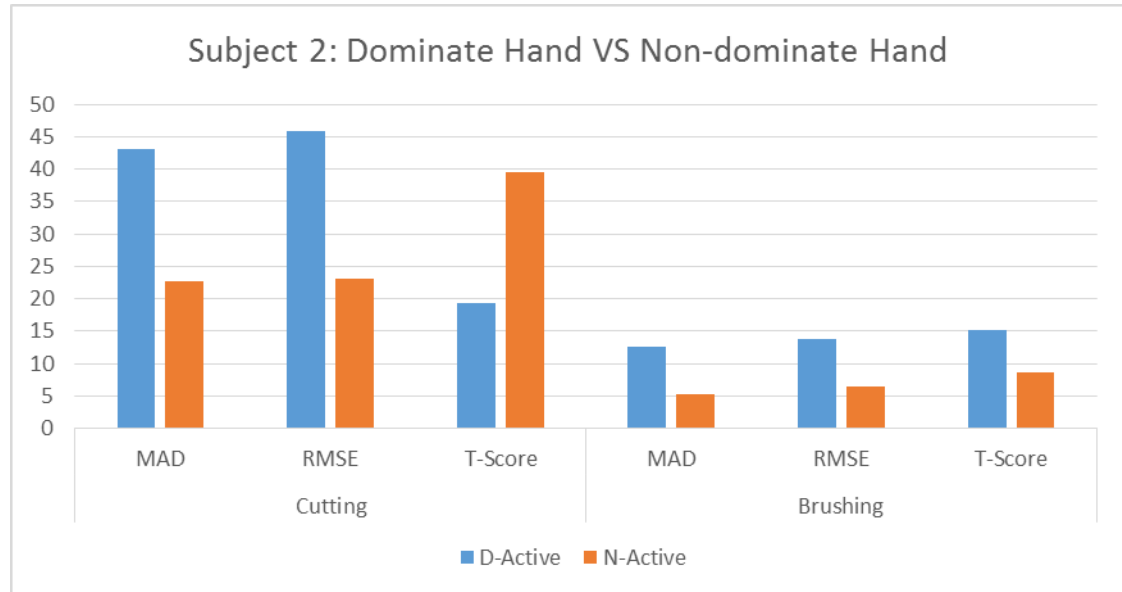
Figure 10. Mean absolute difference and RMSE for Dominate hand active stage and
Non-dominate hand active stage
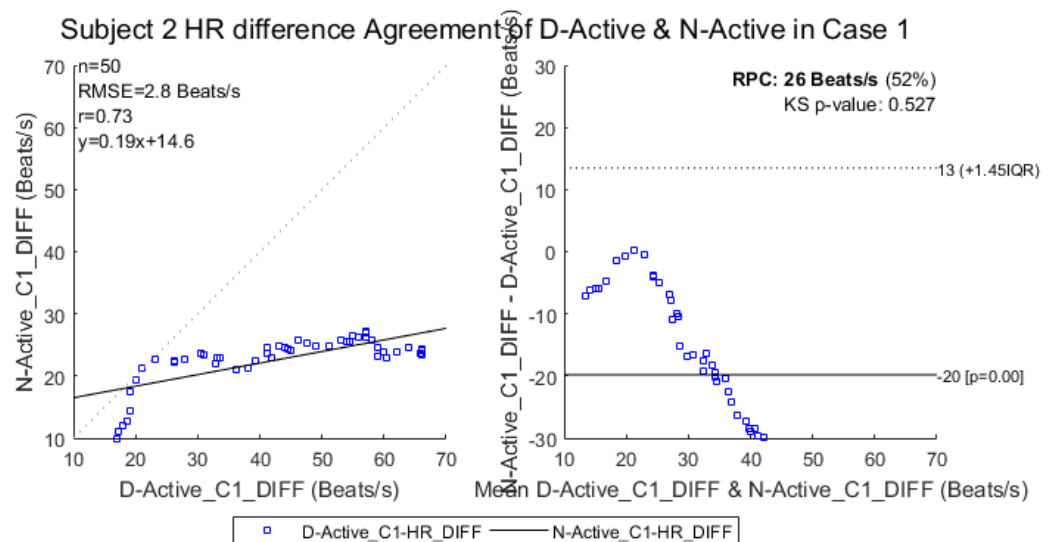
Figure 11. Results for heart rate differences of two devices in D-Active and N-Active
stage of case 1. Correlation between two stages and Bland-Altman Plots indicating
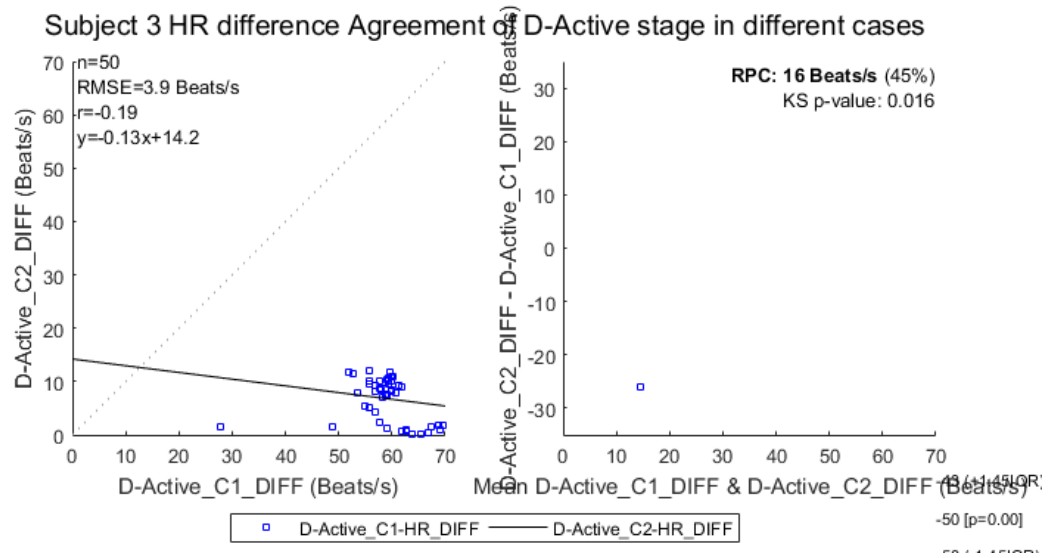mean bias scores and 95% limits of agreement.
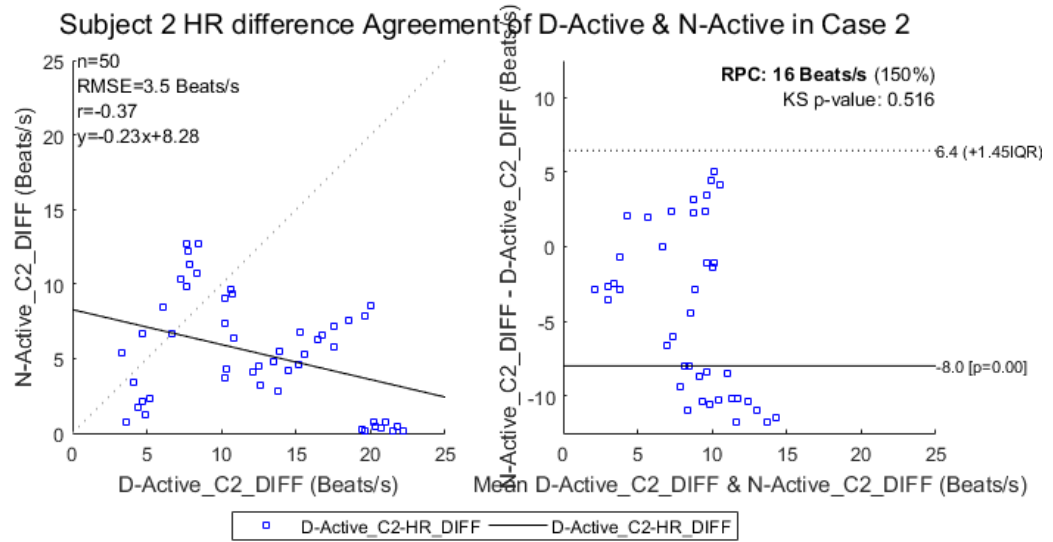
Figure 12. Results for heart rate differences of two devices in D-Active and N-Active stage of case 2. Correlation between two stages and Bland-Altman Plots indicating mean bias scores and 95% limits of agreement.

3.4.4 Relative heart rate accuracy in different skin tone

One more observation of the time synced heart rate data of the Microsoft Band and the Huawei Android smartwatch is that the relative accuracy of their heart rate is related to the skin tone of the participants as well. We divided six participants into three groups based on their skin tone, in which the first group (Group Blue) contains subjects one and two, both of them being Indian. The second group (Group Orange) has subjects three through five, all of which are Chinese. The sixth subject is Caucasian, and in a separate group. The data of third group was not included in this comparison since only one participant is in the group. From Figure 13, we found that for both the vegetable cutting test and the electric tooth brushing test, Group Blue has larger heart rate difference in both the Rest and Active stages. For the vegetable cutting test, the heart rate difference of Group Blue is 1.6 times larger than Group

Orange in the Rest stage, and 1.2 times larger than Group Orange in the Active stage. In the electric tooth brushing test, the heart rate difference of Group Blue is 1.6 times and 1.4 times larger than Group Orange on the Rest stage and the Active stage respectively. One explanation of this skin tone observation is that the optical sensors use the amount of refracted LED lights to determine the blood volume in a vessel. It's likely that the amount of an LED light absorbed by darker skin will be larger than that absorbed by lighter skin. Thus, even with same amount of blood volume, the amount of LED light refracted by darker skin will be much less than lighter skin. Because of this, the heart rate measurement of darker skin may be not as accurate as lighter skin. This chapter only provides some initial findings based on the experiments and participants we have. For more accurate conclusions, more participants should be recruited, and quantity analysis of skin tone should also been conducted as well.
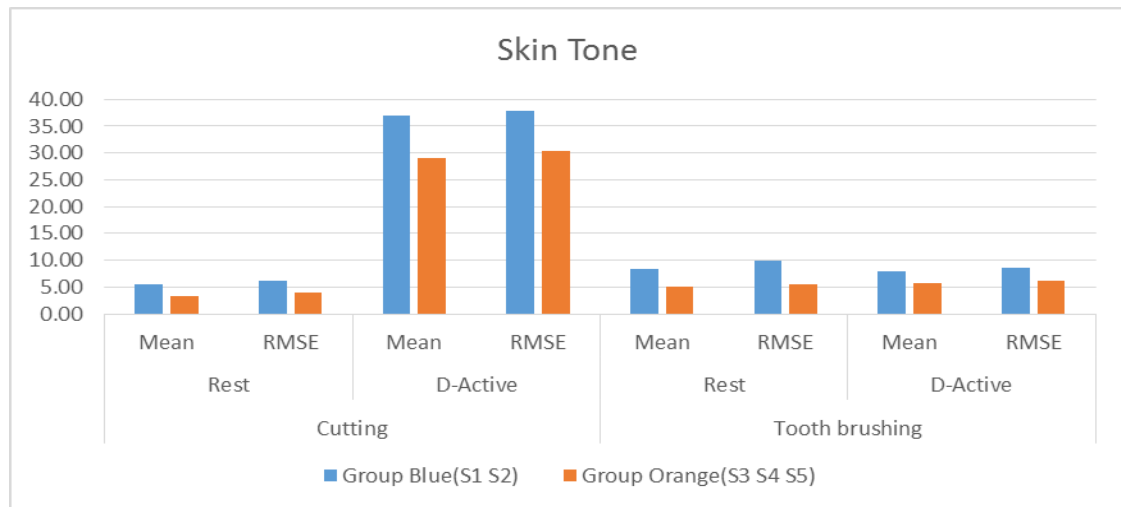
Figure 13. Mean absolute difference and RMSE of different skin tone groups

CHAPTER 4

FURTHER EXPLORATION

From chapter 3, it is obvious that the heart rate measurements of different wearable devices are different from each other, and same factors have different influences on different devices. Therefore, it is confirmed that there always exists some sort of disturbance in heart rate data read from any wearable device. In order to use heart rate data either in personal health monitoring or for clinical usage, it's extremely important to detect an anomaly of heart rate data, and mark that data as unreliable before applying it to any application or diagnosis program.

For a further exploration of the heart rate accuracy of wearable devices, we conducted more experiments on more daily activities and collected heart rate data from the Microsoft Band and a pulse oximeter device. The heart rate read from the pulse oximeter is proved to be very accurate when no activity is involved during the measure. We use this heart rate data as a criterion, and the heart rate data read from the Microsoft Band as test data to detect heart rate anomaly. A CUSUM algorithm was used in this exploration to detect heart rate anomaly. The parameters of CUSUM algorithm were tuned based on the heart rate pairs from both the Microsoft Band and the pulse oximeter, and more tests on heart rate of daily activities were performed to get the accuracy of the CUSUM algorithm.

4.1 Experiment design and procedure

For this experiment, two health participants were recruited, and each participant was informed of the procedures first. They then underwent the same procedures through direct verbal communication. Participants gave verbal informed consent to five different activities, which were hand tooth brushing, electric tooth brushing, chopping, vacuuming, and washing dishes. Just like the previous experiment, these activities were picked to represent common everyday activities with different activity characteristics. An instructor timed each activity, and gave corresponding instructions to the participant throughout the whole experiment. Biosensor data from the Microsoft Band was streamed to a phone application, and stored in separated files automatically once the experiment started.

Each participant wore the Microsoft Band on the wrist of their dominant hand, and a pulse oximeter was clipped to one of participant's toes. The pulse oximeter was kept stable, ensuring that there the pulse oximeter was not moved during all activities in order to get the most accurate heart rate readings as possible. Each activity was performed the same way for both the subject's dominant hand and their non dominant hand, just like the previous experiment. Starting with their dominant hand, the subject began the experiment by resting for 30 seconds. The resting period was followed by a one minute period of performing the specified activity. This cycle was repeated two more times, and ended with a final resting period, where they again rested for 30 seconds. The same procedure was performed by the subjects non dominant hand, totaling 9 minutes for the entire procedure for both hands. The whole procedure is figured in Figure 14 and the data from all of the sensors was streamed to separate

folders in a smartphone which the Microsoft Band was connected to via a reliable Bluetooth communication.
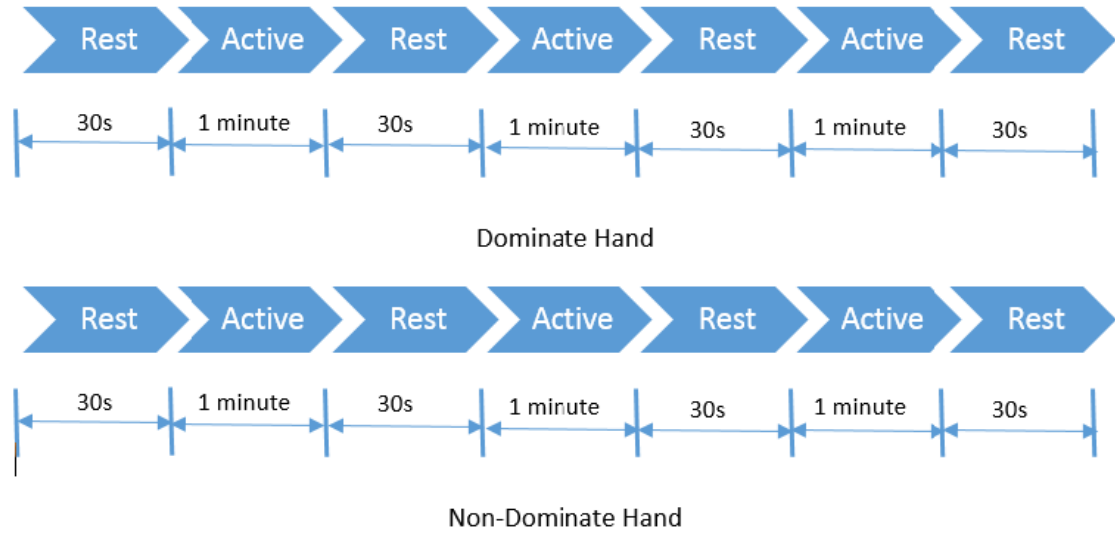
Figure 14. Procedure for anomaly detection experiment

Although we have a reliable method to record and process biosensor data from the Microsoft Band, there was no reliable method to stream the heart rate data from our pulse oximeter to a file. Because of this, we used a camera to record the display on the pulse oximeter throughout the entire experiment and saved the video file. We then read the heart rate reading from the video at an approximate reporting rate of one data point per second. The readings from the oximeter video were then time synced with the heart rate data from the Microsoft Band and resampled to match matrix sizes.
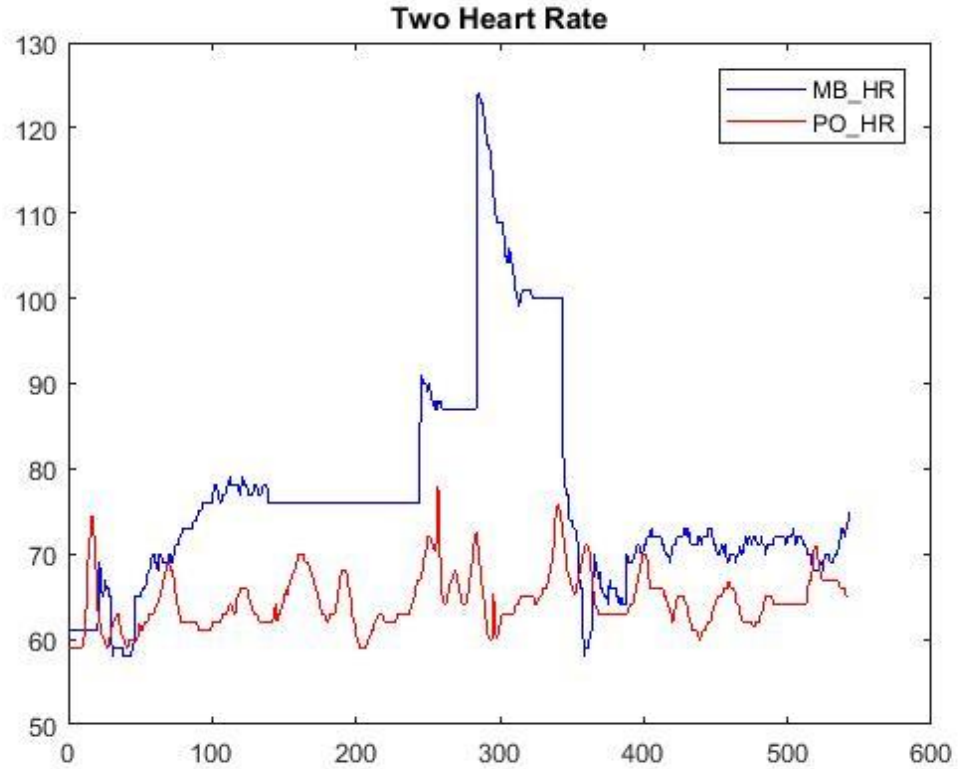
Figure 15. Heart of MB and Pulse Oximeter of hand tooth brushing test case

Figure 15 depicts the heart rate readings from both the MB and pulse oximeter from the hand tooth brushing test case in this experiment. It's clear from comparing the heart rate readings from the pulse oximeter with the smartwatch that the heart rate from Microsoft Band has anomaly readings at around time 250 seconds and 300 seconds, during which the heart rate reading of MB at time 250 seconds has a sudden increase of about 15 bps, and a sudden increase of 30 bps around time 300 seconds. This is just an example of the heart rate where the anomaly is obvious, and can be easily recognized by a consumer. However, there are many cases where the anomaly is not as obvious in this example, and we need a dedicated algorithm to detect and mark the unreliable heart rate data.

4.2 CUSUM anomaly detection

To detect anomalies in a wearable device's heart rate data, we used a recursive Cumulative Sum (CUSUM) algorithm as our first preferred method. The CUSUM detector has two advantages when compared to other change detectors. The first advantage is that CUSUM is not sensitive to the probabilistic distribution of the underlying signal, which is suitable to be applied on heart rate data. The second advantage is that it is proven to be optimal in terms of detecting changes faster than other methods. To have a better understanding of the CUSUM algorithm, we must first introduce the basics of the CUSUM algorithm. The CUSUM involves the calculation of a cumulative sum samples from a process x_n and determine whether the values of x_n has changed. To simplify the algorithm, we assume the distributions of x_n before and after change follow Gaussian distribution and the mean values of these two distributions are u_0 and u_1 respectively. Let x_j denote the jth sample of the data sequence. The basic CUSUM decision function is:

$$G_j = \max(G_{j-1} + (x_j - \frac{u_0 + u_1}{2}), 0) \tag{1}$$

$$T_s = \min\{k : G_j \geq h\} \tag{2}$$

Where G_{j-1} is the decision function at the sample $j-1$ and h is the change detection threshold. In this function, T_s is the stopping time, the time when the detector identifies a change occur and raise an alarm. Each time when $G_j \leq 0$ or $G_j \geq h$, there's a change detect ant the algorithm restarts by setting $G_j = 0$ and a new round of detection begins.

34

Initialization

Set d to the most likely change magnitude

Set the detection threshold $h > 0$

$S[-1] = G[-1] = 0$

Initialize the estimators $\overline{u_0}$ and $\overline{\sigma^2}$

$k = 0$

End

While the algorithm is not stopped **do**

Measure the current sample $x[k]$

Calculate the current estimates $\overline{u_0}[k]$ and $\overline{\sigma^2}[k]$

$$s[k] = \frac{d}{\overline{\sigma^2}[k]}(x[k] - \overline{u_0}[k] - \frac{d}{2})$$

$S[k] = S[k-1] + s[k]$

$G[k] = \max((G[k-1] + s[k]), 0)$

If $G[k] > h > 0$ **then**

$n_d = k$

$n_c = \arg\min_{1 \le n_c \le k} S[n_c - 1]$

Stop or reset the algorithm

End

$k = k + 1$

End

Figure 16. Basic CUSUM algorithm, Gaussian distribution case

Figure 16 depicts a basic CUSUM algorithm of a Gaussian distribution case in which there are several parameters that the user has to correctly set in order to to get optimal CUSUM performance:

1. The detection threshold h.

2. The change magnitude d.

The detection threshold h: The classical way to set this parameter is to use the average run length function, which is the expected number of samples before an action is taken, and more specifically, the mean time between false alarms. The average run length is zero and the mean detection delay. These two specific values of the average run length function depend on the detection threshold h, and can thus be used to set the performance of the CUSUM algorithm to a desired value for a particular application.

The change magnitude d: The user must have a prior knowledge about the signal to correctly set this parameter. Indeed, an efficient setting for the change magnitude is the a priori most likely change magnitude that should appear in the signal. In case several magnitudes of jump are possible, the best choice is the minimum one. In any case, the resulting change detection algorithm is only optimal to sequentially detect the chosen change magnitude.

4.3 Experiment result and discussion

4.3.1 CUSUM parameter tuning

As we initially explored CUSUM anomaly detection, we first tuned the CUSUM parameters based on the comparison of heart rate data from the Microsoft Band and the pulse oximeter. Figure 17 shows a comparison of normal heart rate readings and heart rate readings with anomalies in the hand tooth brushing test case. It's obvious that a normal heart rate ranges from 60bps to 80 bps for hand tooth brushing, while the anomaly heart rate has a sudden increase from 705bps to 90bps around time 250 seconds, and another sudden increase from 88bps to 125bps at time 300 seconds. With the comparison of the MB heart rate and the pulse oximeter, the sudden increase at

250 seconds and 300 seconds are two anomalies that should be marked as unreliable

heart rate readings.
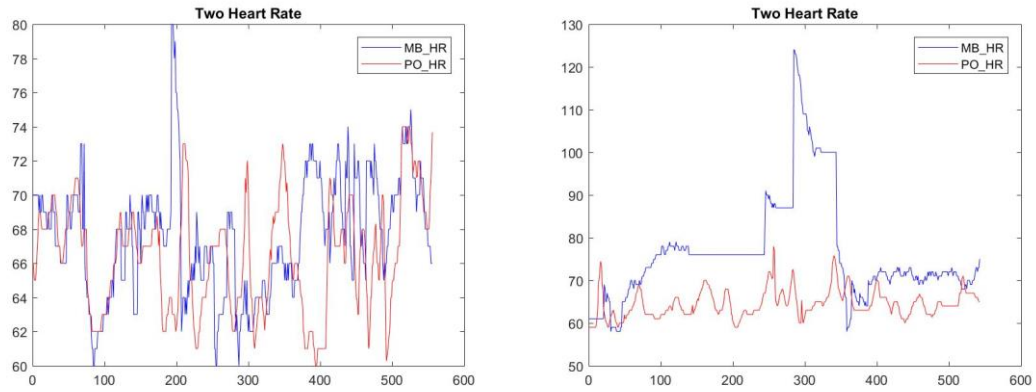
Figure 17. MB and Pulse Oximeter heart rate comparison of hand tooth brushing test
case of two participants with one has normal heart rate reading the other has anomaly
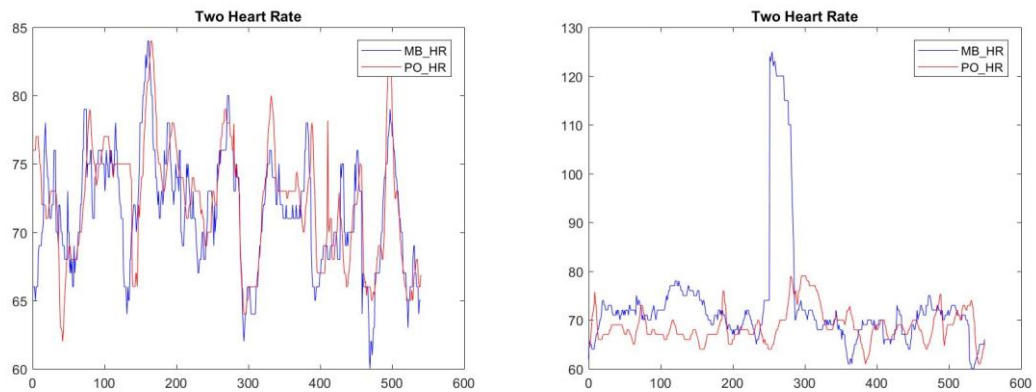
Figure 18. MB and Pulse Oximeter heart rate comparison of water flosser test case of
two participants with one has normal heart rate reading and the other has anomaly

From Figure 18, the comparison of the two heart rate readings from the water

flosser test show that the normal heart rate ranges between 60bps to 85bps, while the

anomaly shows a sudden increase from 70bps to 120bps at time 250 seconds. The goal

of the CUSUM algorithm is to detect these occasions of anomaly, and mark corresponding heart rate reading as unrealizable readings.

To tune CUSUM parameters based on our experiment results, we first consider the change magnitude *d*. We start with setting the slice window size to 5, which means we consider the heart rate readings in any 5 seconds period. To mark an anomaly in the heart rate readings over various activities, the first step is the get the range of normal heart rate changes. By collecting and comparing all training data sets, it's obvious that for all the normal heart readings, the change range is within 15 bps in any 5 second period, while for a heart rate with an anomaly, the change range is over 20bps, and sometimes even as high as 50bps in 5 seconds periods. Therefore, the change magnitude is set to 25. The detection threshold should be the mean time between false alarms of the average run length and the mean detection delay. As we set the slice window to 5, it's reasonable to set the threshold as twice of the slice window, which is 10 in our training data. To evaluate the result of the parameters we choose, Figure 19 demonstrated the CUSUM results with the detect threshold as 10, and change magnitude of 25 based on the heart rate data reading from the biosensors. Initial test results of the training set show that our CUSUM algorithm with tuned parameters can detect heart rate anomaly without false alarms.
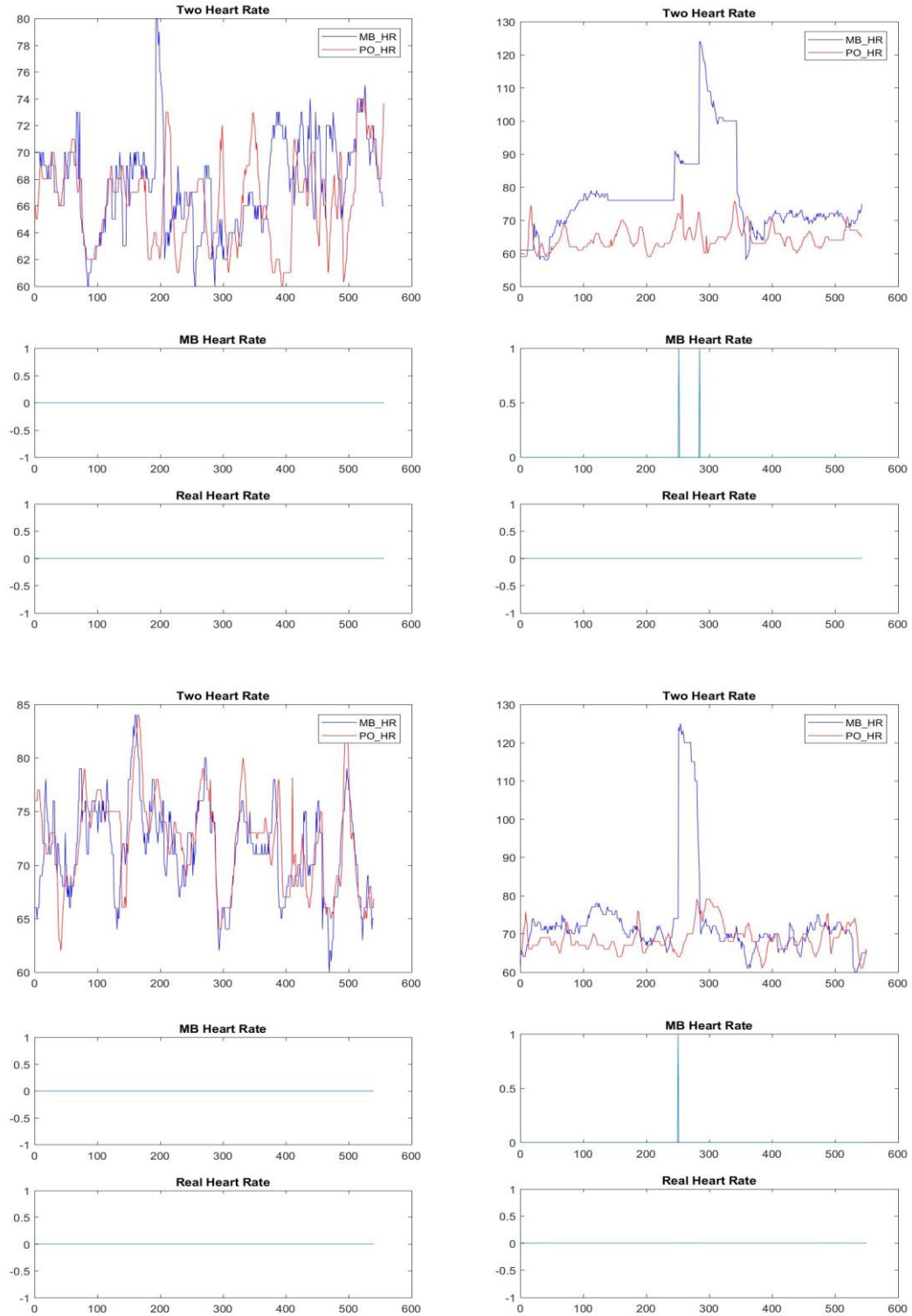
Figure 19, CUSUM anomaly detection result for heart rate shown in Figure17, 18. The results show the tuned CUSUM parameters can detect anomaly without false alarms.

4.3.2 CUSUM anomaly detection performance measurement

To test the performance of our CUSUM algorithm with tuned parameters, we applied the algorithm to all the data collected from our experiments, and all the results are pictured in Figure 20. There are two participants and five daily activities (Chopping (CH), Electric Tooth Brushing (ET), Manual Tooth Brushing (MT), Vacuuming (VA) and Washing Dishes (WD)) tested in this experiment. To examine the performance of the CUSUM algorithm, the data from the same activities was concatenated as one data stream, and the accelerometer data was also included to show the relationship between movement and heart rate anomaly.

For the chopping activity, there's an anomaly detected at time 840 seconds, where the heart rate has a sudden increase from 70 bps to 90 bps. At the moment of the anomaly, the participant was stopping the current activity and setting themselves to their resting position. There's no other heart rate anomaly found from the heart rate signal. We saw that the CUSUM detected the only heart rate anomaly without any false alarm in the chopping test case.

For the electric tooth brushing activity, there are three anomalies detected by the CUSUM algorithm. However, it seems that the first detection at time 260 seconds should be a false alarm since the heart rate only had an increase of 13bps. The other two detections successfully detected the anomaly. Furthermore, the third anomaly detected occurred at the time when the participant was is in resting, which is very abnormal.

For the manual tooth brushing activity, the only anomaly detected was at time 840 seconds. However, there seems to be another anomaly at time 280 seconds which was

missed by the anomaly detection algorithm. The anomaly was detected during the time when the participant transitioned from an active state to a resting state, which matched the anomaly detection in the chopping activity.

For the vacuuming activity, the anomaly detected at time 300 seconds should be a false alarm, where the heart rate only increased by around 12bps while two anomalies at time 680 seconds and 800 seconds were missed by the algorithm. The heart rate increases at these two moments were much larger than the ones detect by the CUSUM algorithm. Also, both of anomalies happened at the time when the participant was changing from an active state to a resetting state.

For the washing dishes activity, the anomaly detected at time 860 seconds is a valid heart rate anomaly, which also occurred at the moment when the participant was in transit from active to resting. However, another obviously anomaly at time 700 seconds was ignored, which has a heart rate at around 40 bps, and is clearly an anomaly.

Overall, the performance of CUSUM with tuned parameters is not as good as expected. Over the course of the experiment, five anomalies were successfully detected, with two false alarms and three anomalies missed by the algorithm. It's clear that a dedicate algorithm should be designed to tune the CUSUM parameters, and more training data is needed for the algorithm.
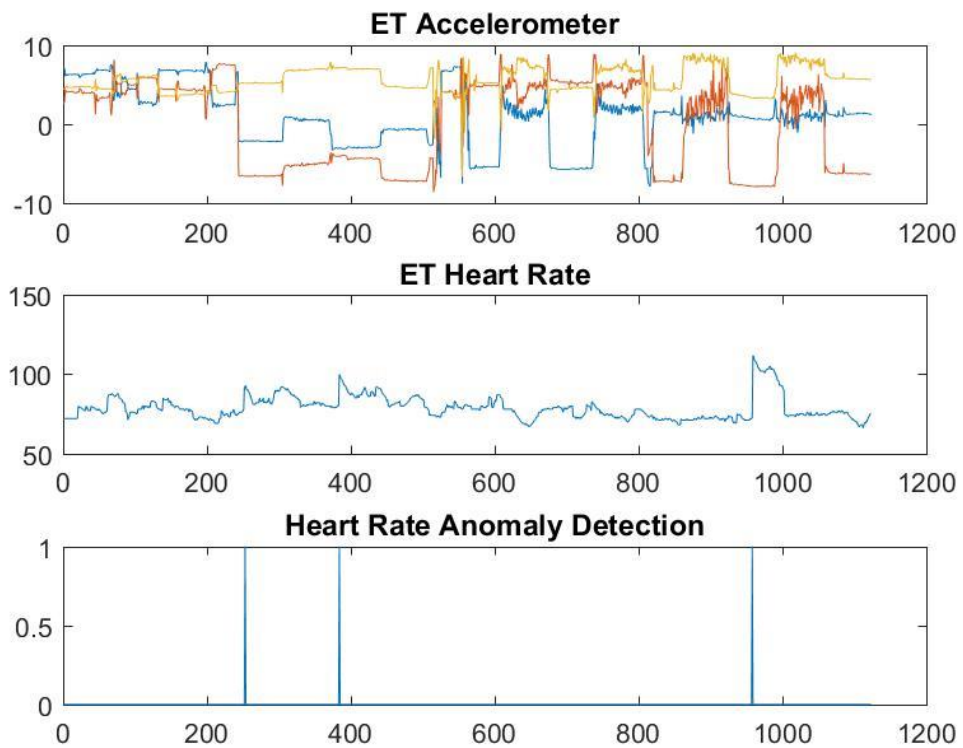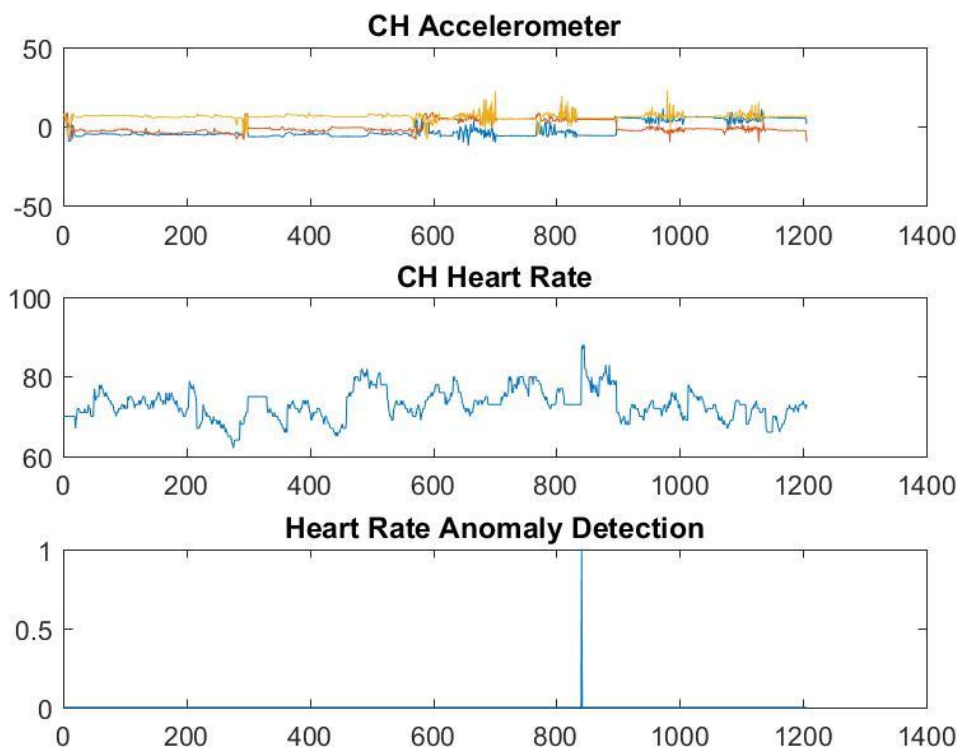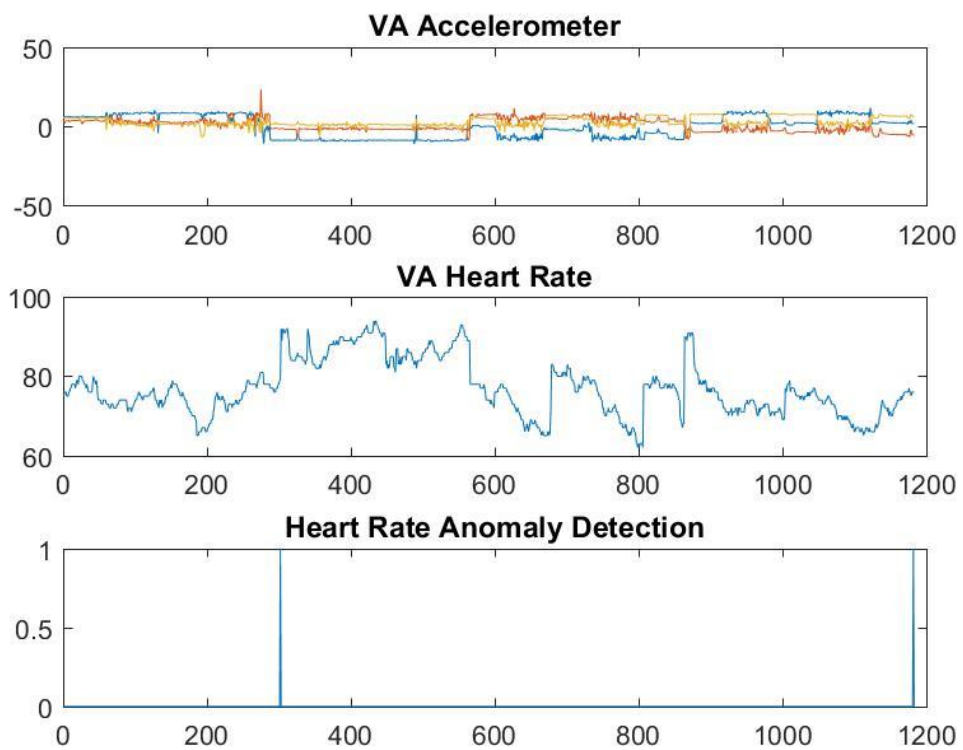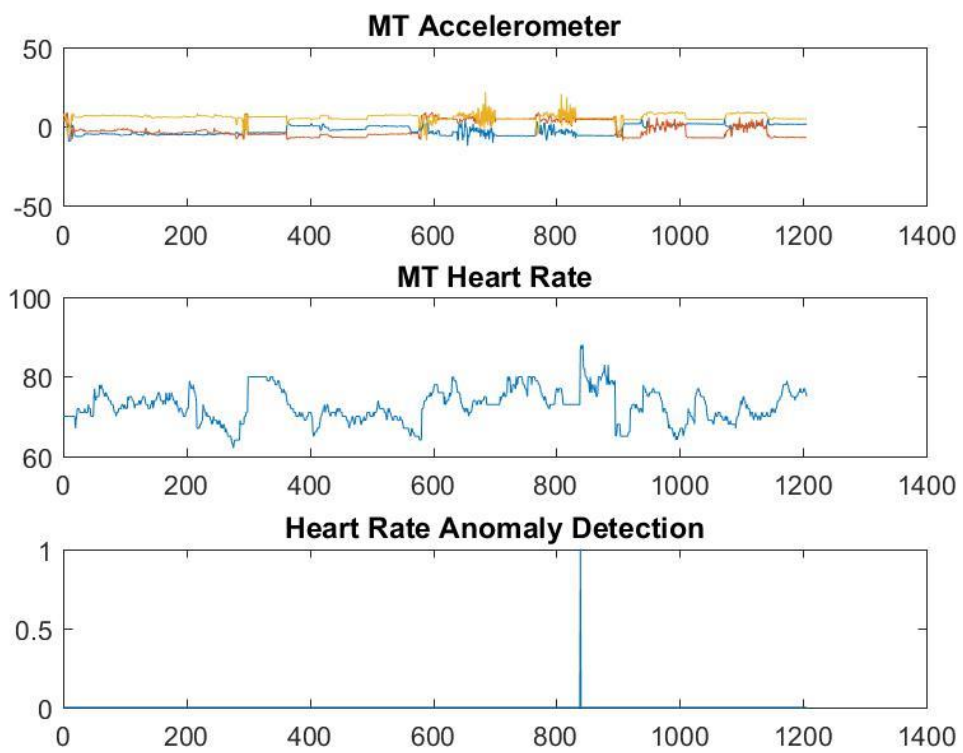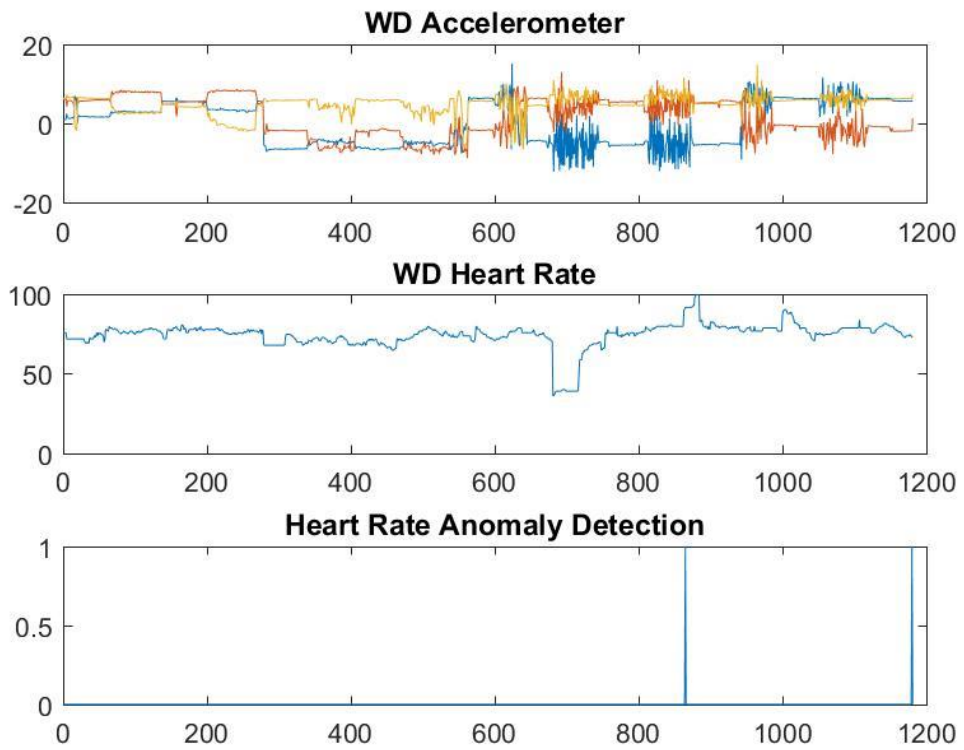
Figure 20. Heart rate anomaly detection results for an experiment with two participants and five daily activities.

CHAPTER 5

CONCLUSION AND DISUCSSION

The main objective of this study was to analyze the accuracy of PPG heart rate sensors from commercially available smartwatches, like the Microsoft Band and the Huawei Android smartwatch. PPG technology is relatively new, and has been applied on wearable devices to obtain consumer level heart rate monitoring. The inherent variability in accuracy may likely exist among various devices. Previous research shows that conditions of low physical exertion elicited the least variability in error values among various trackers. In this study, we focused on the performance of PPG sensors in daily activities, and discussed various factors that have influence on the accuracy of PPG heart rate sensor. From previous studies, it is obvious that movement plays a very important role in the accuracy of a wearable device's optical sensor, and in this thesis, the focus was on the influence of both magnitude and frequency of the movement. Our three test cases mimicked everyone's daily activities, and represented the three different types of movement which were included in the first experiment. Among them the vegetable cutting test has the largest magnitude and moderate frequency. The electric tooth brushing test has the highest movement frequency, but smallest magnitude, while the walking test represented the movements that have large magnitude but small frequency.

When considering different factors that have effects on the accuracy of a wearable device's optical sensor, we evaluated the influence of movement magnitude, movement frequency, and user preference of wearing the device on their dominant hand or non dominant hand, as well as skin tone of the participants. From the result of

our experiment, the Microsoft Band and the Huawei Android smartwatch have very large heart rate differences across all three of the test cases we conducted, with the largest differences being in the vegetable cutting test. The movement in this activity had the largest magnitude and relatively high frequency. The maximum heart rate difference is about 40bps, with both the Mean absolute value and RMSE being 9bps and 13.69 respectively. Meanwhile, the electric tooth brushing test case has the smallest heart rate difference, with the Mean absolute difference and RMSE being 3.86bps and 5.47bps, indicating that frequency has smaller a influence on heart rate accuracy than magnitude. We found that among all these factors, movement plays the most important role, and has the largest impact on the accuracy of optical heart rate data. Within movements, both the magnitude and frequency of the movements affect heart rate accuracy, with magnitude having a significant impact and frequency having barely any impact. We found that when the device was worn on the wrist of the user's dominant hand, the data showed larger heart rate differences, which matched our expectations. However, when the device was worn on the wrist of the user's non dominant hand, the readings were also affected by the movement on their dominant hand, although the influence is not as high. Furthermore, we found that skin tone may also have some impacts on the accuracy of optical heart rate sensors. Experiment results show that the heart rate differences of the participant group with a darker skin tone are larger than that of the participant group with a lighter skin tone in both the vegetable cutting test and the electric tooth brushing test.

Based on the findings from the first experiment, we further explored how to use the CUSUM algorithm to detect heart rate anomaly. More experiments with a variety

of daily activities were included in this to further explore this algorithm, and we first tuned CUSUM parameters with training test data. We then examined the performance of CUSUM on more experimental data. The CUSUM algorithm showed enough accuracy to detect basic heart rate anomaly. However, as part of our future work, more sophisticated algorithms need to be developed to have better detection performance.

LIMITATIONS AND FUTURE WORK

Though our study was successful as a primitive experiment on the optical sensor in wearable devices, and we were able to conclude multiple factors that have impacts on the accuracy of optical heart rate sensors, there are some obvious limitations of this study, and more improvements would need to be included in future work. The first limitation is the sample size, considering the fact that at the time of this study, there was only a small number of participants available. This definitely impairs the statistical significance of our research. With more participants, longer experiment times, and refined experiment protocols, it will be possible to have a valid method to record and calculate skin tone of participants, and to have a more convincing conclusion on the effect of skin tone on optical sensor variability. Also, more participants will provide the potential to group subjects with of different ages, races, health conditions etc., so that we could discover more factors under the hood, and all the conclusions of our study will be more convictive.

In addition to the sample size, more devices and more tests cases can be added to this research to have a more sophisticated and thorough understanding of the differences of each device, and how these devices perform in a much larger range of daily activities. More potential factors that may have influence on optical sensor

accuracy should be experimented and evaluated, such as the surface condition of the skin and how the tightness of smartwatch affects its heart rate accuracy.

Another possible improvement is to add a criterion device to provide golden heart rate data. Not only would we know the accuracy of each wearable device, but we'd also have the ability to identify the time frame and movement in which a heart rate anomaly occurs. In this situation, more mathematical algorithms could be used to detect the heart rate anomaly, and with the latest machine learning algorithms, it's even possible to classify each anomaly into different categories.

With regard to the anomaly detection algorithm, the CUSUM algorithm with tuned parameters is the initial step to further explore the reliability of optical heart rate sensor. To eliminate heart rate anomaly of wearable devices, an algorithm which can train itself with real time heart rate data and mark an unreliable heart rate in a device should be developed. Also, heart rate anomaly should be distinguished with heart rate characters of real heart disease.

APPENDICES

In this Appendix source code for all Matlab code are provided. These code are run on the user's computer and provide variety functionality from data process, metric analysis to figure plot.

1. Source code for sensor data pre-process and segmentation to different stages

```matlab
% Title:        Smart Watch Sensor data parse
% Created by:   Leichen Dai
% Date:         Aug 5th,2017
% Notes:        This file read csv file in data_dir folder.
%               1. remove header line
%               2. convert date time to seconds
%               3. write to xlsx file and plot the data
%               4. resample data based on given base
% no error support, or excel data.

function [] = parsecsv_sgmt(data_dir,sample_base)
%convert time to seconds. 24*60*60=86400
TIMESWITCH = 86400;
p1 = ["001","002"];
p2 = "003";
%# get all csv files in source directory
d = dir(data_dir);
isub = [d.isdir] & ~strcmp({d.name},'.') & ~strcmp({d.name},'..');
%# returns logical vector
subfolders = d(isub);
% Get a list of all files and folders in this folder.
for k = 1 : length(subfolders)
    fprintf('Sub folder #%d=%s \n', k,subfolders(k).name);
    sub_path = fullfile(subfolders(k).folder, subfolders(k).name);
    %# absolute-path filename
    fprintf('Sub folder name:%s \n', sub_path);
    fcsv =  dir(fullfile(sub_path,'*.csv'));
    for i = 1 : length(fcsv)
        fname = fullfile(fcsv(i).folder, fcsv(i).name);
        %# absolute-path filename
        [~,f] = fileparts(fname);
        %# used to name sheets in output
        fid = fopen(fname);
        if fid>0
            switch f
                case 'acc'
                    acc_d = textscan(fid,'%s %s %s %f %f
%f','Delimiter',',','HeaderLines',1);
                    ref = datenum([acc_d{1,2}{1,1} ' '
acc_d{1,3}{1,1}],'mm/dd/yyyy HH:MM:SS.FFF');
                    acc_s =
cell2mat([(datenum(strcat(acc_d{1,2},{' '},acc_d{1,3}),'mm/dd/yyyy
HH:MM:SS.FFF')-ref)*TIMESWITCH, acc_d(:,4:6)]);
```

49

```matlab
                case 'gsr'
                    gsr_d = textscan(fid,'%s %s %s
%f','Delimiter',',','HeaderLines',1);
                    ref = datenum([gsr_d{1,2}{1,1} ' '
gsr_d{1,3}{1,1}],'mm/dd/yyyy HH:MM:SS.FFF');
                    gsr_s = cell2mat([(datenum(strcat(gsr_d{1,2},{'
'},gsr_d{1,3}),'mm/dd/yyyy HH:MM:SS.FFF')-ref)*TIMESWITCH,
gsr_d(:,4)]);
                case 'hr'
                    hr_d = textscan(fid,'%s %s %s %f
%s','Delimiter',',','HeaderLines',1);
                    ref = datenum([hr_d{1,2}{1,1} ' '
hr_d{1,3}{1,1}],'mm/dd/yyyy HH:MM:SS.FFF');
                    hr_s = cell2mat([(datenum(strcat(hr_d{1,2},{'
'},hr_d{1,3}),'mm/dd/yyyy HH:MM:SS.FFF')-ref)*TIMESWITCH,
hr_d(:,4)]);
                case 'hhr'
                    hhr_d = textscan(fid,'%s %s %f
%s','Delimiter',',','HeaderLines',1);
                    ref = datenum([hhr_d{1,1}{1,1} ' '
hhr_d{1,2}{1,1}],'mm/dd/yyyy HH:MM:SS.FFF');
                    hhr_s = cell2mat([(datenum(strcat(hhr_d{1,1},{'
'},hhr_d{1,2}),'mm/dd/yyyy HH:MM:SS.FFF')-ref)*TIMESWITCH,
hhr_d(:,3)]);
                case 'light'
                    light_d = textscan(fid,'%s %s %s
%f','Delimiter',',','HeaderLines',1);
                    ref = datenum([light_d{1,2}{1,1} ' '
light_d{1,3}{1,1}],'mm/dd/yyyy HH:MM:SS.FFF');
                    light_s =
cell2mat([(datenum(strcat(light_d{1,2},{'
'},light_d{1,3}),'mm/dd/yyyy HH:MM:SS.FFF')-ref)*TIMESWITCH,
light_d(:,4)]);
                otherwise
                    disp(['Skip file:' fcsv(i).name]);
            end
            % close the file
            fclose(fid);
        end
    end

    f_acc_o = fullfile(sub_path, 'acc_orig.csv');
    csvwrite(f_acc_o,acc_s);
    f_gsr_o = fullfile(sub_path, 'gsr_orig.csv');
    csvwrite(f_gsr_o,gsr_s);
    f_hr_o = fullfile(sub_path, 'hr_orig.csv');
    csvwrite(f_hr_o,hr_s);
    f_hhr_o = fullfile(sub_path, 'hhr_orig.csv');
    csvwrite(f_hhr_o,hhr_s);
    f_light_o = fullfile(sub_path, 'light_orig.csv');
    csvwrite(f_light_o,light_s);

    acc_num_sec = length(acc_s(:,1))/acc_s(end,1);
```

```matlab
%# unique all data
    acc_s = unique(acc_s,'rows');
    gsr_s = unique(gsr_s,'rows');
    hr_s = unique(hr_s,'rows');
    hhr_s = unique(hhr_s,'rows');
    light_s = unique(light_s,'rows');

    % Get resample target number
    switch sample_base
        case 'acc'
            target = acc_s;
        case 'gsr'
            target= gsr_s;
        case 'hr'
            target = hr_s;
        case 'light'
            target = light_s;
        otherwise
            target = hr_s;
    end
    target_intp = target(:,1);
    %resample and clean resample errors
    acc_sd = interp1(acc_s(:,1), acc_s, target_intp,
'linear','extrap');
    gsr_sd = interp1(gsr_s(:,1), gsr_s, target_intp,
'linear','extrap');
    hr_sd = interp1(hr_s(:,1), hr_s, target_intp,
'linear','extrap');
    hhr_sd = interp1(hhr_s(:,1), hhr_s, target_intp,
'linear','extrap');
    light_sd = interp1(light_s(:,1), light_s, target_intp,
'linear','extrap');
    f_acc = fullfile(sub_path, 'acc_smp.csv');
    csvwrite(f_acc,acc_sd);
    f_gsr = fullfile(sub_path, 'gsr_smp.csv');
    csvwrite(f_gsr,gsr_sd);
    f_hr = fullfile(sub_path, 'hr_smp.csv');
    csvwrite(f_hr,hr_sd);
    f_hhr = fullfile(sub_path, 'hhr_smp.csv');
    csvwrite(f_hhr,hhr_sd);
    f_light = fullfile(sub_path, 'light_smp.csv');
    csvwrite(f_light,light_sd);
    %find match case 1 and case 2
    if endsWith(subfolders(k).name, p1)
        save_data_1(sub_path, acc_sd, 'acc');
        save_data_1(sub_path, gsr_sd, 'gsr');
        save_data_1(sub_path, hr_sd, 'hr');
        save_data_1(sub_path, hhr_sd, 'hhr');
        save_data_1(sub_path, light_sd,'light');
        save_data_o1(sub_path, acc_s, 'acc', acc_num_sec);
    elseif endsWith(subfolders(k).name, p2)
        save_data_2(sub_path, acc_sd, 'acc');
        save_data_2(sub_path, gsr_sd, 'gsr');
        save_data_2(sub_path, hr_sd, 'hr');
        save_data_2(sub_path, hhr_sd, 'hhr');
        save_data_2(sub_path, light_sd, 'light');
```

```matlab
        save_data_o2(sub_path, acc_s, 'acc', acc_num_sec);
    end
end

function save_data_1(path,fdata,name)
f_r1 = fullfile(path, [name '_rest_1.csv']);
csvwrite(f_r1,fdata(11:50, :));
f_t1 = fullfile(path, [name '_tran_1.csv']);
csvwrite(f_t1,fdata(51:70, :));
f_a1 = fullfile(path, [name '_act_1.csv']);
csvwrite(f_a1,fdata(71:110, :));
f_t2 = fullfile(path, [name '_tran_2.csv']);
csvwrite(f_t2,fdata(111:130, :));
f_r2 = fullfile(path, [name '_rest_2.csv']);
csvwrite(f_r2,fdata(131:150, :));
f_r3 = fullfile(path, [name '_rest_3.csv']);
csvwrite(f_r3,fdata(171:200, :));
f_t3 = fullfile(path, [name '_tran_3.csv']);
csvwrite(f_t3,fdata(201:220, :));
f_a2 = fullfile(path, [name '_act_2.csv']);
csvwrite(f_a2,fdata(221:260, :));
f_t4 = fullfile(path, [name '_tran_4.csv']);
csvwrite(f_t4,fdata(261:280, :));
f_r4 = fullfile(path, [name '_rest_4.csv']);
csvwrite(f_r4,fdata(281:295, :));
end

function save_data_2(path,fdata,name)
f_r1 = fullfile(path, [name '_rest_1.csv']);
csvwrite(f_r1,fdata(11:50, :));
f_t1 = fullfile(path, [name '_tran_1.csv']);
csvwrite(f_t1,fdata(51:70, :));
f_a1 = fullfile(path, [name '_act_1.csv']);
csvwrite(f_a1,fdata(71:end-70, :));
f_t2 = fullfile(path, [name '_tran_2.csv']);
csvwrite(f_t2,fdata(end-70:end-50, :));
f_r2 = fullfile(path, [name '_rest_2.csv']);
csvwrite(f_r2,fdata(end-50:end, :));
end

function save_data_o1(path,fdata,name,num)
f_r1 = fullfile(path, [name '_rest_1_orig.csv']);
csvwrite(f_r1,fdata(floor(11*num):floor(50*num), :));
f_t1 = fullfile(path, [name '_tran_1_orig.csv']);
csvwrite(f_t1,fdata(floor(51*num):floor(70*num), :));
f_a1 = fullfile(path, [name '_act_1_orig.csv']);
csvwrite(f_a1,fdata(floor(71*num):floor(110*num), :));
f_t2 = fullfile(path, [name '_tran_2_orig.csv']);
csvwrite(f_t2,fdata(floor(111*num):floor(130*num), :));
f_r2 = fullfile(path, [name '_rest_2_orig.csv']);
csvwrite(f_r2,fdata(floor(131*num):floor(150*num), :));
f_r3 = fullfile(path, [name '_rest_3_orig.csv']);
csvwrite(f_r3,fdata(floor(171*num):floor(200*num), :));
f_t3 = fullfile(path, [name '_tran_3_orig.csv']);
csvwrite(f_t3,fdata(floor(201*num):floor(220*num), :));
```

```
f_a2 = fullfile(path, [name '_act_2_orig.csv']);
csvwrite(f_a2,fdata(floor(221*num):floor(260*num), :));
f_t4 = fullfile(path, [name '_tran_4_orig.csv']);
csvwrite(f_t4,fdata(floor(261*num):floor(280*num), :));
f_r4 = fullfile(path, [name '_rest_4_orig.csv']);
csvwrite(f_r4,fdata(floor(281*num):floor(295*num), :));
end

function save_data_o2(path,fdata,name,num)
f_r1 = fullfile(path, [name '_rest_1_orig.csv']);
csvwrite(f_r1,fdata(floor(11*num):floor(50*num), :));
f_t1 = fullfile(path, [name '_tran_1_orig.csv']);
csvwrite(f_t1,fdata(floor(51*num):floor(70*num), :));
f_a1 = fullfile(path, [name '_act_1_orig.csv']);
csvwrite(f_a1,fdata(floor(71*num):floor(end-70*num), :));
f_t2 = fullfile(path, [name '_tran_2_orig.csv']);
csvwrite(f_t2,fdata(floor(end-70*num):floor(end-50*num), :));
f_r2 = fullfile(path, [name '_rest_2_orig.csv']);
csvwrite(f_r2,fdata(floor(end-50*num):end, :));
end
end
```

2. Source code to plot average heart rate signals

```
% Title:       Smart Watch Sensor data parse
% Created by:  Leichen Dai

function [] = similarity_case_plot(data_dir)
p1 = "001";
p2 = "002";
p3 = "003";
d = dir(data_dir);
%# returns logical vector
isub = [d.isdir] & ~strcmp({d.name},'.') & ~strcmp({d.name},'..');
subfolders = d(isub);
% Get a list of all files and folders in this folder.

    function plot_total(name)
        mh_t1 = zeros(300);
        mh_t2 = zeros(300);
        mh_t3 = zeros(300);
        hh_t1 = zeros(300);
        hh_t2 = zeros(300);
        hh_t3 = zeros(300);
        t1 = 0;
        t2 = 0;
        t3 = 0;
        for k = 1 : length(subfolders)
            fprintf('Sub folder #%d=%s \n', k,subfolders(k).name);
             %# absolute-path filename
            sub_path = fullfile(subfolders(k).folder,
subfolders(k).name);
            mhr_csv = fullfile(sub_path,'hr_smp.csv');
```

```matlab
                mhr = csvread(mhr_csv);
                mhr = mhr(:,2);
                hhr_csv = fullfile(sub_path,'hhr_smp.csv');
                hhr = csvread(hhr_csv);
                hhr = hhr(:,2);

                if endsWith(subfolders(k).name, p1) %find match case 1
                    mh_t_c = min(length(mh_t1),length(mhr));
                    mh_t1 = mh_t1(1:mh_t_c) + mhr(1:mh_t_c)';
                    hh_t_c = min(length(hh_t1),length(hhr));
                    hh_t1 = hh_t1(1:hh_t_c) + hhr(1:hh_t_c)';
                    t1 = t1 + 1;
                elseif endsWith(subfolders(k).name, p2) %find match
case 2
                    mh_t_c = min(length(mh_t2),length(mhr));
                    mh_t2 = mh_t2(1:mh_t_c) + mhr(1:mh_t_c)';
                    hh_t_c = min(length(hh_t2),length(hhr));
                    hh_t2 = hh_t2(1:hh_t_c) + hhr(1:hh_t_c)';
                    t2 = t2 + 1;
                elseif endsWith(subfolders(k).name, p3) %find match
case 3
                    mh_t_c = min(length(mh_t3),length(mhr));
                    mh_t3 = mh_t3(1:mh_t_c) + mhr(1:mh_t_c)';
                    hh_t_c = min(length(hh_t3),length(hhr));
                    hh_t3 = hh_t3(1:hh_t_c) + hhr(1:hh_t_c)';
                    t3 = t3 + 1;
                end
            end
            mh_t1 = mh_t1/t1;
            mh_t2 = mh_t2/t2;
            mh_t3 = mh_t3/t3;
            hh_t1 = hh_t1/t1;
            hh_t2 = hh_t2/t2;
            hh_t3 = hh_t3/t3;

            t1_diff = abs(mh_t1 - hh_t1);
            t2_diff = abs(mh_t2 - hh_t2);
            t3_diff = abs(mh_t3 - hh_t3);

            t1_rmse = sqrt(mean((mh_t1 - hh_t1).^2));
            t2_rmse = sqrt(mean((mh_t2 - hh_t2).^2));
            t3_rmse = sqrt(mean((mh_t3 - hh_t3).^2));

            [pr1, pval1] = corr(mh_t1',hh_t1','type','Pearson');
            [pr2, pval2] = corr(mh_t2',hh_t2','type','Pearson');
            [pr3, pval3] = corr(mh_t3',hh_t3','type','Pearson');

            fprintf('C1: mean_diff=%f RMSE=%f max=%f pr=%f pv=%f\n',...
                mean(t1_diff),t1_rmse,max(t1_diff),pr1,pval1);
            fprintf('C2: mean_diff=%f RMSE=%f max=%f pr=%f pv=%f\n',...
                mean(t2_diff),t2_rmse,max(t2_diff),pr2,pval2);
            fprintf('C3: mean_diff=%f RMSE=%f max=%f pr=%f pv=%f\n',...
                mean(t3_diff),t3_rmse,max(t3_diff),pr3,pval3);
```

```matlab
        % Plot average heart rate for all three cases
        % plot_avg(mh_t1,mh_t2,mh_t3,hh_t1,hh_t2,hh_t3);
        % Plot difference of all heart rate data
        % plot_diff(t1_diff, t2_diff,
t3_diff,t1_rmse,t2_rmse,t3_rmse);
        % Plot Bland Altman heart rate data
        % BA plot paramters
        tit = 'Average Heart Rate Agreement'; % figure title
        territories = {'MB','HW'};
        states = {'HR'};
        % names of groups in data {dimension 1 and 2}
        gnames = {territories, states};
        % Names of data sets
        label = {'MB\_HR\_AVG','HW\_HR\_AVG','Beats/s'};
        % stats to display of correlation scatter plot
        corrinfo = {'n','RMSE','r','eq'};
        BAinfo = {'RPC(%)','ks'}; % stats to display on Bland-
ALtman plot
        limits = 'auto'; % how to set the axes limits
        colors = 'br';
        BlandAltman(mh_t1',
hh_t1',label,tit,gnames,'corrInfo',corrinfo,'baInfo',BAinfo,'axesLi
mits',limits,'colors',colors, 'showFitCI',' on');
        %BlandAltman(mh_t1', hh_t1',label,tit);
    end
    function plot_avg(mh_t1,mh_t2,mh_t3,hh_t1,hh_t2,hh_t3)
        fig1=figure('name','hr_case');
        plot(0:length(mh_t1)-1,mh_t1,'k');
        hold on;
        plot(0:length(mh_t2)-1,mh_t2,'c');
        hold on;
        plot(0:length(mh_t3)-1,mh_t3,'r');
        hold on;
        plot(0:length(hh_t1)-1,hh_t1,':k', 'LineWidth',2);
        hold on;
        plot(0:length(hh_t2)-1,hh_t2,':c', 'LineWidth',2);
        hold on;
        plot(0:length(hh_t3)-1,hh_t3,':r', 'LineWidth',2);
        title('Average HR of each case');
legend('ms\_hr\_C1','ms\_hr\_C2','ms\_hr\_C3','hw\_hr\
_C1','hw\_hr\
_C2','hw\_hr\_C3');
     %   saveas(fig1,[sub_path '/' 'hr_time.jpg']);
    end

    function plot_diff(t1_diff, t2_diff,
t3_diff,t1_rmse,t2_rmse,t3_rmse)
        fig3=figure('name','hr_case_diff');
        plot(0:length(t1_diff)-1,t1_diff,'k');
        hold on;
        plot(0:length(t2_diff)-1,t2_diff,'c');
        hold on;
        plot(0:length(t3_diff)-1,t3_diff,'r');
        hold on;
        rmse_1 = refline([0,t1_rmse]);
        rmse_1.Color = 'k';
        mean_1 = refline([0,mean(t1_diff)]);
```

```matlab
        mean_1.Color = 'k';
        mean_1.LineStyle = '--';
        rmse_2 = refline([0,t2_rmse]);
        rmse_2.Color = 'c';
        mean_2 = refline([0,mean(t2_diff)]);
        mean_2.Color = 'c';
        mean_2.LineStyle = '-.';
        rmse_3 = refline([0,t3_rmse]);
        rmse_3.Color = 'r';
        mean_3 = refline([0,mean(t3_diff)]);
        mean_3.Color = 'r';
        mean_3.LineStyle = ':';
        title('HR difference of each case');

legend('C1\_diff','C2\_diff','C3\_diff','C1\_RMSE','C1\_Mean','C2\_
RMSE','C2\_Mean','C3\_RMSE','C3\_Mean');
    end
plot_total('smp');
end
```

3. Source code of statistical analysis

```matlab
% Title:        Smart Watch Sensor data parse
% Created by:   Leichen Dai

function [] = similarity_compute(data_dir)
p1 = ["001","002"];
p2 = "003";
d = dir(data_dir);
isub = [d.isdir] & ~strcmp({d.name},'.') & ~strcmp({d.name},'..');
%# returns logical vector
subfolders = d(isub);
% Get a list of all files and folders in this folder.
for k = 1 : length(subfolders)
    fprintf('Sub folder #%d=%s \n', k,subfolders(k).name);
    sub_path = fullfile(subfolders(k).folder, subfolders(k).name);
    %# absolute-path filename
    if endsWith(subfolders(k).name, p1) %find match case 1 and case
2
        plot_diff(sub_path, 'rest_1');
        plot_diff(sub_path, 'act_1');
        plot_diff(sub_path, 'rest_2');
        plot_diff(sub_path, 'rest_3');
        plot_diff(sub_path, 'act_2');
%        plot_diff(sub_path, 'rest_4');
    elseif endsWith(subfolders(k).name, p2)
        plot_diff(sub_path, 'rest_1');
        plot_diff(sub_path, 'act_1');
        plot_diff(sub_path, 'rest_2');
    end

end
```

```matlab
    function plot_diff(sub_path,name)
        mh_name = ['hr_' name '.csv'];
        hh_name = ['hhr_' name '.csv'];
        mhr_r1 = fullfile(sub_path,mh_name);
        mhr_r1 = csvread(mhr_r1);
        hhr_r1 = fullfile(sub_path,hh_name);
        hhr_r1 = csvread(hhr_r1);
% Calculate mean absolute difference
        r1_diff = mean(abs(mhr_r1(:,2) - hhr_r1(:,2)));
% Calculate RMSE
        r1_rmse = sqrt(mean((mhr_r1(:,2) - hhr_r1(:,2)).^2));
% Calculate Pearson-Correlation
        [pmh, pval] =
corr(mhr_r1(:,2),hhr_r1(:,2),'type','Pearson');
%        smh = corr(mhr_r1(:,2),hhr_r1(:,2),'type','Spearman');
%        kmh = corr(mhr_r1(:,2),hhr_r1(:,2),'type','Kendall');
        fprintf('%s: Mean=%f RMSE=%f Pearson-correlation=%f P-
value=%f \n',...
            name,r1_diff,r1_rmse,abs(pmh), pval);
    end
end
```

4. Source code for CUSUM algorithm

```matlab
function [alarms, nc] = CUSUM( x, h, k, window, d )
%CUSUM Uses CUmulative SUM approach to detect change in a series or
signal.
%   [alarms, nc] = CUSUM( x, h, k, window, d ) takes the signal x,
the
%   detection threshold h, the change magnitude d, the window size
window
%   and the no. of change points to set an alarm, k as arguments.
%
%   The function returns a change point vector nc, containing ones
%   where there is a change point. Similarly, an alarms vector,
containing
%   the alarm positions depending on the value of k is also
returned.
%
%   h: set the detection threshold h > 0
%   k: Minimum No. ST Elevations to detect for an alarm
%   window: Sliding window size for CUSUM algorithm
%   d: set d to the most likely change magnitude

%///////////////////////////////////////////////////////////////////

alarms = zeros(length(x), 1);
nc = zeros(length(x), 1);        % Change points in x[n]
start = 1;
```

```matlab
new_min = zeros(length(x), 1);
%   end initialization


%//////////////////////////////////////////////////////////////////

%   main loop
while start < length(x)
G = zeros(length(x), 1);          % Decision Function
s = zeros(length(x), 1);          % Instantaneous Log-Liklihood Ratio
S = zeros(length(x), 1);          % Comulative Sum
start = start + 1;
for n = start : 1 : length(x)
    sigma = std(x(start:n));      % Current Standard deviation
    mu = mean(x(start:n));        % Current Mean for hypothesis H0

  % Calculate the instantaneous log-liklihood ratio, s[n]
    s(n) = (d/(sigma*sigma))*(x(n)-mu-(d/2));

  % Calculate the Decision Function, G[n] and Cumulative Sum, S[n]
    if n == 1
        S(n) = s(n);
        G(n) = max(s(n), 0);
    else
        S(n) = S(n-1) + s(n);
        G(n) = max((G(n-1) + s(n)), 0);

        min_array = find(S(1:n-1) == min(S(1:n-1)));
        if not(isempty(min_array))
            new_min(min_array(length(min_array))) =
min_array(length(min_array));
        end
    end

% Find the change point, when hypothesis switches from H0 to H1,
% characterized by G[n] being greater than the threshold value h
    if G(n) > h
        nc(find(new_min, 1, 'last')) = 1;

        if n > window
            if sum(nc(n-window:n)) >= k
                alarms(find(nc(n-window:n), 1, 'first') + n -
window - 1) = 1;
            end
        end

        % Reset the algorithm/ main loop
        start = n;
        break
    end
end
end
%   end main loop

end
```

5. Source code of CUSUM training parameters plot

```matlab
function [] = cusum_real_hr(data_dir)
d = dir(data_dir);
isub = [d.isdir] & ~strcmp({d.name},'.') & ~strcmp({d.name},'..');
%# returns logical vector
subfolders = d(isub);
mb_hr_name = 'hr.xlsx';
real_hr_name = 'real_hr.xlsx';
acc_name = 'acc.xlsx';
for k = 1 : length(subfolders)
    sub_path = fullfile(subfolders(k).folder, subfolders(k).name);
    mb_xlsx = fullfile(sub_path,mb_hr_name);
    mb_hr = xlsread(mb_xlsx);
    mbhr = mb_hr(:,2);

    real_xlsx = fullfile(sub_path,real_hr_name);
    real_hr = xlsread(real_xlsx);

    acc_xlsx = fullfile(sub_path,acc_name);
    acc_sd = xlsread(acc_xlsx);

    % Resample real heart rate to have same size with MB heart rate
    rlhr_s = interp1(real_hr(:,1), real_hr, mb_hr(:,1),
'linear','extrap');
    rlhr = rlhr_s(:,2);

    hr_delay = finddelay(mbhr, rlhr);
    fprintf('Heart rate delay of MB and Real is:%d \n',hr_delay);

    [alarms_mbhr, nc_mbhr] = CUSUM(mbhr,10, 2, 5, 25 );
    [alarms_rlhr, nc_rlhr] = CUSUM(rlhr,10, 2, 5, 25 );

    figure_name1 = [subfolders(k).name ' two heart rate'];
    % Plot mbhr and real hr in one figure
    fig1=figure('name',figure_name1);

    plot(mb_hr(:,1),mb_hr(:,2),'b');
    hold on;

    plot(rlhr_s(:,1),rlhr_s(:,2),'r');
    title('Two Heart Rate');
    legend('MB\_HR','PO\_HR');
    saveas(fig1,[sub_path '/' 'two_hr.jpg']);

    figure_name2 = [subfolders(k).name 'CUSUM two heart rate'];
    % Plot mbhr and real hr cusum in one figure
    fig1=figure('name',figure_name2);
    subplot(2,1,1)
    plot(mb_hr(:,1),nc_mbhr);
    title('MB Heart Rate');
```

```
    subplot(2,1,2)
    plot(rlhr_s(:,1),nc_rlhr);
    title('Real Heart Rate');
    saveas(fig1,[sub_path '/' 'cusum_2hr.jpg']);
end
end
```

6. Source code of CUSUM test plot

```
% Title:        Smart Watch Sensor data parse
% Created by:   Leichen Dai

function [] = cusum_test_plot(data_dir)
%case1 = '200';
%case2 = '300';
%case3 = '400';
%case4 = '500';
%case5 = '600';
%case6 = '700';
%case7 = '800';
case1 = 'CH';
case2 = 'ET';
case3 = 'MT';
case4 = 'VA';
case5 = 'WD';
d = dir(data_dir);
isub = [d.isdir] & ~strcmp({d.name},'.') & ~strcmp({d.name},'..');
%# returns logical vector
subfolders = d(isub);

function  cat_case(test_case)
    fprintf('Calculate case=%s \n', test_case);
    hr_name = ['hr.xlsx'];
    hr_t = [];
    acc_name = ['acc.xlsx'];
    acc_t = [];
    count = 0;

    for k = 1 : length(subfolders)
        if startsWith(subfolders(k).name, test_case) %find match
given
            fprintf('Find match case=%s \n', subfolders(k).name);
            %# absolute-path filename
            sub_path = fullfile(subfolders(k).folder,
subfolders(k).name);

            hr_xlsx = fullfile(sub_path,hr_name);
            hr_s = xlsread(hr_xlsx);
            hr_t = vertcat(hr_t, hr_s);
```

60

```matlab
            acc_xlsx = fullfile(sub_path,acc_name);
            acc_s = xlsread(acc_xlsx);
            acc_sd = interp1(acc_s(:,1), acc_s, hr_s(:,1),
'linear','extrap');
            acc_t = vertcat(acc_t, acc_sd);
            count = count + 1;
        end
    end

    [alarms_hr, nc_hr] = CUSUM(hr_t(:,2),10,5, 10, 20 );
    figure_name = [test_case '_CUSUM'];
    disp(figure_name);
    % Plot mbhr and real hr cusum in one figure
    fig1=figure('name',figure_name);
    subplot(3,1,1)
    plot(0:length(acc_t)-1, acc_t(:,2),0:length(acc_t)-1,
acc_t(:,3),0:length(acc_t)-1, acc_t(:,4));
    title([test_case ' Accelerometer']);

    subplot(3,1,2)
    plot(0:length(hr_t)-1, hr_t(:,2));
    title([test_case ' Heart Rate']);

    subplot(3,1,3)
    plot(0:length(hr_t)-1,nc_hr);
    title('Heart Rate Anomaly Detection');
    saveas(fig1,[data_dir '/' test_case '_cusum.jpg']);

end

cat_case(case1);
cat_case(case2);
cat_case(case3);
cat_case(case4);
cat_case(case5);
%cat_case(case6);
%cat_case(case7);

end
```

BIBLIOGRAPHY

Achten, J., & Jeukendrup, A. E. (2003). Heart rate monitoring - Applications and limitations. Sports Medicine, 33(7), 517-538. doi:10.2165/00007256-200333070-00004

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement, 28(3), R1-R39. doi:10.1088/0967-3334/28/3/r01

Asare, L., Kviesis-Kipge, E., Ozols, M., Spigulis, J., & Erts, R. (2012). MULTI-SPECTRAL OPTOELECTRONIC DEVICE FOR SKIN MICROCIRCULATION ANALYSIS. Lithuanian Journal of Physics, 52(1), 59-62.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1(8476), 307-310.

B. Reeder, A. David, Health at hand: a systematic review of smart watch uses for health and wellness, J. Biomed. Inform. 63 (2016) 269–276.

Chen, K. Y., & Bassett, D. R. (2005). The technology of accelerometry-based activity monitors: Current and future. Medicine and Science in Sports and Exercise, 37(11), S490-S500. doi:10.1249/01.mss.0000185571.49104.82

Chuah, S. H. W., Rauschnabel, P. A., Krey, N., Nguyen, B., Ramayah, T., & Lade, S. (2016). Wearable technologies: The role of usefulness and visibility in smartwatch adoption. Computers in Human Behavior, 65, 276-284.

Doherty, S. I., & Oh, P. (2012). A Multi-Sensor Monitoring System of Human

    Physiology and Daily Activities. Telemedicine and E-Health, 18(3), 185-192.

    doi:10.1089/tmj.2011.0138

Dolezal, B. A., Boland, D. M., Carney, J., Abrazado, M., Smith, D. L., & Cooper, C.

    B. (2014). Validation of heart rate derived from a physiological status monitor-

    embedded compression shirt against criterion ECG. Journal of Occupational

    Environmental Hygiene, 11(12), 833-839.

    doi:10.1080/15459624.2014.92511434

El-Amrawy F, Nounou MI. Are currently available wearable devices for activity

    tracking and heart rate monitoring accurate, precise, and medically beneficial?

    Healthc Inform Res 21: 315–320, 2015. doi:10.4258/hir.2015.21.4.315.

Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. Current

    cardiology reviews, 8(1), 14-25.

Fukushima, H., Kawanaka, H., Bhuiyan, M. S., & Oguri, K. (2012). Estimating heart

    rate using wrist-type Photoplethysmography and acceleration sensor while

    running. Conference proceedings : ... Annual International Conference of the

    IEEE Engineering in Medicine and Biology Society. IEEE Engineering in

    Medicine and Biology Society. Annual Conference, 2012, 2901-2904.

    doi:10.1109/embc.2012.6346570

Gao, L., Bourke, A. K., & Nelson, J. (2014). Evaluation of accelerometer based multi-

    sensor versus single-sensor activity recognition systems. Medical Engineering

    & Physics, 36(6), 779-785. doi:10.1016/j.medengphy.2014.02.012

IDC. (2017) *IDC Forecasts Shipments of Wearable Devices to Nearly Double by 2021 as Smart Watches and New Product Categories Gain Traction*. https://www.idc.com/getdoc.jsp?containerId=prUS43408517

Keytel, L. R., Goedecke, J. H., Noakes, T. D., Hiiloskorpi, H., Laukkanen, R., van der Merwe, L., & Lambert, E. V. (2005). Prediction of energy expenditure from heart rate monitoring during submaximal exercise. Journal of Sports Science, 23(3), 289-297. doi:10.1080/02640410470001730089

Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res. 2016 Sep 20;18(9):e253. doi: 10.2196/jmir.6025.

Kroll, R.R.; McKenzie, E.D.; Boyd, J.G.; Sheth, P.; Howes, D.; Wood, M.; Maslove, D.M. Use of wearable devices for post-discharge monitoring of ICU patients: A feasibility study. J. Intens. Care 2017, 5, 64, doi:10.1186/s40560-017-0261-9.

Lee, J., Matsumura, K., Yamakoshi, K.-i., Rolfe, P., Tanaka, S., & Yamakoshi, T. (2013). Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2013, 1724-1727. doi:10.1109/embc.2013.6609852

Lewis K. Validation of Wearable Biofeedback Technology for Heart Rate Tracking
Via Reflective Photoplethymorgraphy. California State Polytechnic University;
Pomona: 2017.

Lyons, E. J., Lewis, Z. H., Mayrsohn, B. G., & Rowland, J. L. (2014). Behavior
change techniques implemented in electronic lifestyle activity monitors: a
systematic content analysis. Journal of Medical Internet Research, 16(8), e192.
doi:10.2196/jmir.3469

M. Haescher, D. J. C. Matthies e B. Urban. "Anomaly Detection with Smartwatches
As an Opportunity for Implicit Interaction". Em: Proceedings of the 17th
International Conference on Human-Computer Interaction with Mobile
Devices and Services Adjunct. 2015, pp. 955–958.

Maeda, Y., Sekine, M., & Tamura, T. (2011). The Advantages of Wearable Green
Reflected Photoplethysmography. Journal of Medical Systems, 35(5), 829-834.
doi:10.1007/s10916-010-9506-z36

Mercer, K., Giangregorio, L., Schneider, E., Chilana, P., Li, M., & Grindrod, K.
(2016). Acceptance of Commercially Available Wearable Activity Trackers
Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods
Evaluation. JMIR mHealth and uHealth, 4(1), e7. doi:10.2196/mhealth.4225

O. Salem, Y. Liu, and A. Mehaoua, "Anomaly detection in medical wireless sensor
networks," Journal of Computing Science and Engineering, vol. 7, no. 4, pp.
272–284, 2013.

O. Salem, Y. Liu, A. Mehaoua, R. Boutaba, Online anomaly detection in wireless

    body area networks for reliable healthcare monitoring, IEEE J. Biomed. Health

    Inform. 18 (5) (2014) 1541–1551.

Parak, J., & Korhonen, I. (2014). Evaluation of wearable consumer heart rate monitors

    based on photopletysmography. Conf Proc IEEE Engineering in Medicine

    Biology Society, 2014, 3670-3673. doi:10.1109/EMBC.2014.6944419

Rafolt, D., & Gallasch, E. (2004). Influence of contact forces on wrist

    photoplethysmography--prestudy for a wearable patient monitor. Biomed Tech

    (Berl), 49(1-2), 22-26. doi:10.1515/BMT.2004.005

Pelizzo,G., Guddo,A. Aurora P, Annalisa D. S, (2018). Accuracy of a Wrist-Worn

    Heart Rate Sensing Device during Elective Pediatric Surgical Procedures,

    Children, 5,38;doi:10.3390

Stahl, S. E., An, H.S., Dinkel, D. M., Noble , J. M., & Lee , J.-M. (2016). How

    accurate are the wrist-based heart rate monitors during walking and running

    activities? are they accurate enough? (Vol. 2): BMJ Open Sport &

    Exercise Medicine.

Teng, X. F., & Zhang, Y. T. (2004). The effect of contacting force on

    photoplethysmographic signals. Physiological Measurements, 25(5), 1323-

    1335.

Terbizan, D., Dolezal, B., & Albano, C. (2002). Validity of Seven Commercially

    Available Heart Rate Monitors. Measurement in Physical Education and

    Exercise Science, 6(4), 243-247.

Vanderlei, L. C. M., Silva, R. A., Pastre, C. M., Azevedo, F. M., & Godoy, M. F. (2008). Comparison of the Polar S810i monitor and the ECG for the analysis of heart rate variability in the time and frequency domains. Brazilian Journal of Medical and Biological Research, 41(10), 854-859. doi:10.1590/s0100-879x200800500003937

V. P. Cornet and R. J. Holden, "Systematic review of smartphone-based passive sensing for health and wellbeing," Journal of Biomedical Informatics, vol. 77, pp. 120–132, 2018.

Wallen, M. P., Gomersall, S. R., Keating, S. E., Wisloff, U., & Coombes, J. S. (2016). Accuracy of Heart Rate Watches: Implications for Weight Management. Plos One, 11(5), 11. doi:10.1371/journal.pone.0154420

Z. Li, W. Xu, A. Huang, and M. Sarrafzadeh, "Dimensionality reduction for anomaly detection in electrocardiography: A manifold approach," in Proc. IEEE Wearable Implantable Body Sens. Netw., London, U.K., Jun. 2012, pp. 161–165

Zhu, C.; Sheng, W.; Liu, M. Wearable sensor-based behavioral anomaly detection in smart assisted living systems. IEEE Trans. Autom. Sci. Eng. 2015, 12, 1225–1234.