

Removal of Motion Artifacts from Photoplethysmogram Data for Feature Based Diagnosis of Atrial Fibrillation

BS Electrical Engineering

Session: 2020 - 21

By:

Mohammad Afaq 21100214

Sheraz Hassan 21100285

Salaar Ali Assad 21100087

Supervised by:

**Dr. Muhammad Tahir,
Assistant Professor, EE**



**Department of Electrical Engineering
Lahore University of Management Sciences**

Abstract

As a substitute of the traditional method of electrocardiogram (ECG) to monitor the heart activity of a patient, photoplethysmogram (PPG) has become an important alternative in the medical industry today. PPG holds immense advantages over the conventional ECG in terms of cost and flexibility. The data of a patient can be easily recorded through a non-invasive PPG sensor device without requiring the need of a medical supervision unlike ECG. Moreover, the recorded PPG data is useful in extracting vital cardiovascular parameters such as Heart Rate (HR), R-R interval, as well as respiration rate (RR) for identifying Atrial Fibrillation (AFib), a common cardiovascular disease showing symptoms of irregular heart behavior. However, the biggest challenge in extracting these parameters from the PPG data is interference due to artifacts. The artifacts this paper will explore can be categorized as either motion artifacts, which are caused due to physical movements or the changes in blood tissue volume, known as perfusion variation. These artifacts introduce noise in the data drastically affecting its quality, accuracy and reliability. This project will identify and address the above-mentioned challenge to remove artifact-induced noise from raw PPG data, ensuring that the original nature of the PPG waveform is not lost. The noise-free and cleansed PPG data can then be used to extract vital cardiovascular parameters with accuracy. Moreover, this project also aims to utilize the perfusion variation, extracted from the raw PPG signal for the approximation of vital health parameters namely; breathing rate, stress, SpO₂ (oxygen saturation level in the blood), mood etc. Our approach will start of by first comparing the raw PPG data from a low-end device with that of the pre-processed data from a high-end device with pre-implemented low frequency noise artifact removal algorithm. From this comparison, the perfusion variation can be approximated and subtracted from the raw PPG data. Next, this project will propose data processing algorithms to account for the emergence of motion artifacts from movements in the PPG data by procuring readings from both the PPG and its corresponding 3 axis accelerometer readings. Finally, it will process the cleansed PPG data and compute its cardiovascular parameters to accurately predict the possibility of whether the patient is exhibiting symptoms of AFib.

Acknowledgments

The authors wish to express sincere appreciation to Dr. Muhammad Tahir for his dedicated supervision during the course of this senior year project and in the preparation of this manuscript.

Originality Certificate

“We the undersigned certify that this submission is the original work of members of the group and meets the Faculty's Expectations of Originality”:

1. Mohammad Afaq [2021-10-0214]

Signed: _____ 21100214@lums.edu.pk

2. Sheraz Hassan [2021-10-0285]

Signed: _____ 21100285@lums.edu.pk

3. Salaar Ali Assad [2021-10-0087]

Signed: _____ 21100087@lums.edu.pk

Advisor

Dr. Muhammad Tahir

tahir@lums.edu.pk

Contents

Abstract	i
Acknowledgments	ii
Originality Certificate	iii
List of Figures	vi
List of Tables	viii
1 Introduction	1
1.1 Motivation	1
1.2 Problem Statement	2
1.3 General Block Diagram	2
1.4 Social Benefits and Relevance	3
1.5 Goals and Objectives	3
1.6 Outcomes	4
1.7 Timeline and Distribution of Work	4
2 Background	6
2.1 Literature Review	6
2.1.1 Photoplethysmogram	6
2.1.2 PPG Waveform	6
2.1.3 Perfusion Waveform	8
2.1.4 Atrial Fibrillation Detection	8
2.1.5 Motion Artifacts	9
2.1.6 Existing Methods	10
2.1.7 Data Processing Algorithms	11
3 Methodology and Tools	12
3.1 System Level Design	12
3.1.1 Perfusion Artifacts	12
3.1.2 Motion Artifacts	12
3.1.3 Noise	13
3.2 Tools and Instruments	14
3.2.1 Simulation Software Packages	14

3.2.2	Hardware Instruments	14
3.2.2.1	Empatica E4	14
3.2.2.2	Maxim	15
4	Project Implementation	17
4.1	Data Availability	17
4.2	Data Quality Comparison E4 vs MAXREFDES103	18
4.3	Separating Perfusion and PPG waveform from Raw PPG Signal	23
4.3.1	Approximate Perfusion Waveform using a Polynomial	24
4.3.2	Filtering PPG Waveform	26
4.4	Motion Artifacts Removal	28
4.4.1	Analysis using the Accelerometer Signal	28
4.4.2	Deep Learning	32
4.5	Parameter Estimation	32
4.5.1	Breath Rate	32
4.5.2	Pulse Wave Duration Interval	34
4.6	Results Analysis and Future work	34
4.6.1	Motion Artifact Removal	34
4.6.2	Health Parameter Approximation from Perfusion Variations	35

List of Figures

1.1	A general block diagram showing different stages of our project pathway	2
1.2	General timeline for our project	5
2.1	PPG waveform notation [1]	7
2.2	BVP for a healthy individual	8
2.3	BVP for an Afib patient	8
2.4	BVP for a healthy individual without motion artifacts	9
2.5	BVP for a healthy individual with motion artifacts	9
3.1	System Level Design	13
3.2	Front and Back view of Empatica E4	15
3.3	MAXREFDES103	15
3.4	MAX86140EVSYS [2]	16
4.1	The Raw PPG signal obtained using Green, Red and Infra-Red light can be observed here. Green 1 and Green 2 data is obtained using a single source green LED light source but detected from two photodiode channels	19
4.2	Plots of Raw PPG signal obtained under different configuration using MAXREFDES103 as mentioned in table 4.1.	21
4.3	Plot of PPG signal obtained under a 30 second window from the same individual with (a) MAXREFDES103 and (b) E4	23
4.4	Raw PPG Signal (Blue) along with the approximated Perfusion Waveform (Red)	25
4.5	PPG Waveform obtained along with its Frequency Plot	26
4.6	Raw PPG Signal and its Frequency Plot from MAXREFDES103	27
4.7	PPG Waveform and Perfusion Waveform obtained by filtering using a bandpass filter	28
4.8	BVP of a healthy individual at rest along with its magnitude spectrum using the E4	29
4.9	BVP of a healthy individual at with motion along with its magnitude spectrum using the E4	30
4.10	Filtered BVP using frequency of motion in the accelerometer data of E4	30
4.11	Comparison of inverted PPG waveform before and after eliminating frequencies of motion using MAXREFDES103	31

List of Figures

4.12 Respiratory Signal at 0.3 Hz with Local Maxima obtained	33
--	----

List of Tables

4.1	Different configuration for MAXREFDES103 to achieve a sampling average rate of 64 Hz	20
-----	--	----

Chapter 1

Introduction

1.1 Motivation

Disorders pertaining to the blood vessels and the heart are categorized as cardiovascular diseases (CVDs) such as; coronary heart disease (CHD), hypertension, angina, myocardial infarction, rheumatic heart disease and cardiac arrhythmias namely atrial fibrillation (Afib) etc., resulting from sedentary behaviour, poor dietary choices, smoking and pollution, brewing against a background of genetic susceptibility. Data from the World Health Organization states that CVDs are the primary cause of death world-wide among non-communicable diseases, and reasonable predictions state that the situation will become worse. In Pakistan, according to WHO statistics in 2014, approximately one fifth (19%) of the entire population was had CVDs, making it the leading non-communicable disease in the country [3]. In this scenario, a growing demand of medical assistance implies a large number of populations require assistance in hospitals where Electrocardiogram (ECG) signal recording remain the most prevalent clinical standards of care for cardiac health assessment and monitoring. However, ECG has not always been the first choice of patients due to its uncomfortable position of application, high cost and availability issues in the hospitals [4]. As a result, majority of the patients ignore the little to no symptoms of CVDs, resulting in a diagnosis only after a heart attack or a severe heart irregularity is experienced. Fortunately, with advancements in technology, the cardiovascular data of the patient can now be monitored real time without requiring any medical supervision. By having patients wear a wrist worn device which contains photoplethysmogram (PPG) sensors, data from the PPG signals can be collected and analyzed using data processing algorithms. Through this analysis, abnormal heart rhythms, which if left alone can evolve into major cardiac disorders, can be detected and addressed in the timely manner, potentially

saving the lives of many.

1.2 Problem Statement

The precise nature of the proposed project is to deal with the problem of remote and ambulatory cardiac health monitoring and assessment, more specifically for Atrial Fibrillation (Afib). The PPG signals will be recorded using a wearable device worn on wrist. The challenge in PPG based classification framework is that, first of all, the PPG data is corrupted under the influence of motion artifacts which originate as a result of movement of the device sensor over the skin surface, and secondly, the corrupted PPG data can super impose on the vital heart rhythmic data waveform that may denote the possibility of a cardiac event hence making it difficult to analyze and diagnose for any cardiac related disease [5]. Hence, there is a dire need to address these issues and develop data processing algorithms for a robust diagnosis and classification of any troubling cardiac event [5].

1.3 General Block Diagram

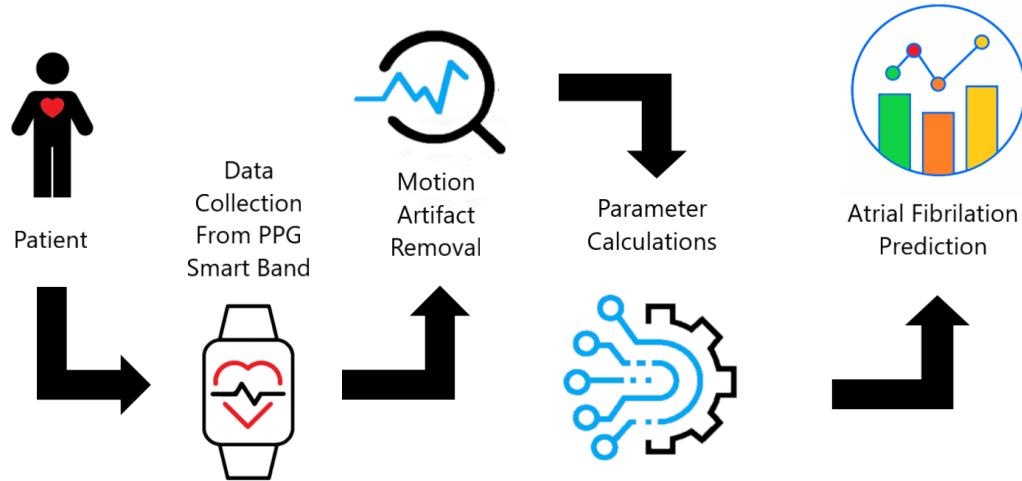


FIGURE 1.1: A general block diagram showing different stages of our project pathway

Figure 1.1 gives us a brief overview of the different stages involved in our project, with the first stage comprising of PPG data collection. The data collected will be collected from two PPG smart bands, mentioned in Chapter 3. We will also analyse and extract the perfusion waveform from the raw PPG signal obtained as part of the data collection and pre-processing of the signal. The second stage will be responsible for the cleansed PPG signal extraction from the corrupted signal by detecting and removing motion artifacts using a data processing algorithm.

This will be followed by parameter calculations to for useful feature extraction from the cleansed PPG waveform. The final stage involves the model accurately predicting whether a potential patient suffers from CVDs despite motion artifacts being introduced in the recorded PPG data.

1.4 Social Benefits and Relevance

As previously stated, statistics from WHO state that, among non-communicable diseases, CVDs are considered the primary cause of deaths. With the situation predicted to deteriorate as time passes, developing nations such as Pakistan are in dire need of a solution which can cater to this ever-increasing demand of cardiac medical assistance. This project, therefore, holds immense social benefits and relevance, as its application not only caters to local population but the national community as well.

Currently, medical resources, such as ECG signal recording instruments, are far too limited in Pakistan. They also require the patient to be present in the hospital hindering their daily life and, further straining doctors and medical staff at the hospital. But perhaps the most important aspect is that, despite the fact many cardiac disorders can be treated and possibly cured should there be timely intervention, the large medical expense such treatments incur act as a deterrent, particularly for those belonging to low income communities, making early detection of cardiac diseases difficult. Therefore, by providing an alternative that can detect early symptoms of AFib, is cheaper, requires minimal assistance from doctors and does not hinder the daily activities of the patients, this project can not only save lives, but also, minimize the expenses and time spent on hospital trips, medication, treatments, etc.

Furthermore, this project can also help doctors more accurately detect symptoms which would, under normal circumstances, go unnoticed as continuous monitoring of every patient at the hospital is impractical. The data of a patient during any hour of the day will be readily available for the medical staff hence allowing them to cater to a greater number of patients while at the same time offer reliable and accurate treatment.

1.5 Goals and Objectives

Our goals can be divided into three parts:

- Compare the differences between raw data from a low-end wearable PPG sensor device and processed data from a high-end PPG device.

- Understanding and analysis of the limits in data collections and its nature from low end PPG device such as PPG feature, artifacts, noise etc., with reference to data collected from high end device.
- Develop data processing algorithms that will take into account and adjust for motion artifacts present in the raw input data, such as motion artifacts.
- Create and use the cleansed PPG database to compute vital cardiac parameters using data science to identify potential cases of Afib.

1.6 Outcomes

Our project will present the following outcomes:

- Application of data processing algorithms on the raw PPG data to remove motion artifacts efficiently.
- Diagnosis of Afib on the cleansed PPG data by extracting and processing cardiovascular features
- If time permits, live simulation of the PPG data of an individual using a software algorithm that removes motion artifacts, extracts and computes PPG features and analyses them to predict red flags for any irregular heart activity under a time window T .

1.7 Timeline and Distribution of Work

Figure 1.2 shows the general timeline for our project. For Goal 1, an extensive review of existing literature began over the summer break and throughout September to establish a firm grasp on the understanding of the nature and the features of the PPG data. Literature linking AFib and PPG signals and the limitations in its data accuracy and collection was also reviewed during this time and further research on these topics will continue throughout the current semester.

For the second goal, device comparison, the processed data from the high end PPG device will be compared to the raw data from a low-end PPG sensing device. This comparison will allow for a greater understanding of the limitations of PPG data collection in low-end devices, creating a more holistic view for the practical and commercial implementation of wearable PPG smart bands for AFib detection. Goal 1, which began in September, will be concluded in December 2020 for the Senior Project 1 report.

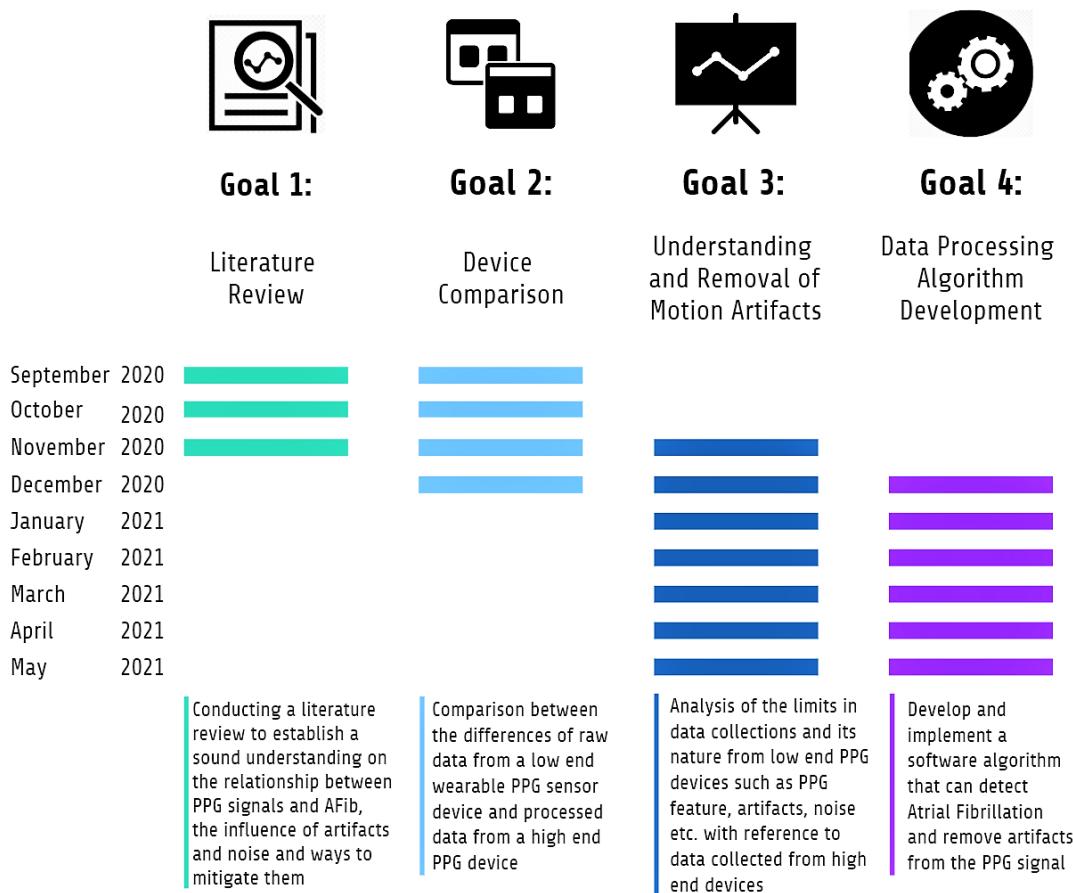


FIGURE 1.2: General timeline for our project

Goal 3 is the investigation of the nature of PPG features along with the various types of motion artifacts and noise that will be present in the raw data collected. The raw data will be analyzed to not only understand the challenges faced when collecting PPG signals, particularly from low-end devices, but also, determine potential solutions that can mitigate these problems. Goal 3 will begin in November and expand over both semesters till May 2021 where it will be concluded in the Senior Project 2 report.

The last goal will focus on the development and implementation of a software algorithm that can successfully detect and remove artifacts from the raw PPG signal, primarily motion artifacts, using data processing techniques. Moreover, the algorithm will also be able to correctly identify cases of individuals suffering from AFib. Research on exploring the various techniques previously used in existing literature has been carried out along with multiple attempts to simulate these results. The investigation and development of the software algorithm will start towards the end of the first semester and will conclude in May 2021.

Chapter 2

Background

2.1 Literature Review

2.1.1 Photoplethysmogram

Photoplethysmography (PPG) is a popular technique used extensively in the medical industry to obtain a bio optical signal by measuring the variations in blood flowing through a tissue using the properties of reflection of light through a photo diode channel. It is measured using a PPG sensor at the skin surface and unlike ECG, is a low cost non invasive alternative which operates close to green and red infrared frequency. The way how a PPG sensor works is through LED and a photo diode. The blood tissue is first illuminated using a light source (LED) and a photo diode near the light source measures the variations in the amount of light being reflected from the blood flowing through the tissue. These volumetric changes of blood in the blood vessels produce variation in voltage through which the PPG waveform is produced with respect to time. The variation of the blood volume in the with respect to the heart activity forms the variation in the amplitude of the voltage hence forming PPG waveform over time.

2.1.2 PPG Waveform

PPG waveform is the most popular clinical waveform as of today. A simple technology, PPG has dominate ECG in the past years due to its increased sensitivity and specificity in rapid detection and diagnosis of the heart activity [6]. It is considered to be relatively less complex than ECG. A lot of work has already been performed to measure heart rate and blood pressure PPG signals. PPG can give also give a good estimation of many other heart related diseases other than AF. Researchers are exploring the ways to use the power of modern machine learning

techniques as well as deep learning for the improved accuracy of Cardiovascular diseases (CVDs) detection and screening from the PPG signal[7–9].

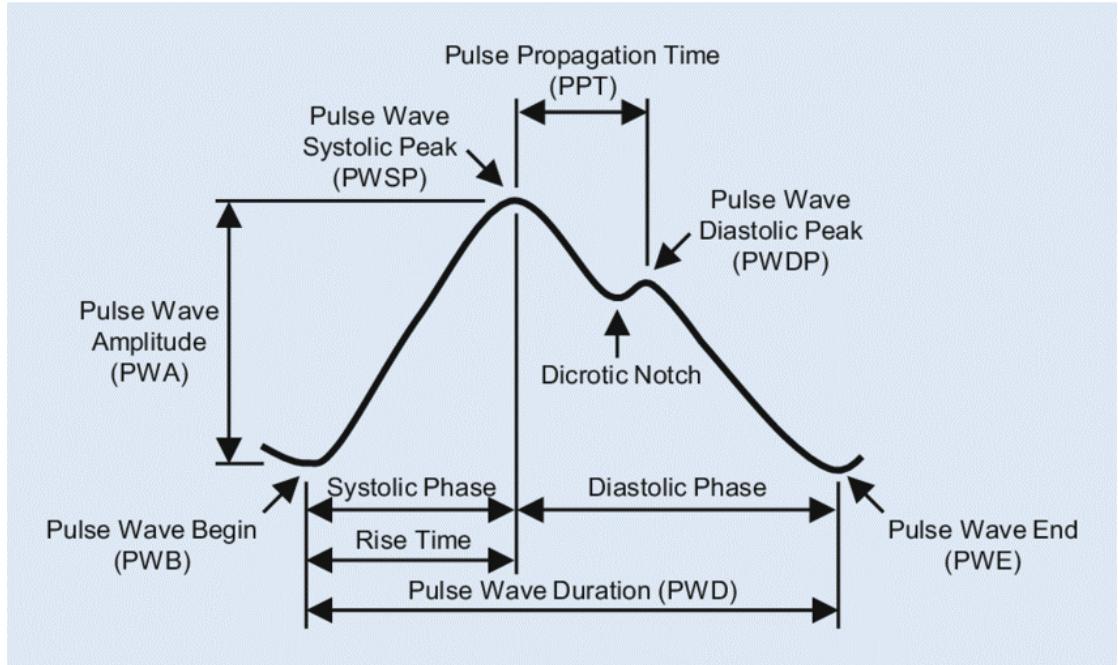


FIGURE 2.1: PPG waveform notation [1]

A cardiac cycle generated in the PPG waveform is shown in Figure 2.1. The signal is generated as the heart contracts and this contraction spreads through the vascular tree [10]. As the left ventricle contracts, the blood is pumped into the arterial tree, which is denoted as the positive gradient of the PPG waveform in figure 2.1. Further the closing of the aortic valves and the separation of the systolic and diastolic phase's results in a decrease in amplitude of the waveform [10]. These specific events which form part of the cardiac cycle together are a part of a PPG waveform for one heartbeat. PPG waveform is believed to hold a range of extractable features which are immensely useful in detecting Afib. Among them, the most relevant to our project is the R-R interval or Inter Beat Interval (IBI) which measures the difference in time between successive peaks (one heartbeat). Logically, with high metabolic activity, the R-R interval decreases as the heart rate increases and the peak magnitude remains somewhat constant. This means that the peaks of each successive PPG waveform gets closer together but this does not happen spontaneously. A certain pattern in the shifting of R-R interval is observed for a healthy individual. The case is opposite for a patient diagnosed with Afib.

2.1.3 Perfusion Waveform

Perfusion, or more specifically, blood perfusion, refers to the blood flow through the vast capillary network in the blood tissue to transport oxygen and nutrients to the living cells along with playing a vital role in homeostasis [11]. The variation of this blood flow is influenced by various factors such as respiration, stress, breathing rate, and anxiety, resulting in a baseline wander in the raw PPG signal due to variation in intensity of light being reflected from the tissue and the capillary network [12]. This baseline wander will be termed as perfusion variation in this report which can be analysed through its respective perfusion waveform. This perfusion waveform is a general plot which introduces a shift in the DC component of the PPG waveform over time which is spontaneous.

2.1.4 Atrial Fibrillation Detection

For this project, our focus lies on detecting Afib by analyzing the difference of successive R-R intervals as well as the variation in the magnitude of the PPG signal since Afib diagnosed patients show the symptoms of ill regular rhythm of the heart activity lasting for more than 30 seconds [10]. Figure 2.2 and 2.3 shows the Blood Volume Pulse (BVP) data acquired from a PPG sensor of a healthy patient with that of a patient diagnosed with AF under a 30 second time window. We will constantly use the term BVP in this report to denote PPG data since it is commonly known as BVP in scientific literature.

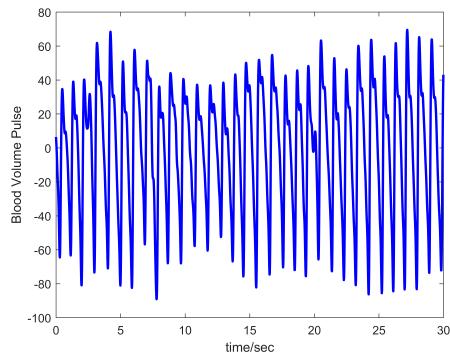


FIGURE 2.2: BVP for a healthy individual

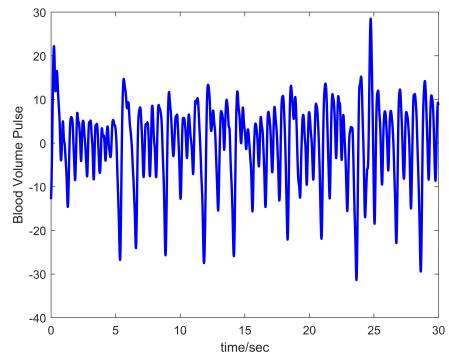


FIGURE 2.3: BVP for an Afib patient

Each peak in the BVP plot represents a heartbeat whereas the time difference in each successive peak represents the R-R interval. It can be clearly observed that for a patient diagnosed with AF in figure 2.3, the R-R interval are ill regular with random variations in the amplitudes of the BVP plot as well whereas for a healthy

patient in figure 2.2 , both the parameters of R-R interval and amplitude remain constant throughout the 30 second window.

2.1.5 Motion Artifacts

Motion artifacts today remains the biggest challenge in recording reliable BVP data. Before reviewing literature on how such motion artifacts can be removed, it is vital to understand how noise due to motion artifacts is introduced in the signal. Motion artifacts occur not only by the disturbance in the contact made between the skin and the sensor, but also by the movement of the body which can be as small as movement through breathing [13]. Hence, making it a difficult task to obtain an accurate BVP reading through sensors. Motion artifacts and noise introduced into the readings by sensors make the BVP signal unreliable and inaccurate. Hence the parameters obtained through these readings cannot be relied on to detect AF.

With the introduction of motion artifacts, the recorded BVP signal is corrupted due to added noise making it difficult to analyze and extract vital information. These motion artifacts are random, meaning they are introduced by random movement of the user hence will be inevitably introduced in the BVP signal recording of a patient whose data is being taken simultaneously throughout the day. As a result, the signal obtained will be unsatisfactory and must be processed using data processing algorithms to eliminate these motion artifacts before accurate and reliable data can be obtained for analysis. To better understand the situation, it is important to visualise the BVP data from a healthy individual both with and without motion artifacts. To ensure that the BVP data was free from motion artifacts, the recording was taken with the individual at rest position.

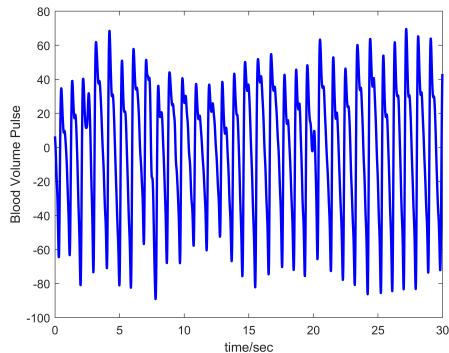


FIGURE 2.4: BVP for a healthy individual without motion artifacts

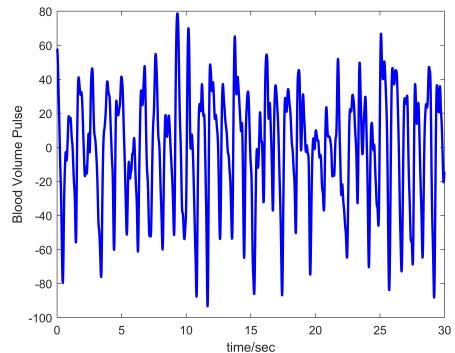


FIGURE 2.5: BVP for a healthy individual with motion artifacts

Figure 2.4 shows the recorded BVP of the individual without motion artifacts in a 30 second interval and figure 2.5 shows the BVP of the same individual but with motion artifacts in a 30 second interval. Figure 2.5 depicts how raw BVP signal for a patient will become noisy subject to motion artifacts hence making it difficult to analyze to extract information from it. Therefore, the major problem that we intend to solve in this project is to accurately remove the motion artifacts from the BVP signal such that no vital information is lost and the signal retains its shape which is extremely important for post analysis of the signal. This includes accurately detecting BVP signal cardiac parameters and using them to accurately predict the heart condition of the patient.

2.1.6 Existing Methods

As of today, there exists several existing methods which are capable of eliminating motion artifacts and cleansing BVP signal widely used in scientific research. Among these methods extensively studied in literature are independent component analysis (ICA), adaptive filtering and deep learning technique to extract physiological features from the BVP signal [13–15]. ICA algorithm is based on the probability statistical theory where the complex BVP signal is separated into its sub components corresponding to the BVP of the variation of blood volume in the vessels, motion artifacts and noise etc. These sub components corresponding to motion artifacts and noise can then be filtered out leaving behind noise free BVP signal [13, 16]. However, the ICA algorithm makes a strong assumption regarding the independence of these sub components, leading to inaccuracy of the results and loss of information which might otherwise prove to be useful due to the violation of this assumption of the ICA algorithm [13, 17, 18].

Another widely used method is adaptive filtering which can achieve satisfactory results by suppressing in band frequencies of motions that might have been introduced in the corrupted BVP signal as motion artifacts [13, 19, 20]. This means that a correlation between the BVP and its corresponding 3 axis accelerometer data is assumed [13]. As a result, this method proposes several disadvantages e.g. a bad quality reference signal through which error is calculated will degrade the quality of the output PPG signal. Moreover, this method utilizes extensive computational resources and power, therefore not optimum for implementing on a wearable PPG sensor device which doesn't hold such high processors as a computer does.

Another proposed approach relies on deep learning techniques to extract the physiological features from the corrupted BVP signal. The paper in [21] highlights that the difficulty in motion artifacts removal from corrupted BVP is attributed to the

non linearity of the cleansed BVP signal and noise. It is this non-linear relationship that limits the effectiveness of other methods, such as adaptive filtering, due to the linear nature of the additive noise hypothesis, and signal decomposition algorithms. Therefore, this technique proposes an algorithm based on Signal-Noise Interaction modeling (SniMA) for complex BVP signals prone to motion artifacts [21]. This algorithm will make use of Envelope Filtering (EF) as well as Time-Delay Neural Network (TDNN) in order to model signal-noise interaction as opposed to directly modeling PPG signals [21]. The EF algorithm is applied to normalize data from the PPG device and eliminate the training imbalances induced by respiration. Whereas, the TDNN aims to not only model the PPG signal noise interaction as a non-linear process, but also, introduce time-dependence information for artifact removal.

Hence, motion artifact removal from a corrupted PPG signal is the primary objective of this project. We will study the existing literature and propose a technique that able to remove the motion artifacts effectively without disrupting the nature of our original PPG signal such that vital cardio parameters can be estimated with accuracy.

2.1.7 Data Processing Algorithms

With the increased number of populations around the world, estimated to reach 8.5 billion by the end of 2030, meeting health care needs of its citizens has become a real challenge for every country. These significant numbers present significant amount of data available of patients which has rapidly increased the demand of machine intelligence in the health industry. Data science in medicine has proven to be immensely resourceful in identifying potential disease infections and can drastically improve the accuracy of diagnosis reducing the human error that presents itself at certain occasions. Hence, in this project, we intend to understand the nature of PPG at the fundamental level to research and implement algorithms to make the PPG signal more robust to motion artifacts. This robustness will allow us to extract features with greater accuracy and reliability and use them for the detection of any serious heart condition that may be detoriorous to one's health.

Chapter 3

Methodology and Tools

3.1 System Level Design

Our research areas revolves around the raw PPG data which contain artifacts, divided into three categories, namely; motion, perfusion and other (irreducible noise) as shown in 3.1.

3.1.1 Perfusion Artifacts

Removal of the perfusion variation is important as it not only results in inconsistencies and unreliability in the PPG waveform, but it also contains a lot of hidden information, such as, breathing rate, stress levels and O₂ saturation levels which can be extracted and analyzed. Our approach will have the use of two PPG devices, a high-end device (Empatica E4) and a low-end device (Maxim MAXREFDES103). The Empatica E4 removes perfusion using its own inbuilt software and will be used as reference for the low-end device. The low-end device, although less accurate, gives us the raw PPG data which contains perfusion, therefore, allowing us to separate the PPG waveform from the perfusion waveform.

3.1.2 Motion Artifacts

PPG data is prone to motion artifacts which can greatly affect the accuracy and reliability of the results. Therefore, we aim to remove these artifacts from the raw PPG signal. Removal of these motion artifacts will be the primary focus of SPROJ II. Once they are removed, the R-R interval of the processed PPG signal will be compared to a clean PPG signal as a metric to determine the performance of the motion artifact removal technique.

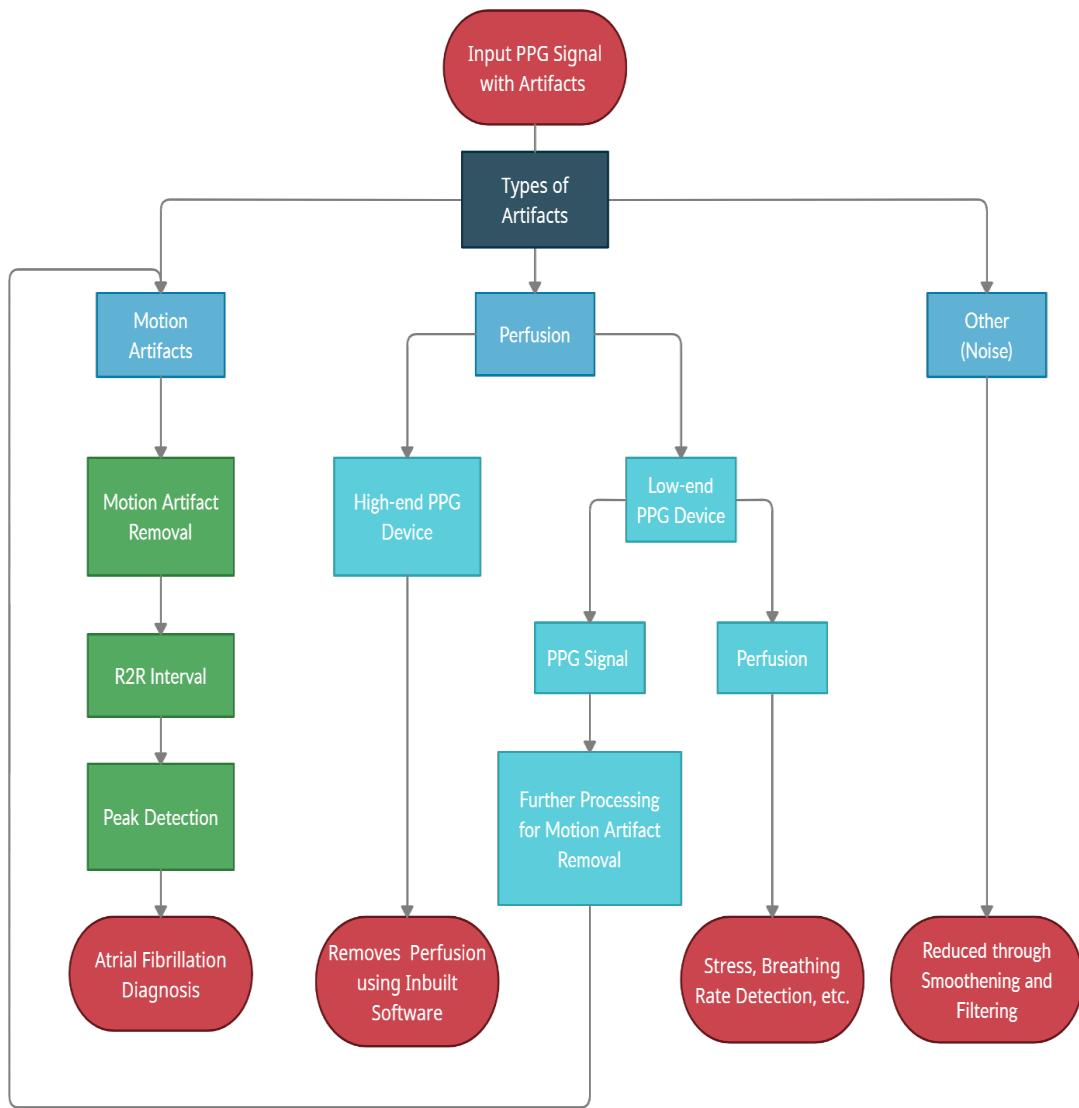


FIGURE 3.1: System Level Design

3.1.3 Noise

Any artifacts that remain in the PPG signal after removing motion and perfusion artifacts, is the irreducible random noise in the signal. It is the background noise that will always be introduced in our system. These artifacts cannot be removed entirely, however, by using smoothing and filtration techniques, they can be reduced.

3.2 Tools and Instruments

3.2.1 Simulation Software Packages

We used many different software and tools in our project as per our need and convenience. Language of implementation of our project is MATLAB and Python. All of these software and tools are listed below along with their uses.

- **MATLAB**

MATLAB is being used majorly for signal processing and analysis. We found signal processing to be more developed and convenient in MATLAB.

- **Jupyter Notebook**

We used Jupyter notebook environment for all of our python-based working. Among the many libraries used in python; NumPy, Pandas, Keras and SciPy are a few.

- **Machine learning / Deep learning models**

Although not implemented during SProj I, we worked with machine/deep learning method of motion artifact removal. We aim to pursue this in SProj II by using preexisting and our own data sets to train our model.

- **MAXIM UI/ Empatica UI**

UI of Empatica was very straight forward to interact with as Empatica doesn't offer a lot of freedom but MAXIM UI gives free hand to the user in terms of flexibility towards sampling rate and sampling average and requires a complete comprehension to be used properly.

3.2.2 Hardware Instruments

3.2.2.1 Empatica E4

Empatica E4 is a high-end wrist wearable device which is used clinically for real time data acquisition. E4 has the ability to plot the PPG signal in real time along with storing the data in the cloud to be used later. Its PPG sensor uses 4 LED's to get the PPG data (also known as Blood volume Pulse or BVP) and works at a fixed sampling rate of 64Hz. Two LED's operates at green and the other two at red wavelengths. Both lights pass through the skin and get absorbed in blood differently. The obtained signals are then processed by passing them through pre-processing algorithms for the removal of perfusion waveform and less prominent

motion artifacts. This limits the extent of experimentation that can be done as the data is not truly raw in the obtained nature and has been through motion artifact removal algorithm. E4 also contains 3-axis accelerometer corresponding to the PPG signal.



FIGURE 3.2: Front and Back view of Empatica E4

3.2.2.2 Maxim

We used two different Maxim devices. These devices are low end and less costly as compared to the Empatica E4.

- First device is MAXREFDES103. It is a Health Band and a wrist wearable device just like Empatica E4. But unlike E4 this band has only one red and one green LED along with an IR emitter and two photo diodes. MAXREFDES103 can also plot the PPG signal in real time but it also provides us with raw data that can be used for proper experimentation. Its data consists of red, green and IR readings. It is also capable of operating at a variable sampling rate.



FIGURE 3.3: MAXREFDES103

- The other Maxim device we used is MAX86140EVSYS which is an evaluation kit containing the MAX86140 and MAX86141 Sensor. The MAX86140 has a single photodiode channel whereas the MAX86141 has a double photodiode channel. Similar to the MAXREFDES103, it gives us the flexibility of optimizing its parameters such as sampling rate, sensitivity, and LED output current as per our needs.

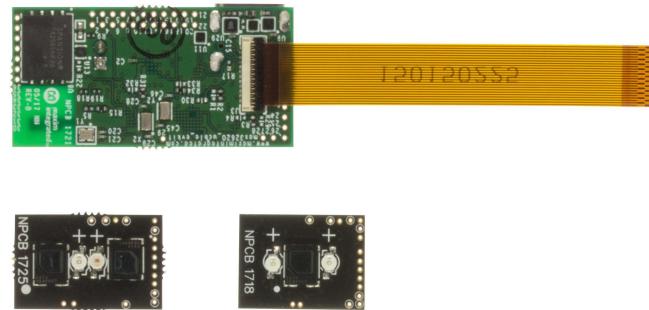


FIGURE 3.4: MAX86140EVSYS [2]

So far we did not require any graphics processing units for processing the data, all processing and analysis was done on 6th generation Intel core i7.

Chapter 4

Project Implementation

This chapter outlines the implementations involved in our project and their results so far. It will describe the techniques we used for our data collection using Maxim MAXREFDES103 and Empatica E4 in addition to the other preexisting data sets that we used. That will be followed by different methods of cleansing and processing of this acquired data along with the analysis of the final outcomes.

4.1 Data Availability

Details of the data sets we used are detailed as follows:

- We used pre-existing data sets of BVP of 46 individuals that was collected using Empatica E4 under a certain protocol to maintain homogeneity last year. These 46 data sets included both healthy people (20) and people suffering from Atrial fibrillation (26). The 26 unhealthy data sets were collected from patients who were interviewed last year by our fellow seniors under clinical standards. The 20 healthy individuals were from within the LUMS community. All of these data sets were taken under supervision to make sure minimum to no motion artifact was introduced in the signal.
- Our data collection was limited to the university students due to the risks involved with Covid-19. We used E4 to get data sets of healthy students but with some specific motion artifact introduced as follows where the individual was asked to sit at an upright position in a calm posture. The Data was then recorded for 5 minutes 30 seconds with the following controlled movement:
 - Rest position to gain a signal for reference
 - Moving the arm up and down in a controlled pace

- Moving the arm left and right in a controlled pace
- Rotating the wrist clockwise and then returning to the original position of the wrist. This movement was repeated throughout the recording interval
- Fist making and relaxing to understand in what way did muscle contraction affected our signal

We also used MAXREFDES103 health band wrist wearable device to get different data sets of students in the university. These sets were both, without motion artifacts and a specific introduced motion exactly like Empatica E4 as mentioned above.

- We also used a preexisting data set from IEEE signaling cup. This data set included the PPG signal along with the corresponding 3 axis accelerometer signal and the Heart rate (analyzed from the PPG signal). We used this as the trainer data in the machine/deep learning model we tried to implement (to be continued in SProj II).

We will be collecting more data in our SProj II and diversify our data set across gender and age in a controlled motion inducing environment. We will test our data processing algorithms in this updated data set to account for their accuracy and reliability in SProj II.

4.2 Data Quality Comparison

E4 vs MAXREFDES103

Our first approach towards this project was to better understand what hidden information a pure raw PPG signal from the MAXREFDES103 gives through its photodiode channels in comparison to E4. Unlike E4, the MAXREFDES103 gives us the flexibility of configurating its sample rate along with the sampling average. The terminology of sampling average is extremely useful in removing redundant noise. Computationally, it takes the first n^{th} samples in a frame where n is the sampling average. It then computes their average which is then stored as sample 1. Next, the frame then takes the next n^{th} samples and denotes sample 2 as their computed average. Hence, the total averaged samples computed is given by (4.1).

$$\text{Total Averaged Samples} = \frac{\text{Total Samples}}{\text{Sampling Average}} \quad (4.1)$$

Where the Averaged Sampling Rate is provided by (4.2)

$$\text{Averaged Sampling Rate} = \frac{\text{Sampling Rate}}{\text{Sampling Average}} \quad (4.2)$$

Figure 4.1 below shows a raw PPG signal obtained from a healthy individual at rest position over the 60 second interval window under the default parameters with sampling rate at 100 Hz and sampling average as 4. Hence, giving us the averaged sampling rate as 25 Hz.

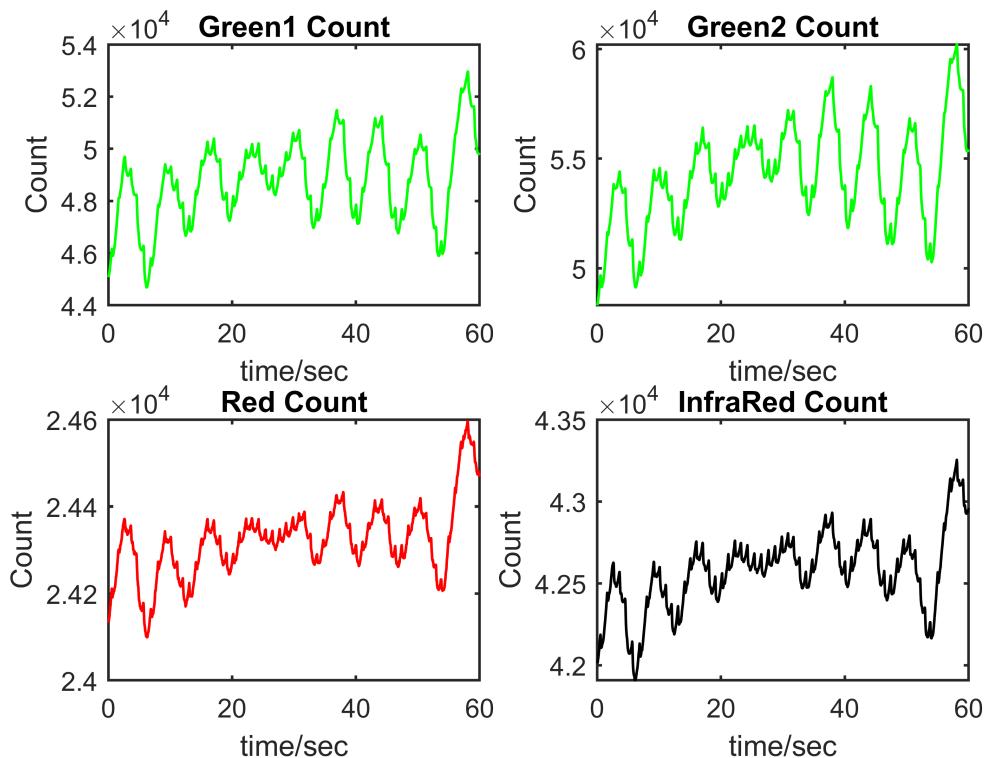


FIGURE 4.1: The Raw PPG signal obtained using Green, Red and Infra-Red light can be observed here. Green 1 and Green 2 data is obtained using a single source green LED light source but detected from two photodiode channels

The PPG signal in figure has been obtained using three different light sources from the MAXREFDES103 namely green, red and Infrared (IR). However, green₁ and green₂ plots are obtained using the same LED source but two photo diode channels. At rest, the individual seems to present no motion artifacts that might interfere with the PPG signal. However, as it can be observed, there is perfusion variation in the PPG signal on top of which we can observe the PPG waveform that is generated from the rhythm of the heart. The shape of the perfusion waveform for the green₁ and green₂ plots follow the same general trend with much higher

variation depicted as compared to the Red and IR plots. This is because the green wavelength at 530 NM gives us the most suitable data of PPG due to its relative freedom from noise [22]. Comparing the DC components for the green plots, it can be observed that for green₂ count, the DC component is much higher as compared to the green₁ count. This is due to the position of the photo diode channels from the LED and the major vein through which this variation in blood flow is being observed. For the remainder of this report, we will be using the green count of the MAXREFDES103 for analysis purpose due to its potential in providing more reliable information with much less susceptibility to artifacts as per on going research in PPG. However, we will utilize the red wavelength PPG waveform towards our future work in SPROJ II where our main focus will be on the removal of motion artifacts.

Since the E4 has a fixed sampling rate of 64 Hz, in order to justify our comparison, we decided to configure our averaged sample rate for MAXREFDES103 to be 64 Hz as well. However, in order to achieve this averaged sampling rate, the MAXREFDES103 can be configured in different ways as shown in table 4.1.

Sampling Rate / Hz	Sampling Average	Averaged Sampling Rate / Hz
64	1	64
128	2	64
256	4	64
512	8	64
1024	16	64
2048	32	64

TABLE 4.1: Different configuration for MAXREFDES103 to achieve a sampling average rate of 64 Hz

To better understand which configuration is optimal in obtaining a smooth signal from the PPG sensor, different readings under each configuration were taken from a healthy individual at rest from a thirty second window and their plots were then compared visually using the waveform obtained from green LED count.

From figure 4.2, PPG signal obtained with a sample rate of 64 Hz with a sampling average of 1 is prone to noise which follows our theory that PPG is prone to artifacts hence obtaining signals under this configuration will produce variation

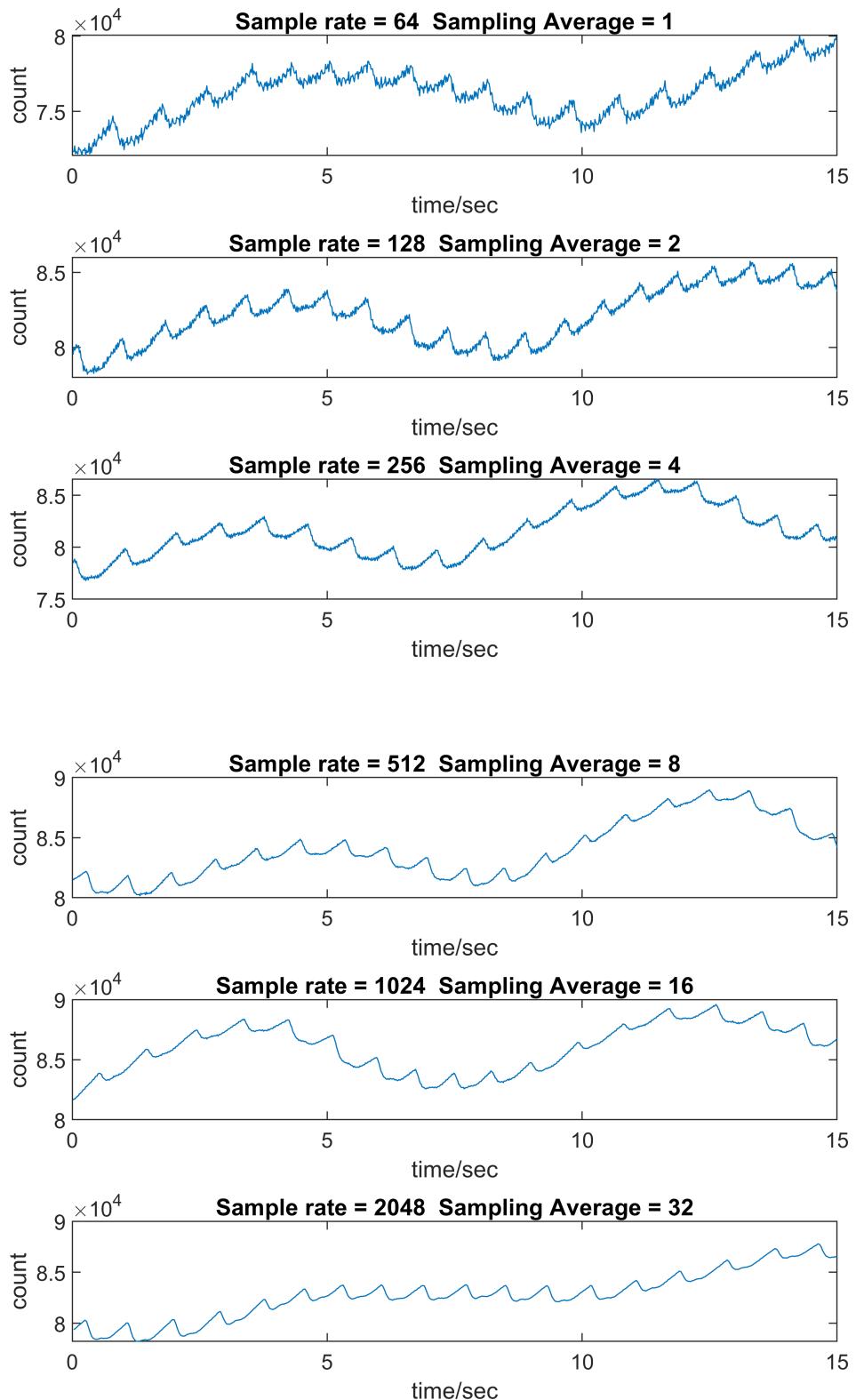


FIGURE 4.2: Plots of Raw PPG signal obtained under different configuration using MAXREFDES103 as mentioned in table 4.1.

in our PPG waveform. As we increase the sample rate and sample average, the signal becomes smoother with this redundant noise being cancelled as the average of samples is computed. Figure 4.2 also shows that as the sample rate exceeds 512 samples /sec with a sampling average of 8, there is not much difference in the quality of smoothness of the signals in terms of removal of unwanted noise. Although, the process becomes computationally expensive. Hence, a trade-off is established. Where low sampling rate with sampling rate to average ratio being held constant introduces noise whereas a high sampling rate makes the process computationally expensive hence not the optimal procedure for a wrist worn device which has limited processing capabilities. For our project, we decided on keeping the sample rate at 512 Hz with a sampling average of 8 to be our optimal configuration as figure shows. Not only does it get rid of background noise which corrupt our signal but is also computational efficient as well. All the plots obtained from Maxim devices in the rest of this report will follow this configuration.

With the correct configuration being determined, we now need to create a general equation for the PPG signal obtained using both MAXREFDES103 and E4 to break down our signal into different components and determine the commonality between the two. One way that we decided to approach was to take data from the MAXREFDES103 and the E4 simultaneously from a healthy individual at rest position over a period of time. The individual wore MAXREFDES103 on his left wrist whereas E4 was worn on the right arm wrist. Special attention was given to the individual to ensure that the individual remains at comfortable rest position such as to avoid any motion artifacts that might disrupt our signal. The readings were taken over a time interval of 3 min. Hence both devices recorded approximately $(3 \times 60) \times 64 = 11520$ samples out of which the initial 20 samples were discarded as the device is likely be in its transient stage.

Figure 4.3 shows a 30 second window PPG signal obtained for both E4 and MAXREFDES103. It can be fairly observed that the PPG signal from the E4 shows little to no effect of perfusion variation and the only AC component that causes variation in the plots is due to the rhythmic beats of the heart of the individual as well as added noise aggravated under the influence of motion. PPG waveform from the same individual under the same time instant from MAXREFDES103 can also be observed. Perfusion variation dominates the waveform here on top of which lies our PPG waveform due to the rhythm of the heartbeat of the individual. The waveform is also subject to noise and little to no motion artifacts. Hence the general equation for both MAXREFDES103 and E4 signal can be written as follows:

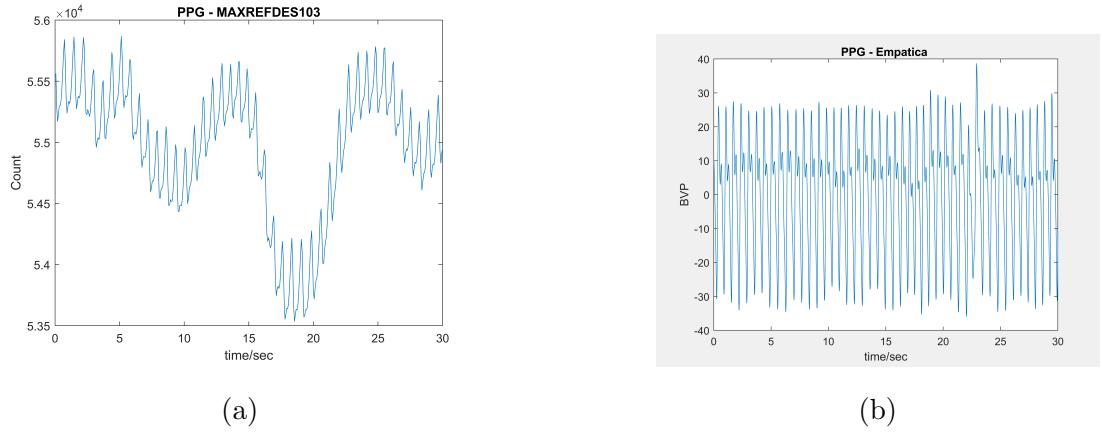


FIGURE 4.3: Plot of PPG signal obtained under a 30 second window from the same individual with (a) MAXREFDES103 and (b) E4 .

$$MAXREFDES103 \text{ Signal} = \text{Perfusion waveform} + \text{PPG waveform} + \text{noise} \quad (4.3)$$

$$E4 \text{ Signal} = \text{PPG waveform} + \text{noise} \quad (4.4)$$

Where noise contains irreducible noise as well as motion artifacts which corrupt our data. Hence from the equation we can see, that the MAXREFDES103 PPG signal can be broken down into its perfusion waveform and the PPG signal, both which contains vital information related to cardio parameters.

4.3 Separating Perfusion and PPG waveform from Raw PPG Signal

To separate the both the perfusion waveform as well as the PPG waveform from the raw PPG signal obtained from MAXREFDES103 as generalized in equation 4.3, we can either first approximate the perfusion waveform and then subtract it from our raw PPG signal to obtain our PPG waveform or filter out the PPG waveform and then use it to compute our perfusion waveform. In this section, we will analyze both methods in detail.

4.3.1 Approximate Perfusion Waveform using a Polynomial

As depicted in figure 4.3, we can mathematically approximate the shape of the curve of the perfusion waveform such that it joins the midpoint between each rhythmic rise and fall of the PPG waveform. However, as we have observed through several trials, the perfusion waveform is subject to variation depending upon several factors such as the breathing rate, stress, mood, hormones, activity etc. Therefore, approximating the complete PPG waveform would be susceptible to errors due to its randomness over the time period on which the data has been taken. In order to overcome this, we can treat the rhythmic PPG waveform on top of the perfusion waveform as irregularities that need to be removed to obtain a smooth perfusion waveform. In order to accomplish this, we use the Savitzky-Golay smoothing filter. The Savitzky-Golay is a finite impulse response (FIR) filter which calculates the polynomial fit of order n under a positive odd numbered frame length (samples) of the input signal and returns the polynomial curve that best approximates the input signal. Savitzky-Golay filter can be implemented in MATLAB using the command `sgolayfilt()`.

Savitzky-Golay filter was applied to a raw PPG signal obtained from a healthy individual at rest with a 2nd order polynomial under the frame length of $4 \times$ *averaged sampling rate* (64 Hz) + 1 = 257 samples. The approximate perfusion waveform under a 45 second window was achieved in figure 4.4.

This perfusion waveform is then subtracted from our raw PPG signal to obtain our PPG waveform. To remove any unwanted noise, we apply a pass band filter in the range of 0.5 – 4Hz as the pulse wave frequency values of the PPG waveform lies within this range [23].

Figure 4.5 shows the PPG waveform obtained along with its frequency spectrum. As can be observed, each peak in the PPG waveform depicts one heartbeat with the time difference between two successive peaks as the R-R interval. However, one interesting thing to note here is that the PPG waveform is inverted as compared to the universal PPG waveform in figure 2.1. This is because the PPG signal measured by the photo diode channels is measured as changes in the volume of blood flowing through the capillary network in the tissue. With the contraction of heart during the systole phase, more blood is pushed through the arteries. With more blood flowing, a much greater proportion of light is absorbed and less is reflected to the photo diodes as a result, the output light current decreases. This is the systole phase which is denoted as the positive gradient in figure 2.1.

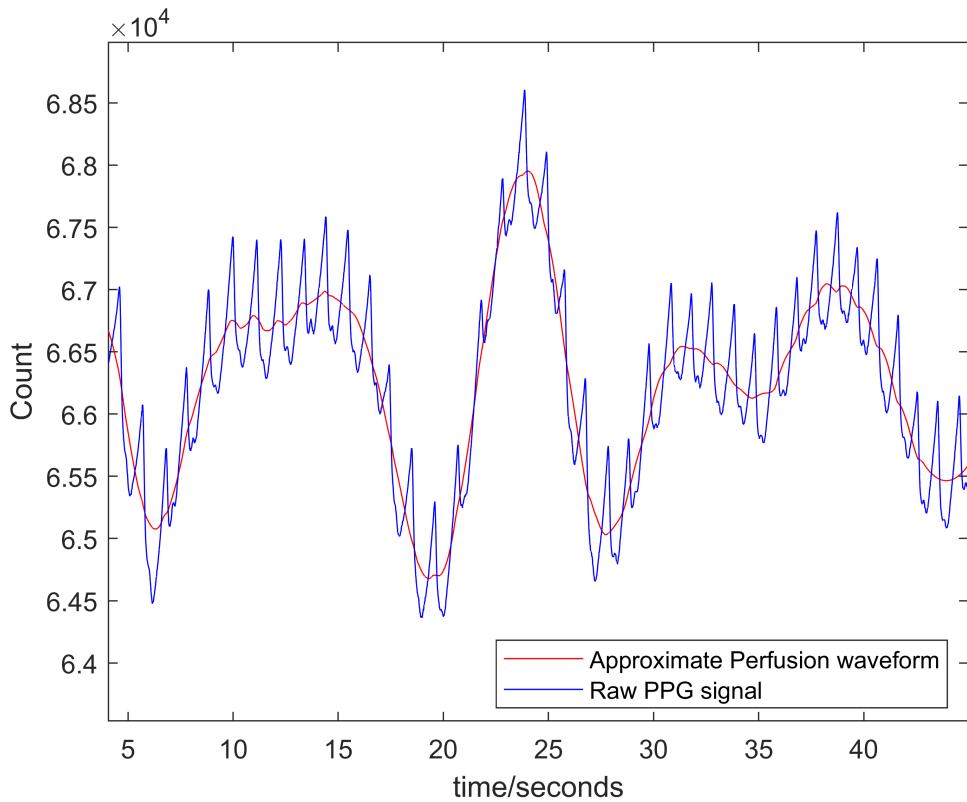


FIGURE 4.4: Raw PPG Signal (Blue) along with the approximated Perfusion Waveform (Red)

Similarly, during the diastole phase, more light is reflected as there is now less blood in the artery and hence a greater light current is produced. This is depicted by the positive gradient in the obtained MAXREFDES103 PPG waveform whereas a negative gradient in the universal PPG waveform shown in figure 2.1.

Therefore, to invert it, we simply need to subtract the PPG waveform from a threshold value. However, as the PPG signal is symmetric about 0, we can simply multiply it with -1 to achieve our PPG waveform. Since we are more interested in the R-R intervals, which can be found using the inverted PPG waveform, we do not need to invert the PPG signal as the peak finder algorithm will work fairly well in determining peaks with the diastolic peak out of the way. Usually the systolic and diastolic peak results in two local maxima under a define frame length and makes it harder for the algorithm to find the systolic peak only as the local maxima for the computation of R-R interval. The inverted PPG waveform will then provide us with the Pulse Wave Duration (PWD) which is proportional to the R-R interval as vital feature for post analysis of our PPG waveform.

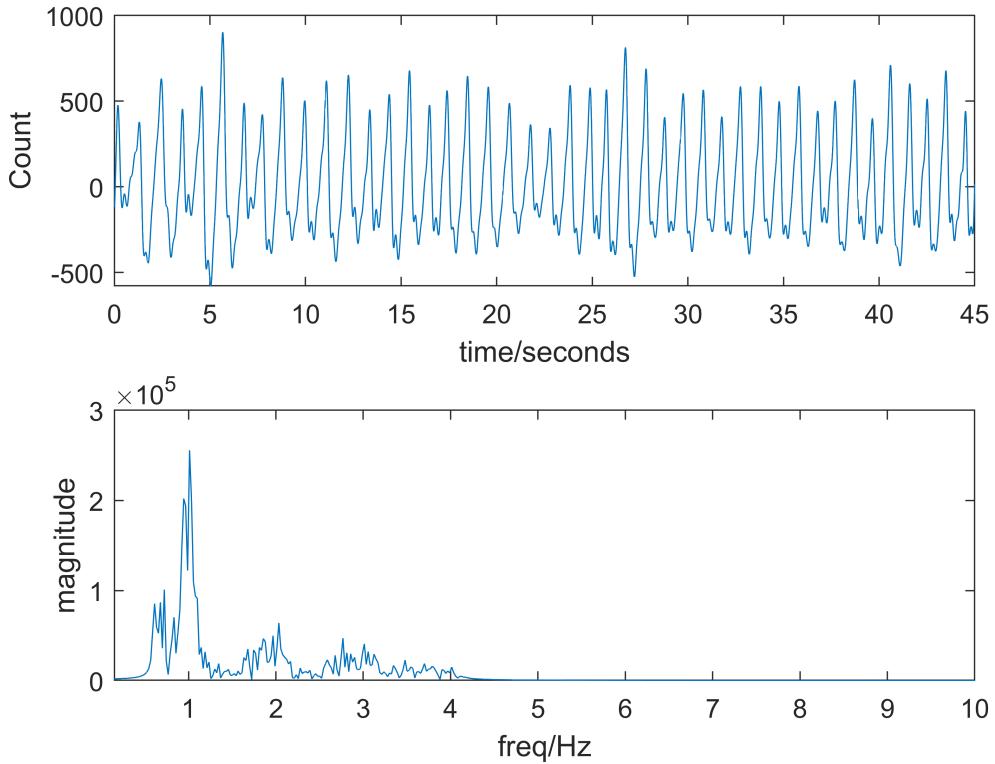


FIGURE 4.5: PPG Waveform obtained along with its Frequency Plot

4.3.2 Filtering PPG Waveform

This method prioritizes the removal of the PPG waveform from the raw PPG signal first and then using it to compute the perfusion waveform. Figure 4.6 shows us the time domain as well as the frequency domain plot of our raw PPG signal for the same individual under the same time window of 45 seconds. The frequency domain plot depicts frequencies of perfusion waveform that dominate the region ≤ 0.5 Hz. The peaks at approximately 1 Hz and 2 Hz correspond to the rhythmic motion of the heart.

Firstly, we remove the DC component from our raw PPG signal by subtracting the median of the data set from all sample points. This step is important as it ensures that once the filter is applied on the signal, the impulse response of the filter itself is not shown at the output as a transient response. Next, a band pass filter is applied with $f_l = 0.5$ Hz and $f_h = 4$ Hz to obtain the PPG waveform. This filtered signal is then subtracted from the original signal to obtain the corresponding perfusion waveform as depicted in equation 4.3. The waveform is then smoothed to remove any unwanted noise using Savitzky-Golay filtering with a 2nd order polynomial at a frame length of 257 samples. The results are shown in figure 4.7

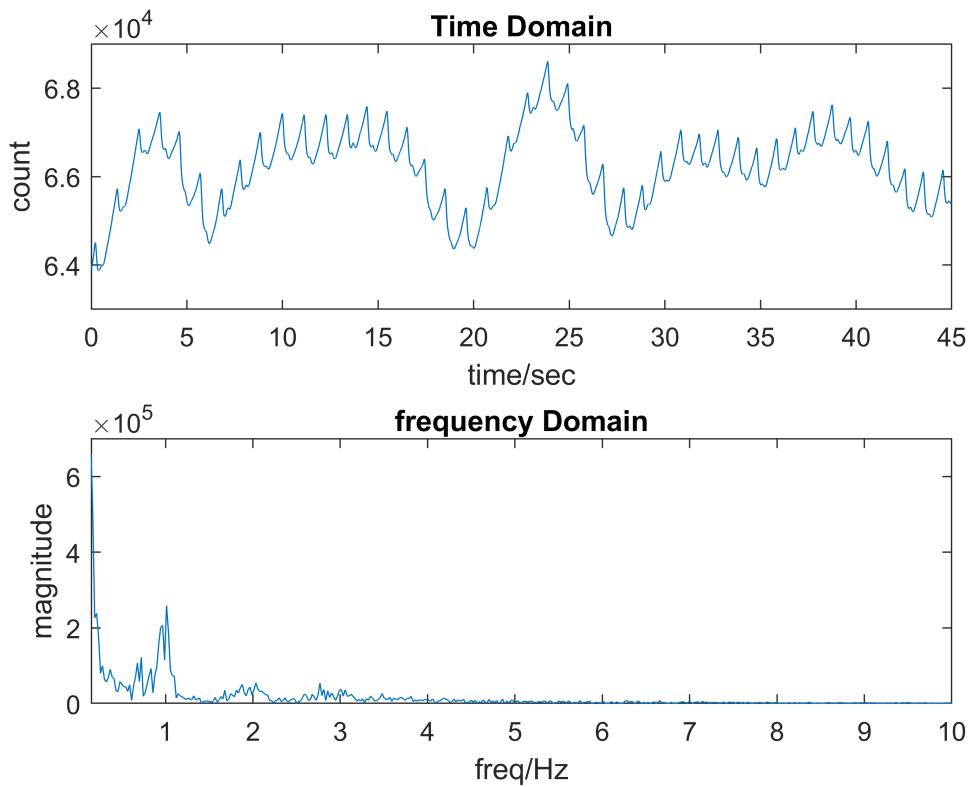


FIGURE 4.6: Raw PPG Signal and its Frequency Plot
from MAXREFDES103

Note here that the PPG waveform is inverted. Both methods work equally well in separating the perfusion and the PPG waveform from the raw PPG signal. However, as the perfusion variation is most susceptible to motion artifacts, there are sudden spikes in the variation of perfusion which might leave method 1 unable to accurately approximate the perfusion waveform. Moreover, since the PPG device maxem record the readings under skin contact, the waveform is disrupted and falls to zero when the skin contact is broken as a result of unintentional movement by the user. Therefore, method 2 of filtering out the PPG waveform in its frequency range is more preferred under such circumstances. Moreover, the PPG waveform that we have separated is subject to noise as seen in the frequency spectrum shown in. We will address this issue in our SPROJ II while applying data algorithms to account for motion artifacts.

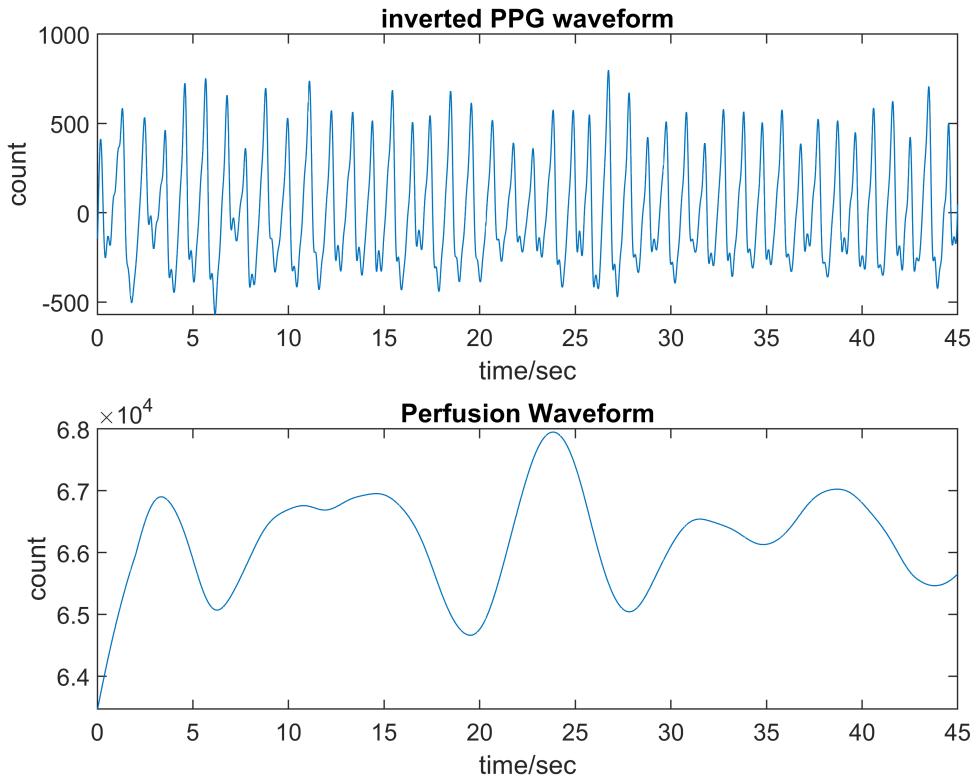


FIGURE 4.7: PPG Waveform and Perfusion Waveform obtained by filtering using a bandpass filter

4.4 Motion Artifacts Removal

4.4.1 Analysis using the Accelerometer Signal

The PPG waveform is prone to motion artifacts and a method needs to be implemented to account to make the PPG more robust. Initially, we start off by finding a correlation between the PPG waveform and the frequencies of motion that are induced as a result of the movement from a healthy individual. For this analysis, we used the BVP recorded through the E4 along with the corresponding data of its 3-axis accelerometer. The 3-axis accelerometer of the E4 records the readings at 32 Hz. As previously stated, the data was recorded in a controlled environment by the individuals to induce certain motions in a controlled pace over a time interval of approximately 5 minute and 30 seconds. The data sets were then plotted along with their magnitude spectrum to visualize the frequencies of motion. Before jumping straight to the analysis of the BVP with motion artifacts, Figure 4.8 shows the BVP plot taken from an individual at rest position under a 30 second window along with its magnitude spectrum with prominent frequencies

at approximately 1 Hz and 2 Hz. Moreover, the BVP signal also shows some variation due to perfusion which is depicted by the less prominent frequencies ≤ 0.5 Hz. These frequencies can be filtered out using a band pass filter with $f_l = 0.5$ Hz and $f_h = 4$ Hz as mentioned previously while separating perfusion waveform from the raw signal.

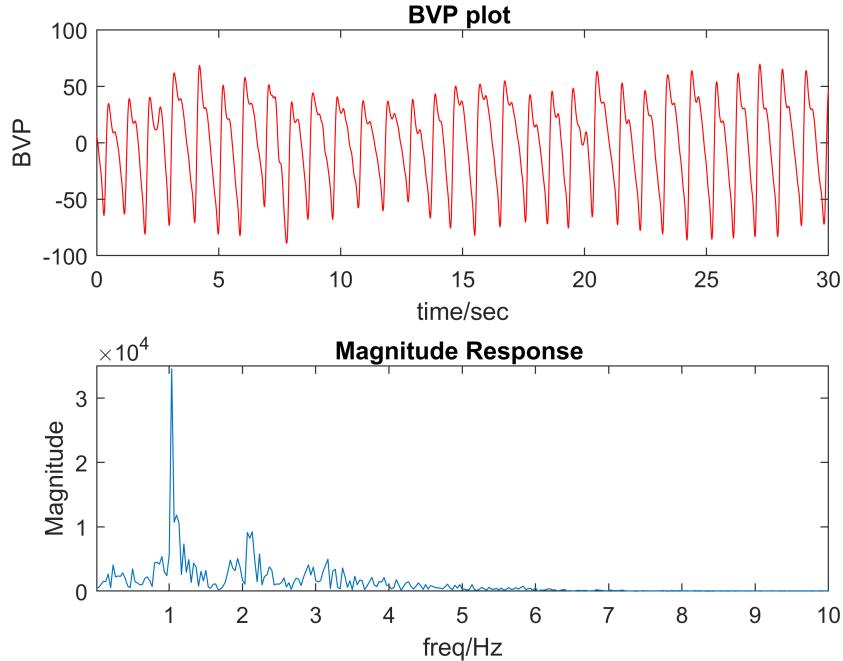


FIGURE 4.8: BVP of a healthy individual at rest along with its magnitude spectrum using the E4

The BVP plot of the same individual with a controlled vertical motion (arm movement up and down in a controlled pace) under a time window of 30 seconds is shown in figure 4.9. Introduction of motion has left our BVP signal subject to noise and unwanted disruption.

As can be observed from the accelerometer data and its corresponding frequencies, the frequency at approximately 0.43 Hz is introduced in the Y axis accelerometer plot more prominently due to the nature of movement. Let's denote this frequency of motion as FM_1 . Observing the magnitude response of the BVP, there is an added frequency more prominent at frequency = FM_1 corresponding to the frequency of motion from the accelerometer signal in comparison to our reference signal in figure 4.8 . In order to remove the frequency of motion at FM_1 , we apply an FIR Equiripple band stop filter of order 100 with bandwidth $\frac{FM_1}{4}$. The results are shown in Figure 4.10.

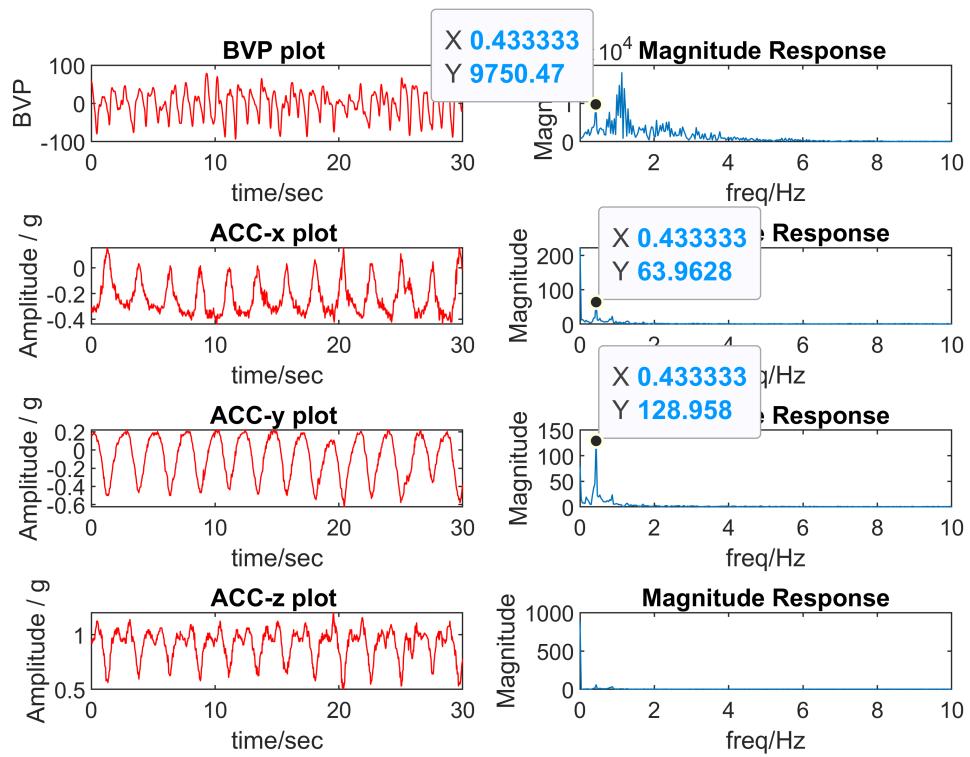


FIGURE 4.9: BVP of a healthy individual with motion along with its magnitude spectrum using the E4

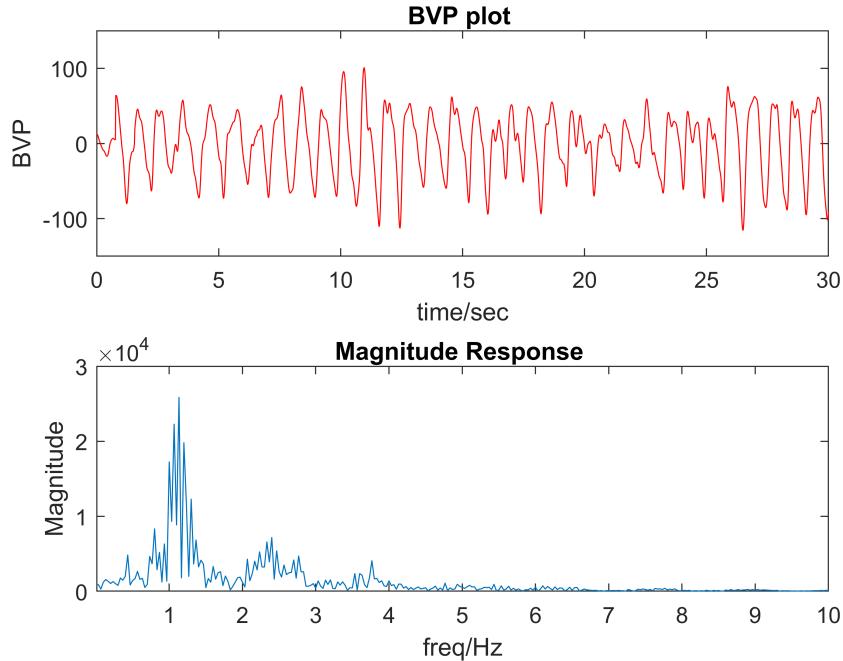


FIGURE 4.10: Filtered BVP using frequency of motion in the accelerometer data of E4

The BVP signal obtained after applying the band stop filter does seem to have improved in quality however, due to added frequencies of motion in between the PPG signal frequency range, the signal remains disrupted and unusable for analysis. The same implementation was applied on the rest of the data sets under different movements. Similar analysis was done on a data set taken using the MAXREFDES103 whose at first, inverted PPG waveform was extracted using the polynomial fitting method described in the previous section. The corresponding data from the 3 axis accelerometer was then used to detect frequencies of motion outside the frequency range for the PPG signal which were then filtered out using the FIR band stop filter. The results are shown in Figure 4.11

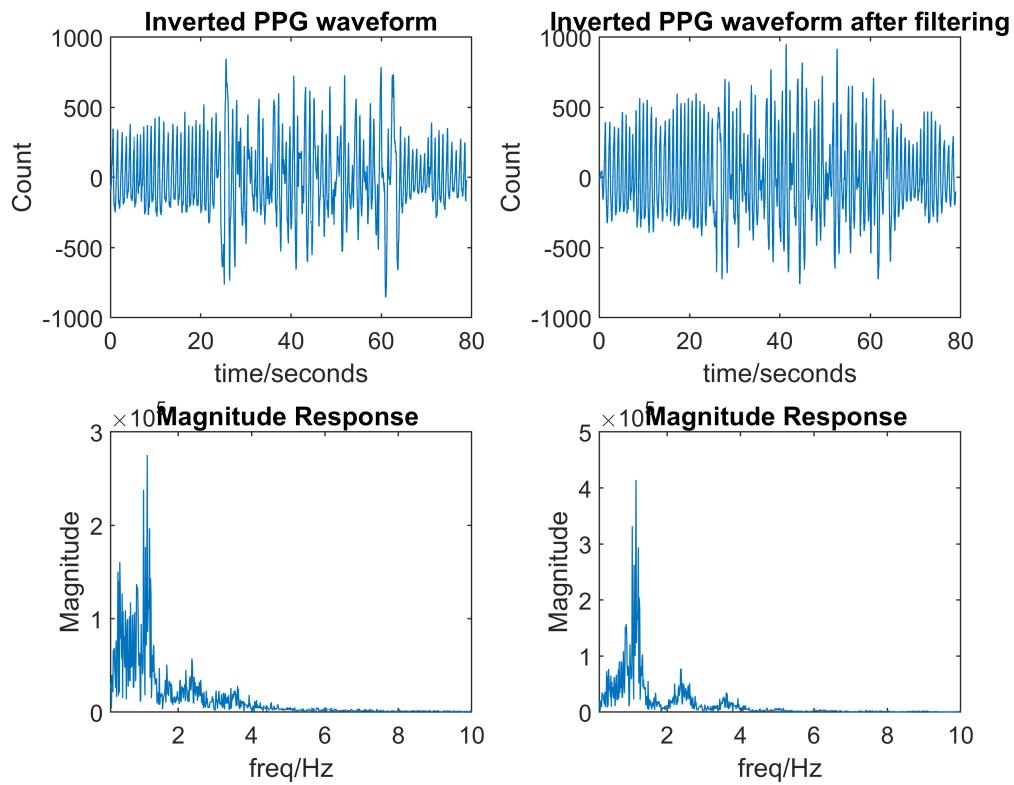


FIGURE 4.11: Comparison of inverted PPG waveform before and after eliminating frequencies of motion using MAXREFDES103

As per our analysis, all the prominent frequencies of motion that were filtered out lied in the range ≤ 1 Hz. However, the obtained BVP with motion artifacts showed added noise frequencies even above 1 Hz which cannot be filtered out using a band stop filter as there is a risk of losing vital PPG information. The results do show some improvement but requires more in depth analysis and processing of the PPG signal within its frequency range to get rid of these motion artifacts. Understanding the frequencies of motion and filtering them out such to ensure that

no vital information for our PPG signal is lost is one step towards the removal of motion artifacts.

4.4.2 Deep Learning

The approach is based on the paper written by Ke Xu, et al. which proposes that motion artifacts can be removed through modeling a non-linear trend between a cleansed PPG signal and motion artifact induced noise [21]. As explained in chapter 2, the paper first utilizes an envelope filter to normalize the training data and to remove training imbalances caused by respiration. Next, a TDNN is developed to model PPG signal-noise interaction and introduce time dependence information for motion artifact removal. Currently, we have begun developing an algorithm to simulate the results achieved by the above mentioned paper, but, as of yet, have not reached any conclusive outcomes. We aim to continue exploring this approach in conjunction with investigating other potential data processing algorithms that can eliminate motion artifacts from raw PPG data and then compare the quality of the processed PPG signal produced by each approach to determine the most effective motion artifact removal algorithm.

4.5 Parameter Estimation

4.5.1 Breath Rate

As mentioned earlier, the perfusion waveform from the raw PPG signal originates due to variation in the blood flow through the capillary network in the blood tissue subject to an individual breathing rate, anxiety levels, hormone levels, stress, sleep quality, and many other factors. Hence, making the perfusion waveform a pool of hidden information that can be used to compute vital cardiovascular and other health parameters. Among these is the breathing rate, which is measured in breaths per minute. The respiratory signal that corresponds to the breathing is dominated by frequencies at 0.3 Hz [24]. Therefore, a bandpass filter is applied at $FM = 0.3$ Hz with a frequency bandwidth of $FM/4$. This is followed by a peak finder algorithm of MATLAB under a window of 64 samples to find local maxima of the respiratory signal. Figure 4.12 shows the respiratory signal of a healthy individual over a period of 70 seconds at rest position.

The distance between each peak represents the time taken for one breath in samples. To approximate the breath rate, the distances between successive peaks is found and stored in a vector \vec{v} in number of samples. The following formula

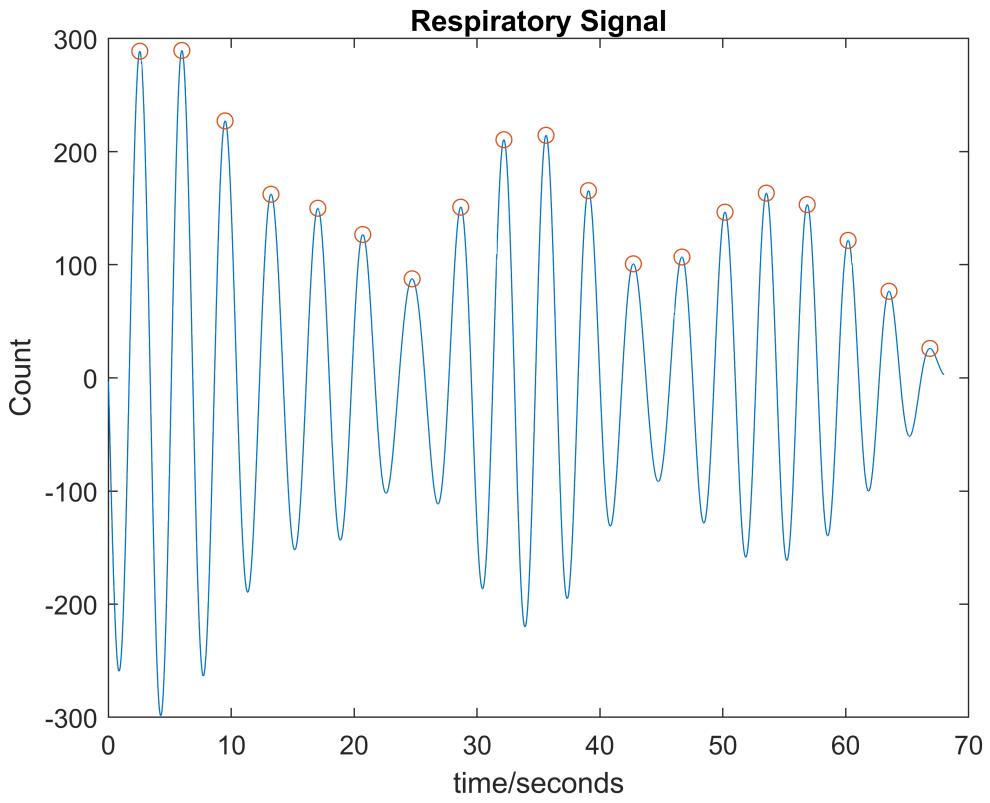


FIGURE 4.12: Respiratory Signal at 0.3 Hz with Local Maxima obtained

converts the average of vector \vec{v} into time format by:

$$\text{Breath rate per second} = \frac{\text{sample rate}}{\text{mean breath rate in samples}} \quad (4.5)$$

Where **sample rate** is 64 Hz. The breath rate per second can be converted to minute by multiplying 60 by it and rounding off to the nearest integer. Hence, we came up with an approximation for our breath rate per minute using the perfusion waveform of an individual.

This algorithm was run on the raw PPG signal of 4 test subjects. The subjects were asked to count the number of breaths they take over the recording time period in seconds. Their breath rate in minutes was then mathematically computed and stored as a reference value. The raw PPG signal was then separated into its perfusion waveform using the above-mentioned approach, on which then the breath rate algorithm ran to approximate their breath rates in breaths per minute. Overall, the breath rate was within the range of ± 1 breath per minute. We planned on testing this algorithm on more data sets, but, due to the ongoing situation of COVID-19, we were not able to take data set recording of individuals on campus. However, we hope to create a more diverse data set to compare our results and

test for the accuracy of the scores achieved. Moreover, the same data set will also be used to approximate other hidden parameters in the perfusion waveform in our SPROJ II.

4.5.2 Pulse Wave Duration Interval

Similarly, the Pulse Wave Duration can be found using the peak finder command in MATLAB with *MinPeakDistance* set as $\frac{\text{Sampling Rate}}{2}$.

The theory behind this analogy is that for an individual at rest, the average heart rate never exceeds the boundary of 60 to 100 beats per minute. Therefore, the maximum number of beats per second that can be expected is $\frac{100}{60}$ i.e. approximately 1.6 beats hence one beat in every 40th sample which is greater than and approximately close to $\frac{\text{Sampling Rate}}{2}$ i.e. 32. However, when the PPG data of an individual is recorded under the influence of motion artifacts, the peak finder algorithm will not work as expected due to a disrupted waveform. Therefore, in order to accurately approximate the Pulse Wave Duration using an adaptive algorithm, it is vital that the disruptions/noise due to motion are first extracted ensuring that the shape of the PPG waveform is retained for analysis.

4.6 Results Analysis and Future work

In SPROJ I we have been able to separate the perfusion waveform and the PPG waveform from the raw PPG signal using polynomial fitting and filtering, whilst comparing the PPG data collected from a high-end and low-end PPG devices. We also established a relation between the PPG data along with its corresponding accelerometer data under the influence of motion artifacts by analyzing the frequencies of motion. Although, filtering out these frequencies of motion do clean our data to some extent, it does not fully remove the noise from the PPG waveform. In SPROJ II we aim to further our progress on this approach by continuing our working on investigating methods for the removal of motion artifacts. Moreover, we will also continue our working on analyzing and extracting vital health parameters from the perfusion variation.

4.6.1 Motion Artifact Removal

The main highlight for SPROJ II will be the removal of motion artifacts from PPG signals. In SPROJ I we have already begun work in this regard by investigating the use of deep learning algorithms, such as, the TDNN to achieve this. In SPROJ II, we intend to analyze different data processing algorithms in order to develop a data processing algorithm that can recognize and extract these motion inducing noise

from the PPG without losing any vital information .The cleansed more robust PPG signal can then be used to compute cardiac parameters from which it can be possible to identify potential cases of Afib and other cardiac conditions.

4.6.2 Health Parameter Approximation from Perfusion Variations

As mentioned before, the perfusion variation is directly influenced by characteristics, including, stress levels, sleep quality, hormonal levels, and SpO2. While we have been able to approximate the breathing rate, we intend to take this a step further by extracting other hidden information from perfusion variations so that other vital health parameters can be approximated as well. Therefore, an in-depth analysis will be conducted as we explore various methodologies been utilized for the extraction of such health parameters.

References

- [1] C. Fischer, M. Glos, T. Penzel, and I. Fietze, “Extended algorithm for real-time pulse waveform segmentation and artifact detection in photoplethysmograms,” *Somnologie*, vol. 21, pp. 1–11, 05 2017.
- [2] “Max86140evsys - evaluation board, max86140 optical pulse oximeter and heart-rate sensor, data logging capabilities.” [Online]. Available: <https://il.farnell.com/maxim-integrated-products/max86140evsys/eval-board-optical-data-acquisition/dp/2778131>
- [3] S. Naseem, U. K. Khattak, H. Ghazanfar, and A. Irfan, “Prevalence of non-communicable diseases and their risk factors at a semi-urban community, Pakistan,” *Pan Afr Med J*, vol. 23, p. 151, 2016.
- [4] S. Lee, G. Ha, D. Wright, Y. Ma, E. Sen-Gupta, N. Haubrich, P. Branche, W. Li, G. Huppert, M. Johnson, B. Mutlu, K. Li, N. Sheth, J. Wright, y.-s. Huang, M. Mansour, J. Rogers, and R. Ghaffari, “Highly flexible, wearable, and disposable cardiac biosensors for remote and ambulatory monitoring,” *npj Digital Medicine*, vol. 1, 12 2018.
- [5] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, “A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology,” in *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2017.
- [6] A. Alian and K. Shelley, “Photoplethysmography,” *Best Practice Research Clinical Anaesthesiology*, vol. 28, 09 2014.
- [7] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, “Deep learning for healthcare applications based on physiological signals: A review,” *Comput Methods Programs Biomed*, vol. 161, pp. 1–13, Jul 2018.

- [8] A. Aliamiri and Y. Shen, “Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor,” in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2018, pp. 442–445.
- [9] C. Yang, C. Garcia, J. Rodriguez-Andina, J. Farina, A. Iñiguez, and S. Yin, “Using ppg signals and wearable devices for atrial fibrillation screening,” *IEEE Transactions on Industrial Electronics*, vol. PP, pp. 1–1, 01 2019.
- [10] T. Pereira, N. Tran, K. Gadhouni, M. Pelter, D. Do, R. Lee, R. Colorado, K. Meisel, and X. Hu, “Photoplethysmography based atrial fibrillation detection: a review,” *npj Digital Medicine*, vol. 3, 12 2020.
- [11] P. L. Ricketts, A. V. Mudaliar, B. E. Ellis, C. A. Pullins, L. A. Meyers, O. I. Lanz, E. P. Scott, and T. E. Diller, “Non-invasive blood perfusion measurements using a combined temperature and heat flux surface probe,” *International Journal of Heat and Mass Transfer*, vol. 51, no. 23-24, p. 5740–5748, 2008.
- [12] M. Nitzan, I. Faib, and H. Friedman, “Respiration-induced changes in tissue blood volume distal to occluded artery, measured by photoplethysmography,” *Journal of Biomedical Optics*, vol. 11, no. 4, p. 040506, 2006.
- [13] S. Nabavi and S. Bhadra, “A robust fusion method for motion artifacts reduction in photoplethysmography signal,” *IEEE Transactions on Instrumentation and Measurement*, vol. PP, pp. 1–1, 07 2020.
- [14] B. Sun, C. Wang, X. Chen, Y. Zhang, and H. Shao, “Ppg signal motion artifacts correction algorithm based on feature estimation,” *Optik*, vol. 176, pp. 337 – 349, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030402618313883>
- [15] A. Tharwat, “Independent component analysis: an introduction,” *Applied Computing and Informatics*, 08 2018.
- [16] B. Kim and S. k. Yoo, “Motion artifact reduction in photoplethysmography using independent component analysis,” *Biomedical Engineering, IEEE Transactions on*, vol. 53, pp. 566 – 568, 04 2006.
- [17] Jianchu Yao and S. Warren, “A short study to assess the potential of independent component analysis for motion artifact separation in wearable pulse oximeter signals,” in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, pp. 3585–3588.

- [18] Y. Ye, Y. Cheng, W. He, M. Hou, and Z. Zhang, “Combining nonlinear adaptive filtering and signal decomposition for motion artifact removal in wearable photoplethysmography,” *IEEE Sensors Journal*, vol. 16, no. 19, pp. 7133–7141, 2016.
- [19] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy, “A novel approach for motion artifact reduction in ppg signals based on as-lms adaptive filter,” *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 5, pp. 1445–1457, 2012.
- [20] P. Regalia, *Adaptive IIR filtering in signal processing and control*. Routledge, 2018.
- [21] K. Xu, X. Jiang, and W. Chen, “Photoplethysmography motion artifacts removal based on signal-noise interaction modeling utilizing envelope filtering and time-delay neural network,” *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3732–3744, 2020.
- [22] J. Lee, K. Matsumura, K. Yamakoshi, P. Rolfe, S. Tanaka, and T. Yamakoshi, “Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion,” *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2013, pp. 1724–1727, 2013.
- [23] J. L. Moraes, M. X. Rocha, G. G. Vasconcelos, J. E. Vasconcelos Filho, V. H. C. de Albuquerque, and A. R. Alexandria, “Advances in Photoplethysmography Signal Analysis for Biomedical Applications,” *Sensors (Basel)*, vol. 18, no. 6, Jun 2018.
- [24] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. F. Pimentel, A. E. W. Johnson, J. Alastrup, L. Tarassenko, P. J. Watkinson, R. Beale, and D. A. Clifton, “Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review,” *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 2–20, 2018.