

# Removal of Motion Artifacts from Photoplethysmogram Data

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 for the traditional electrocardiogram (ECG) method to monitor a patient's heart activity, Photoplethysmogram (PPG) has become an essential alternative in the medical industry today. PPG data provides immense information to extract vital cardiovascular parameters such as Heart Rate (HR), R-R interval, and respiration rate (RR), which is used to identify common cardiovascular diseases and parameters. However, the biggest challenge in extracting these parameters from the PPG data is interference due to artifacts. The artifacts this project 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 challenge mentioned above to remove artifact-induced noise from raw PPG data, ensuring that the PPG waveform's original nature is retained. 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 to approximate vital health parameters such as breathing rate. Our approach will first compare the raw PPG data from a low-end device with that of the pre-processed data from a high-end device (which has a pre-implemented low-frequency noise artifact removal algorithm). The perfusion variation present within the raw PPG signal will be approximated and subtracted from this comparison's original raw signal. This perfusion signal will be used as the basis for the extraction of vital health parameters such as the one discussed earlier. Next, this project will propose a data processing algorithm to account for and classify the emergence of movement induced motion artifacts in the pre-processed PPG data obtained after removing the perfusion signal. Once all the windows containing the motion artifact within the signal are successfully identified and categorized into near-wrist and far-wrist motions depending on the severity of the motion artifacts detected. Both categories will have their distinct algorithms that will remove motion artifacts from the pre-processed PPG signal, cleansing the motion artifact induced portions of the PPG signal, giving us a clean PPG signal at the output. Finally, the results of the testing of our algorithm on different test subjects will be presented and evaluated by visualizing the distribution of kurtosis and skewness of the cleansed PPG signal and through a distinct performance index.

# **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

<b>Abstract</b>	<b>i</b>
<b>Acknowledgments</b>	<b>ii</b>
<b>Originality Certificate</b>	<b>iii</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Problem Statement . . . . .	2
1.3 Social Benefits and Relevance . . . . .	2
1.4 General Block Diagram . . . . .	3
1.5 Goals and Objectives . . . . .	3
1.6 Outcomes . . . . .	4
1.7 Timeline and Distribution of Work . . . . .	4
<b>2 Background</b>	<b>7</b>
2.1 Literature Review . . . . .	7
2.1.1 Photoplethysmogram . . . . .	7
2.1.2 PPG Waveform . . . . .	7
2.1.2.1 Optical Density of PPG Signals . . . . .	8
2.1.3 Perfusion Waveform . . . . .	9
2.1.4 Motion Artifacts . . . . .	10
2.1.5 Characterization of Motion Artifacts . . . . .	11
2.1.5.1 Effect of Tissue . . . . .	11
2.1.5.2 Effect of Venous Blood Movement . . . . .	11
2.1.6 Existing Methods . . . . .	12
2.1.6.1 Analysis using the Accelerometer Signal . . . . .	13
2.1.7 Data Processing Algorithms . . . . .	16
<b>3 Methodology and Tools</b>	<b>18</b>
3.1 System Level Design . . . . .	18
3.1.1 Pre-Processing . . . . .	19

3.1.2	PPG Signal . . . . .	19
3.1.2.1	Motion artifact detection . . . . .	19
3.1.3	Perfusion Signal . . . . .	19
3.1.3.1	Parameter estimation . . . . .	19
3.2	Tools and Instruments . . . . .	20
3.2.1	Simulation Software Packages . . . . .	20
3.2.2	Hardware Instruments . . . . .	20
3.2.2.1	Empatica E4 . . . . .	20
3.2.2.2	Maxim . . . . .	21
<b>4</b>	<b>Project Implementation</b>	<b>23</b>
4.1	Data Collection . . . . .	23
4.2	Data Quality Comparison E4 vs MAXREFDES103 . . . . .	25
4.3	Separating Perfusion and PPG waveform from Raw PPG Signal . . . . .	31
4.3.1	Approximate Perfusion Waveform using a Polynomial . . . . .	32
4.3.2	Filtering PPG Waveform . . . . .	34
4.4	Motion Artifacts Removal . . . . .	35
4.4.1	Detection of Motion Artifacts . . . . .	37
4.4.2	Removal of Motion Artifacts . . . . .	41
4.4.2.1	Far Wrist . . . . .	42
4.4.2.2	Near Wrist . . . . .	43
4.5	Experimentation and Results . . . . .	46
4.5.1	PPG Signal Cleansing . . . . .	46
4.5.1.1	Performance Index . . . . .	47
4.5.2	Parameter Estimation . . . . .	49
4.5.2.1	Breath Rate . . . . .	49
4.6	Graphical User Interface (GUI) . . . . .	51
<b>5</b>	<b>Cost Analysis, Future Work, and Conclusion</b>	<b>55</b>
5.1	Result and Cost Analysis . . . . .	55
5.1.1	Motion Artifact Removal . . . . .	56
5.1.2	Health Parameter Approximation . . . . .	56
5.2	Future work . . . . .	56
5.3	Conclusion . . . . .	57

# List of Figures

1.1	A general block diagram showing different stages of our project pathway	3
1.2	General timeline for our project	5
2.1	PPG waveform notation [1]	8
2.2	BVP for a healthy individual without motion artifacts	11
2.3	BVP for a healthy individual with motion artifacts	11
2.4	BVP of a healthy individual at rest along with its magnitude spectrum using the E4	14
2.5	BVP of a healthy individual at with motion along with its magnitude spectrum using the E4	14
2.6	Filtered BVP using frequency of motion in the accelerometer data of E4	15
2.7	Comparison of inverted PPG waveform before and after eliminating frequencies of motion using MAXREFDES103	16
3.1	System Level Design	18
3.2	Front and Back view of Empatica E4	21
3.3	MAXREFDES103	21
3.4	MAX86140EVSYS [2]	22
4.1	Gender distribution of Subjects	24
4.2	Age distribution among Subjects	24
4.3	Step wise data cleansing process	26
4.4	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	27
4.5	Plots of Raw PPG signal obtained under different configuration using MAXREFDES103 as mentioned in table 4.1.	29
4.6	Plot of PPG signal obtained under a 30 second window from the same individual with (a) MAXREFDES103 and (b) E4 .	31
4.7	Raw PPG Signal (Blue) along with the approximated Perfusion Waveform (Red)	33
4.8	PPG Waveform obtained along with its Frequency Plot	34
4.9	Raw PPG Signal and its Frequency Plot from MAXREFDES103	35

## *List of Figures*

4.10	PPG Waveform and Perfusion Waveform obtained by filtering using a bandpass filter . . . . .	36
4.11	Motion Artifact removal implementation . . . . .	36
4.12	Iteration through the PPG signal using overlapping of windows. The shaded area shows the percentage of overlap . . . . .	38
4.13	Spread of Kurtosis and Skewness for a healthy subject at rest over a period of 25 seconds . . . . .	39
4.14	. . . . .	40
4.15	PPG signal with motion artifacts windows detected . . . . .	41
4.16	Generation of PPG waveform by averaging . . . . .	42
4.17	Synthetically generated PPG waveform . . . . .	43
4.18	Adaptive Filter . . . . .	44
4.19	RLS Adaptive Filter . . . . .	44
4.20	Notch Filter . . . . .	45
4.21	PPG signal before vs after cleaning . . . . .	46
4.22	Spread of kurtosis and Skewness before vs after . . . . .	46
4.23	. . . . .	48
4.24	Respiratory Signal at 0.3 Hz with Local Maxima obtained . . . . .	50
4.25	Start window of GUI . . . . .	51
4.26	Upload window of GUI . . . . .	52
4.27	PPG window in GUI . . . . .	53
4.28	Perfusion window in GUI . . . . .	54

# List of Tables

4.1	Different configuration for MAXREFDES103 to achieve a sampling average rate of 64 Hz . . . . .	28
4.2	PI matrix for Kurtosis. Syntax $j_b$ refers to $j^{th}$ exercise PI before cleansing and Syntax $j_a$ refers to $j^{th}$ exercise PI after cleansing . . .	49
4.3	PI matrix for Skewness. Syntax $j_b$ refers to $j^{th}$ exercise PI before cleansing and Syntax $j_a$ refers to $j^{th}$ exercise PI after cleansing . . .	49

# 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 (AF) 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 by removing the motion artifacts 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 detect and successfully remove the motion artifacts from the noisy PPG signal, which can be of help in many ways but specifically in remote and ambulatory cardiac health monitoring and assessment. The PPG signals will be recorded using a wearable device worn on wrist. The main challenge in PPG based classification framework is 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 removal of such motion artifacts so an appropriate enough result can be obtained for medical use [5].

## 1.3 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, 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.

## 1.4 General Block Diagram

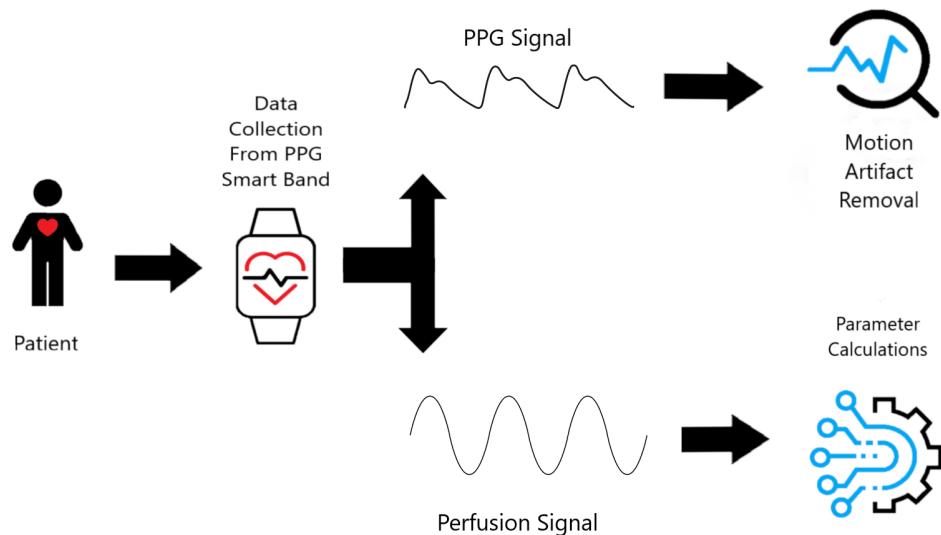


FIGURE 1.1: A general block diagram showing different stages of our project pathway

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 and used alternatively as to our convenience, mentioned in Chapter 3. We will then analyse and extract the perfusion waveform from the raw PPG signal obtained as part of the data collection and processing of the signal. This will be followed by vital health parameter calculations for useful feature extraction from perfusion waveform. The next stage will be responsible for the cleansed PPG signal extraction from the corrupted signal by detecting and removing motion artifacts using data processing algorithms.

## 1.5 Goals and Objectives

Our goals can be divided into following 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 artifacts present in the raw input data.
- Create and use the cleansed signal to compute vital health parameters such as breath rate by applying algorithms on perfusion signal.
- Detection and categorization of motion artifacts into near wrist and far wrist motion to allow for appropriate method selection for motion artifact removal

## 1.6 Outcomes

Our project will present the following outcomes:

- Separation of PPG and perfusion signal from raw PPG data obtained from a low-end device and processing it.
- Application of data processing algorithms on the processed PPG data to remove motion artifacts efficiently.
- Extraction of breath rate parameter from the perfusion signal using extensive algorithms.
- Live simulation of the complete process using PPG data of an individual by using a MATLAB based GUI (Graphical User Interface) that removes motion artifacts along with extracting and computing breath rate from perfusion.

## 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 about 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.

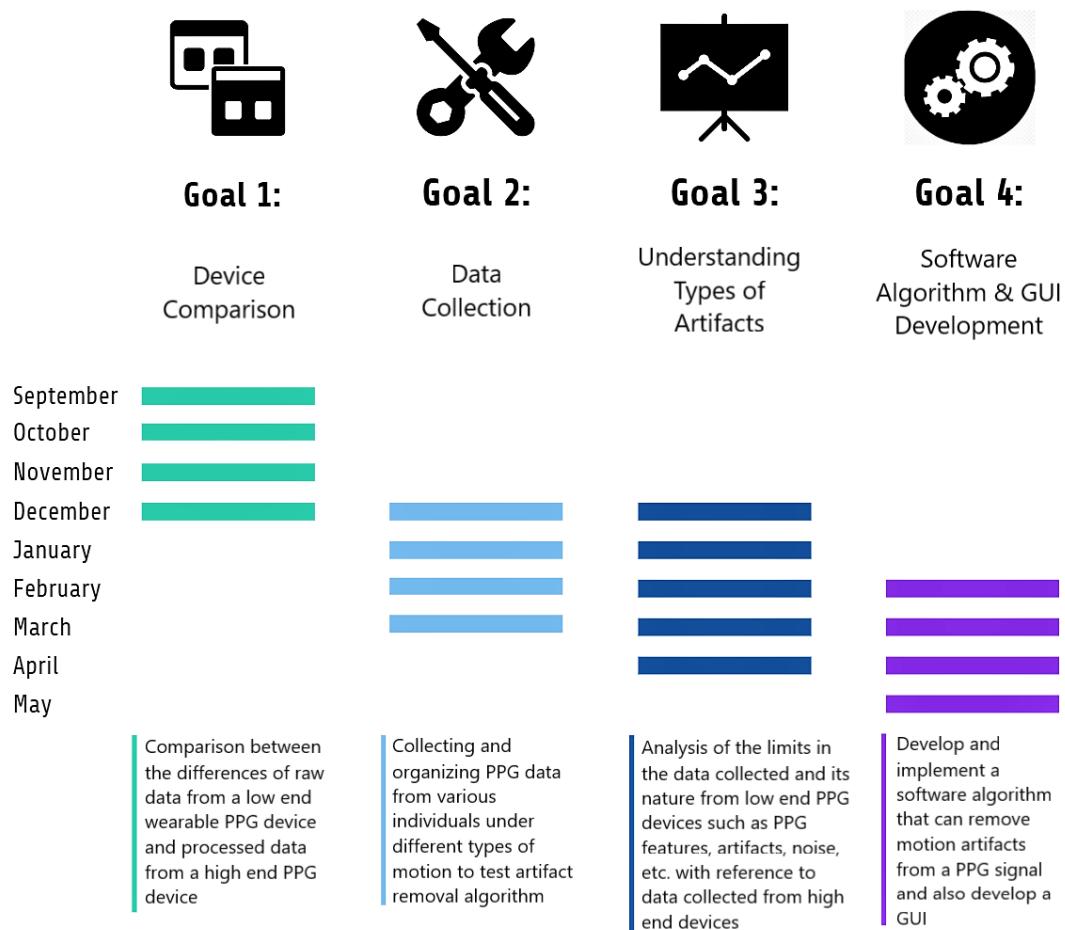


FIGURE 1.2: General timeline for our project

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. Goal 1, which began in September, will be concluded in December 2020 for the Senior Project 1 report.

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 an algorithm

that can successfully detect and remove artifacts from the raw PPG signal, primarily motion artifacts, using data processing techniques. 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 algorithm will start towards the end of the first semester and will conclude in May 2021. This will be accompanied with the GUI formation so that the whole process of separation, cleansing and extraction can be displayed.

# 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 heart related diseases. 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].

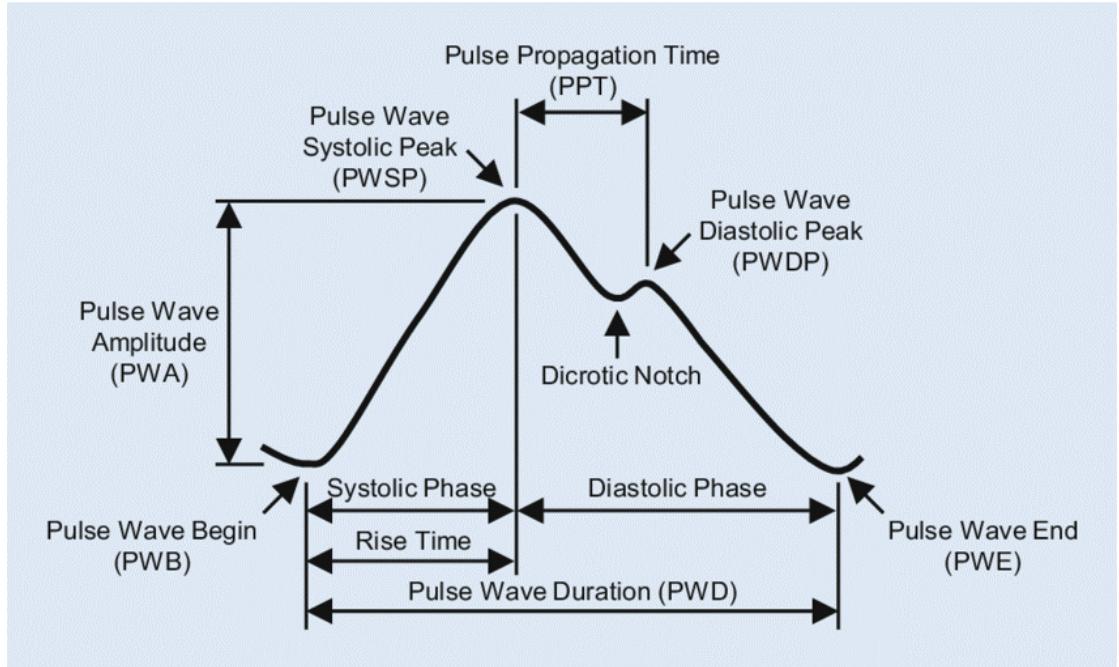


FIGURE 2.1: PPG waveform notation [1]

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 resourceful in detecting CVDs. Among them, one of the most vital extractable feature 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.

### 2.1.2.1 Optical Density of PPG Signals

PPG signals utilize the optical density transmitted and reflected on a material or surface to provide us with information regarding pulse oximetry. By using Schusters's theory on optical scattering, [11] and [12] states the following relationship to

explain the changes in optical density, A, as a function of changes in blood vessel thickness, D, for any particular blood vessel [13]:

$$\Delta A = [\sqrt{E_h(E_h + F)} \text{ Hb} + Z_b] \Delta D_b \quad (2.1)$$

F is defined as the scattering coefficient, Hb as the concentration of hemoglobin in blood,  $Z_b$  as a constant that is dependent on the optical receiver's width and independent of wavelength. Lastly,  $E_h$  is defined as  $E_h \approx SE_o + (1-S)E_r$ , where the extinction coefficients of oxygenated and de-oxygenated hemoglobin are represented by  $E_o$  and  $E_r$  respectively.

However, this equation was modified in [13] to take into consideration of tissue effect, which is a primary source of unwanted fluctuations in the PPG signal [11],[12] to become:

$$\Delta A = \sqrt{E_h(E_h + F)} \text{ Hb} \Delta D_b + Z_b \Delta D_b + Z_t \Delta D_t \quad (2.2)$$

$\Delta D_t$  indicates the change of thickness in tissue and  $Z_t$  is an approximated constant that is independent of wavelength. It should be known that  $Z_t \Delta D_t + Z_b \Delta D_b$  is wavelength independent [13].

Lastly, equation (2.2) is formulated for only one type of blood vessel, however, for this paper we need to take into consideration both venous and arterial blood vessels. As a result, the equation is further adjusted as shown below:

$$\Delta A = \sqrt{E_a(E_a + F)} \text{ Hb}_a \Delta D_a + \sqrt{E_v(E_v + F)} \text{ Hb}_v \Delta D_v + \Delta A_s \quad (2.3)$$

where venous blood vessels are indicated by subscript **v** and arterial blood vessels by **a** [13].  $\Delta A_s$  represents the wavelength independent components of blood and tissue changes. Based on the derived equation (2.3), we intend to propose an algorithm for identifying and cleansing of motion artifacts from PPG.

### 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 [14]. 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 [15]. 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 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 [16]. 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 2.2 shows the recorded BVP of the individual without motion artifacts in a 30 second interval and figure 2.3 shows the BVP of the same individual but with motion artifacts in a 30 second interval. Figure 2.3 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.

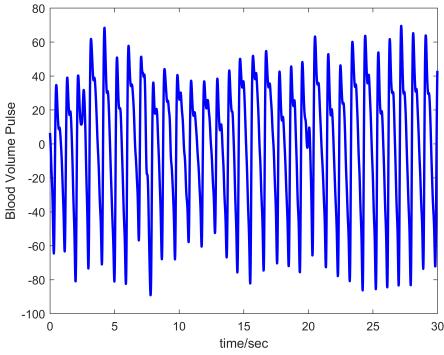


FIGURE 2.2: BVP for a healthy individual without motion artifacts

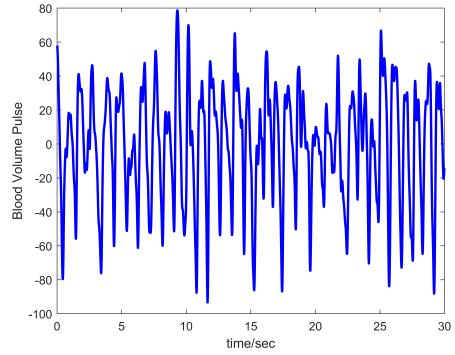


FIGURE 2.3: BVP for a healthy individual with motion artifacts

## 2.1.5 Characterization of Motion Artifacts

The motion artifacts can be characterized in the following types:

### 2.1.5.1 Effect of Tissue

As discussed in the previous section of optical density in PPG signals, the tissue effect is one of the primary sources of noise in a PPG signal. At rest,  $\Delta D_t$  acts as a DC component and can be easily be filtered out. However, in the presence of motion, such as walking or running, this component is no longer constant, resulting in noise-induced fluctuations in the PPG signal, causing errors in calculations of health parameters such as heart rate (HR) [13]. In order to tackle this problem, [13] proposes to remove the component of PPG which is wavelength independent but influenced by the tissue,  $\Delta A_s$ . This is done by subtracting two PPG signals from each other. The resultant equation can be seen below:

$$\Delta A_t = [\sqrt{E_{a1}(E_{a1} + F)} - \sqrt{E_{a2}(E_{a2} + F)}] Hb_a \Delta D_a + [\sqrt{E_{v1}(E_{v1} + F)} - \sqrt{E_{v2}(E_{v2} + F)}] Hb_v \Delta D_v \quad (2.4)$$

The subscripts **1** and **2** indicate the first and second PPG signals used. It should also be noted that when compared to the duration of human motion, the time difference between recordings becomes negligible.

### 2.1.5.2 Effect of Venous Blood Movement

Variation in venous blood movement is another cause of motion-induced error in PPG signals. This change results in inaccuracy when calculating health parameters such as oxygen saturation levels. This can be achieved through the weighted subtraction of the two PPG signals. The first step of the proposed approach in

[13] suggests that by considering  $\Delta A_t$  as our reference signal, we assume that  $\Delta A_s = 0$  for both PPG sources. As a result, the equation for weighted subtraction becomes,

$$\Delta A_{-v} = \Delta A_1 - \beta \Delta A_2 \quad (2.5)$$

where,  $\Delta A_1$  and  $\Delta A_2$  represent the two different PPG sources with  $\Delta A_s = 0$  for both sources. Through proper tuning of  $\beta$  this weighted subtraction can be used to remove either the arterial or venous component from the PPG signal [13], thereby allowing us to remove venous blood movement induced motion artifacts.

### 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 [16–18]. 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 [16, 19]. 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 [16, 20, 21].

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 [16, 22, 23]. This means that a correlation between the BVP and its corresponding 3 axis accelerometer data is assumed [16]. One approach of using adaptive filtering is through the use of an accelerometer as a noise reference for adaptive algorithm. 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, and the non availability of accelerometer data.

Another proposed approach relies on deep learning techniques to extract the physiological features from the corrupted BVP signal. The paper in [24] highlights that the difficulty in motion artifacts removal from corrupted BVP is attributed to the

non linearity of the cleansed BVP signal and noise. This technique proposes an algorithm based on Signal-Noise Interaction modeling (SniMA) for complex BVP signals prone to motion artifacts [24]. 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 [24]. 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 interaction between noise and the PPG signal as a non-linear process, but also, introduce time-dependence to aid in artifact removal. 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.

#### 2.1.6.1 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 2.4 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  $\leq 0.5$  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.

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 2.5. 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

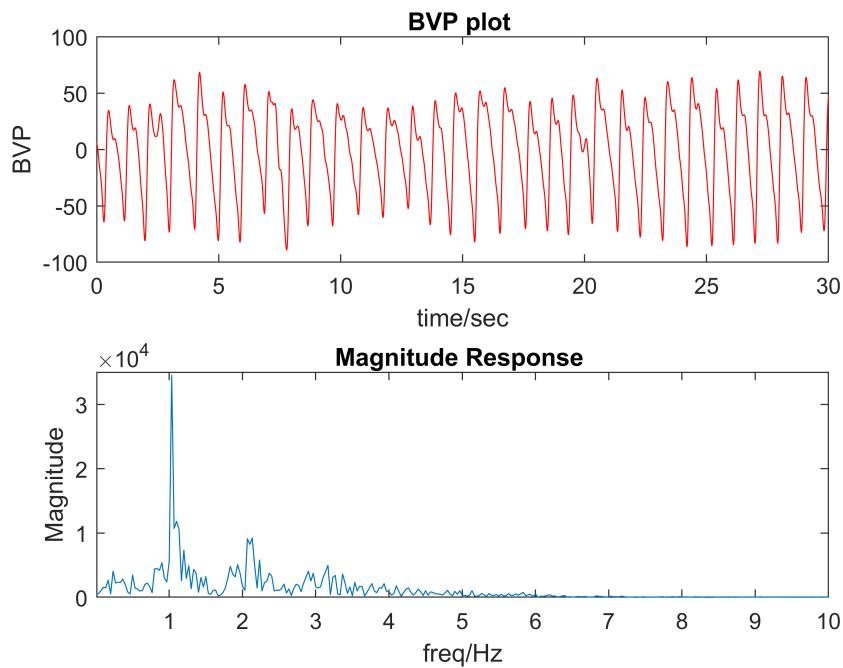


FIGURE 2.4: BVP of a healthy individual at rest along with its magnitude spectrum using the E4

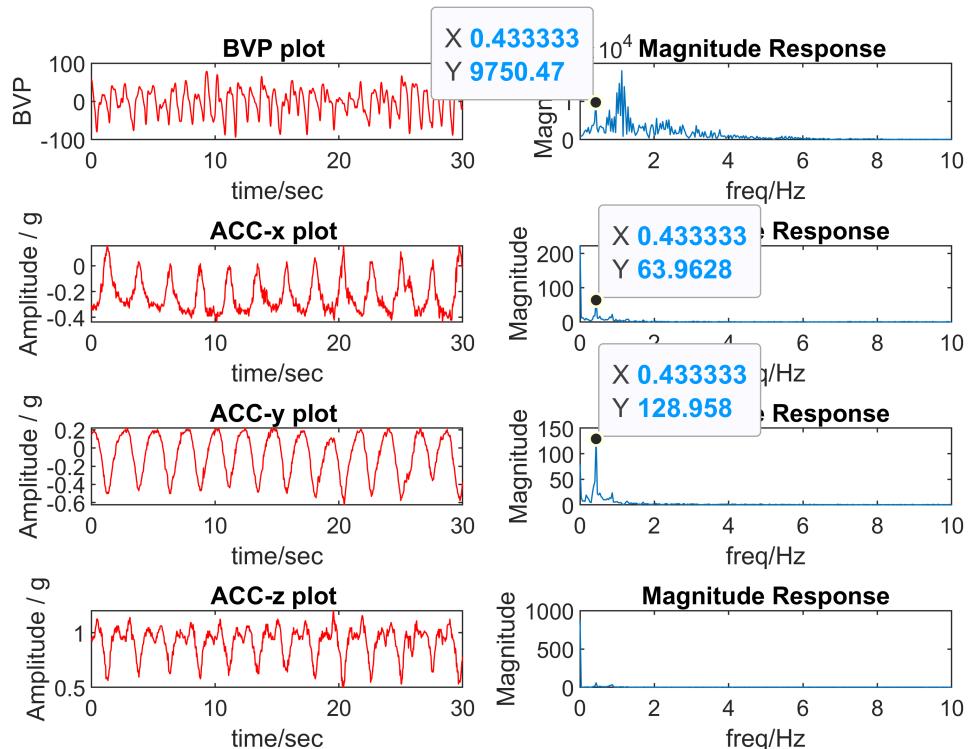


FIGURE 2.5: BVP of a healthy individual with motion along with its magnitude spectrum using the E4

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 2.4 . 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 2.6.

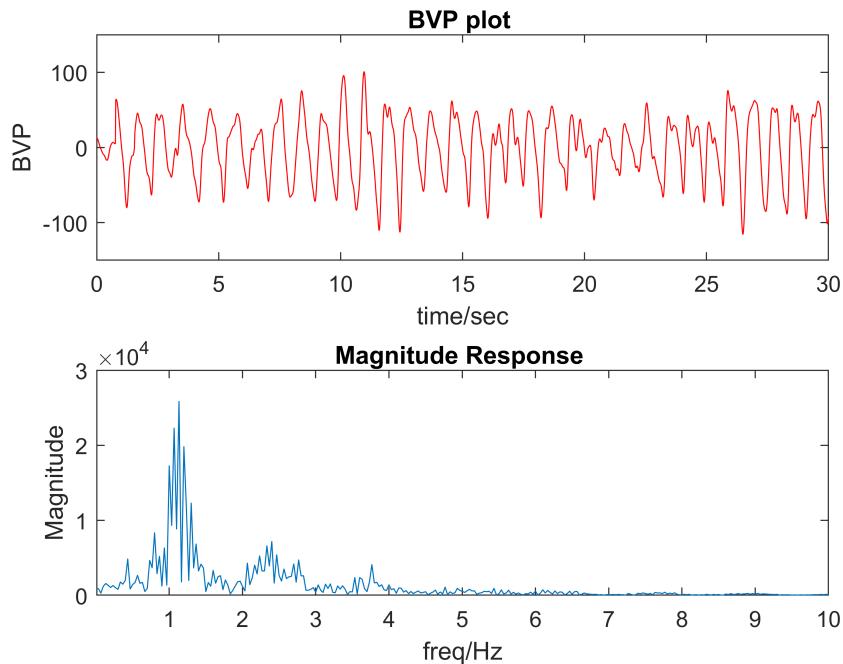


FIGURE 2.6: Filtered BVP using frequency of motion in the accelerometer data of E4

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 2.7

As per our analysis, all the prominent frequencies of motion that were filtered out lied in the range  $\leq 1$  Hz. However, the obtained BVP with motion artifacts

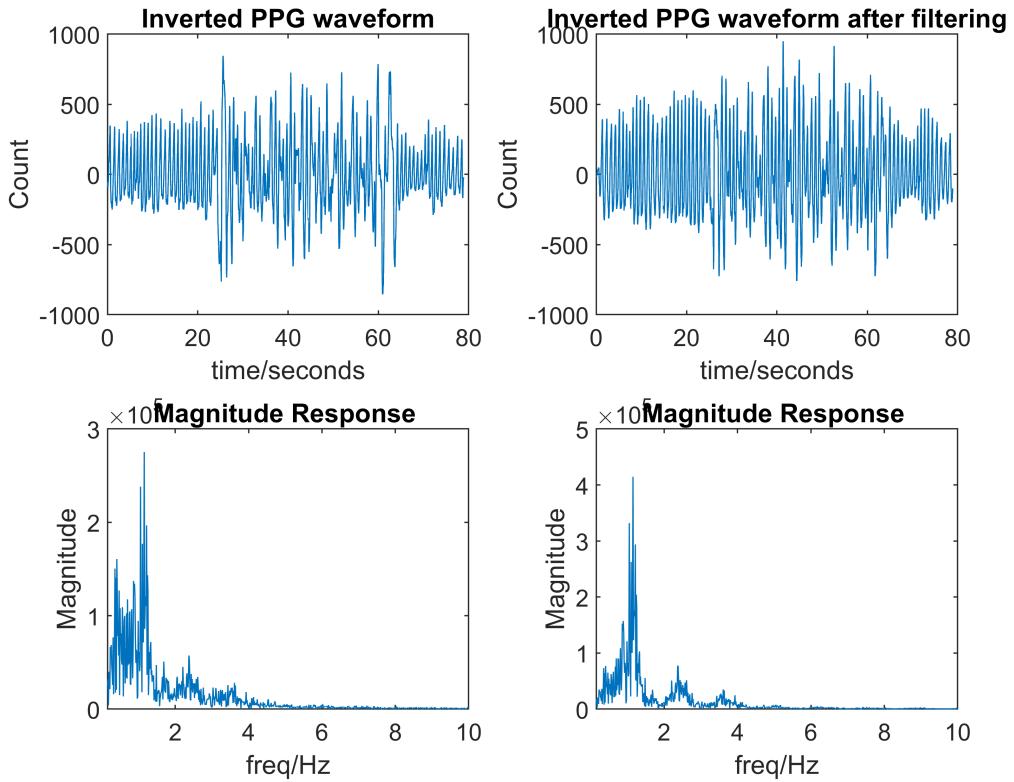


FIGURE 2.7: Comparison of inverted PPG waveform before and after eliminating frequencies of motion using MAXREFDES103

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 result 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.

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 detorius to one's health. Details of these algorithms and their result validation techniques will be discussed in-depth in chapter 4.

# Chapter 3

## Methodology and Tools

### 3.1 System Level Design

Our research areas mostly revolves around the PPG data which contain motion artifacts as well as perfusion which can be obtained from the raw PPG signal as shown in 3.1.

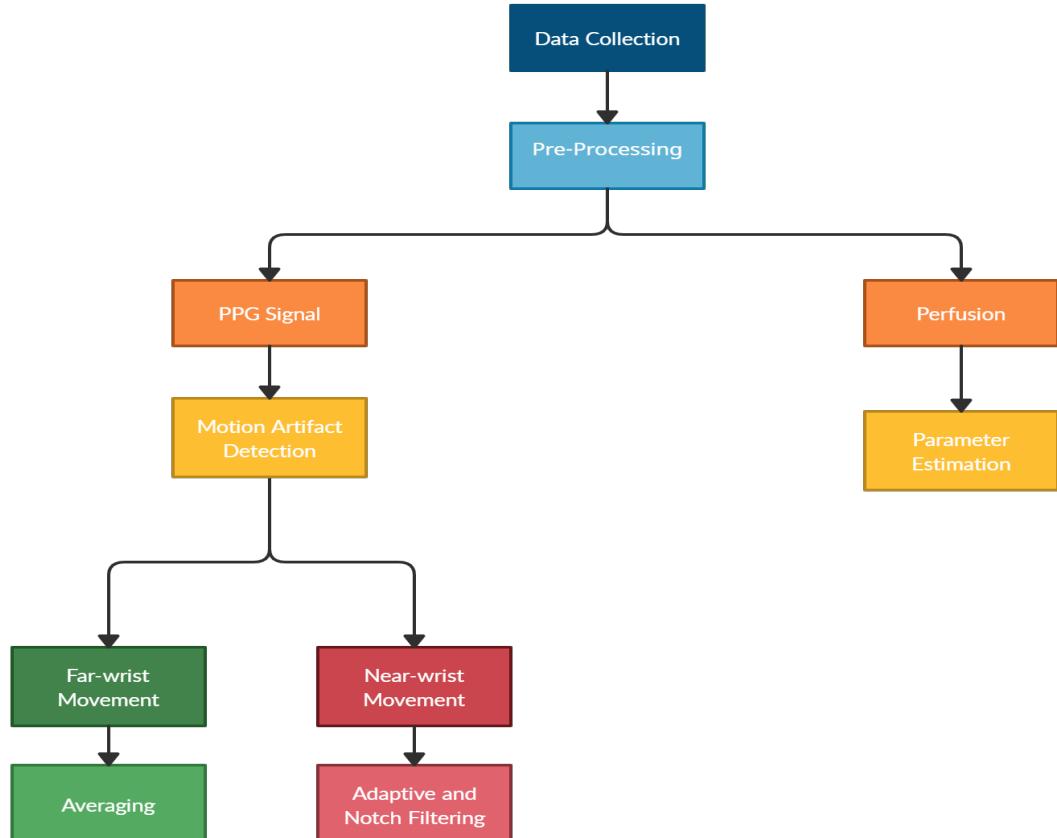


FIGURE 3.1: System Level Design

### 3.1.1 Pre-Processing

PPG signal and perfusion are separated from each other from the raw PPG signal obtained from the low-end PPG device. 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 O<sub>2</sub> 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 (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 PPG Signal

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.

#### 3.1.2.1 Motion artifact detection

Windows containing motion artifacts are detected and passed through an algorithm which removes the motion through one of the 2 different method based on the type of motion detected. If it is far wrist than simple averaging of the nearby signal works effectively and if its near wrist than adaptive filter is used for the removal of motion noise.

### 3.1.3 Perfusion Signal

Perfusion Signal obtained from the raw PPG signal contains information to a lot of vital health parameters such as that of Breath rate, Sp<sub>02</sub> and stress levels.

#### 3.1.3.1 Parameter estimation

Accurate Breath rate estimation was accomplished through the perfusion waveform.

## 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. We also developed our GUI on 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.

- **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.

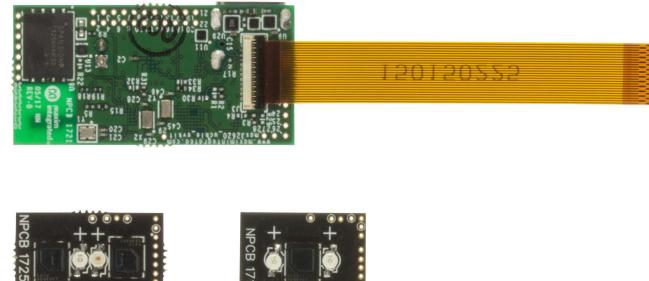


FIGURE 3.4: MAX86140EVSYS [2]

So far we did not require any graphics processing units for processing the data, all processing and analysis was done on 6<sup>th</sup> 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 Collection

Data collection and cleansing process was very extraneous and through with the main objective being obtaining data sets only with our concerned information:

- Our first phase of 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 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 in SPRoj 1.
- The second phase of data collection was much more elaborate and extraneous than the previous. The major change being the use of both MAXREFDES103 and E4 at the same time. This time the data was collected from 10 subjects. The subjects involved in our data collection process belonged to a pretty diverse group of age as age is one of the important factors that come into play when working with PPG and CVDs. Not only that but the pool of subjects was also varying among the gender classification and there was a 60/40 distribution. Figure 4.1 and 4.2 show the accurate distribution and the diversity within the pool of the subjects.

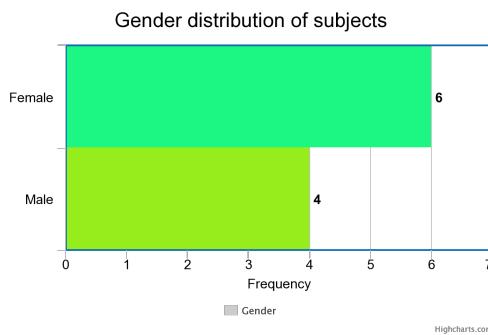


FIGURE 4.1: Gender distribution of Subjects

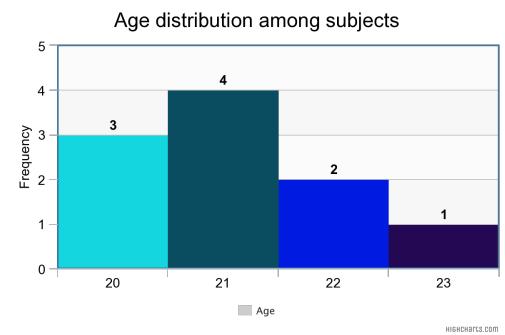


FIGURE 4.2: Age distribution among Subjects

Each subject was put through 6 different motion/exercises. Each motion was carried on for 90 secs with E4 on the non-dominant hand and MAXREFDES103 on the dominating one. Which means with each exercise we had 2 data sets (one for E4 and one for MAXREFDES103) hence for each subject we had 12 data sets ( $6*2=12$ ). This means for 10 subjects it is ( $12*10=120$ ) 120 data sets in total . All types of these motions were completely contrasting to one another and are summarized below:

- Rest position
- Walking
- Moving the dominant arm up and down within a controlled pace
- Moving the non-dominant arm up and down within a controlled pace while keeping the right hand still.

- Typing with the dominant hand (with MAXREFDES103)
- Writing with the dominant hand (hand with MAXREFDES103)

This phase was particularly time consuming as each of the data set were to be cleaned and stored in a way that it can be used directly in the future. The cleansed data should be so that both the data from same exercise was comparable (initially it was not as both devices provide data differently). The E4 data was already processed and free from perfusion whereas that wasn't the case for MAXREFDES103. So, in the first step PPG and perfusion separation was performed for each of the data set from MAXREFDES103 and stored in the same file. The next issue was that of unwanted parameters in the MAXREFDES103 data which were removed so that each file contains only the wanted features and hence can be compared directly to E4 data if required. Since the the data set was pretty large and number of files were an inconvenience to handle (12 files for each individual), we concatenated all the exercises of an individual from one device together into a single file (One file for MAXREFDES103 and one for E4). All if this cleansing was done through Python and it made the whole data handling more convenient. All steps of data cleansing in detail can be seen in figure 4.3

We tested 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 configuring 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  $n^{\text{th}}$  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  $n^{\text{th}}$  samples and denotes sample 2 as their computed average. Hence, the total averaged samples computed is given by (4.1).

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

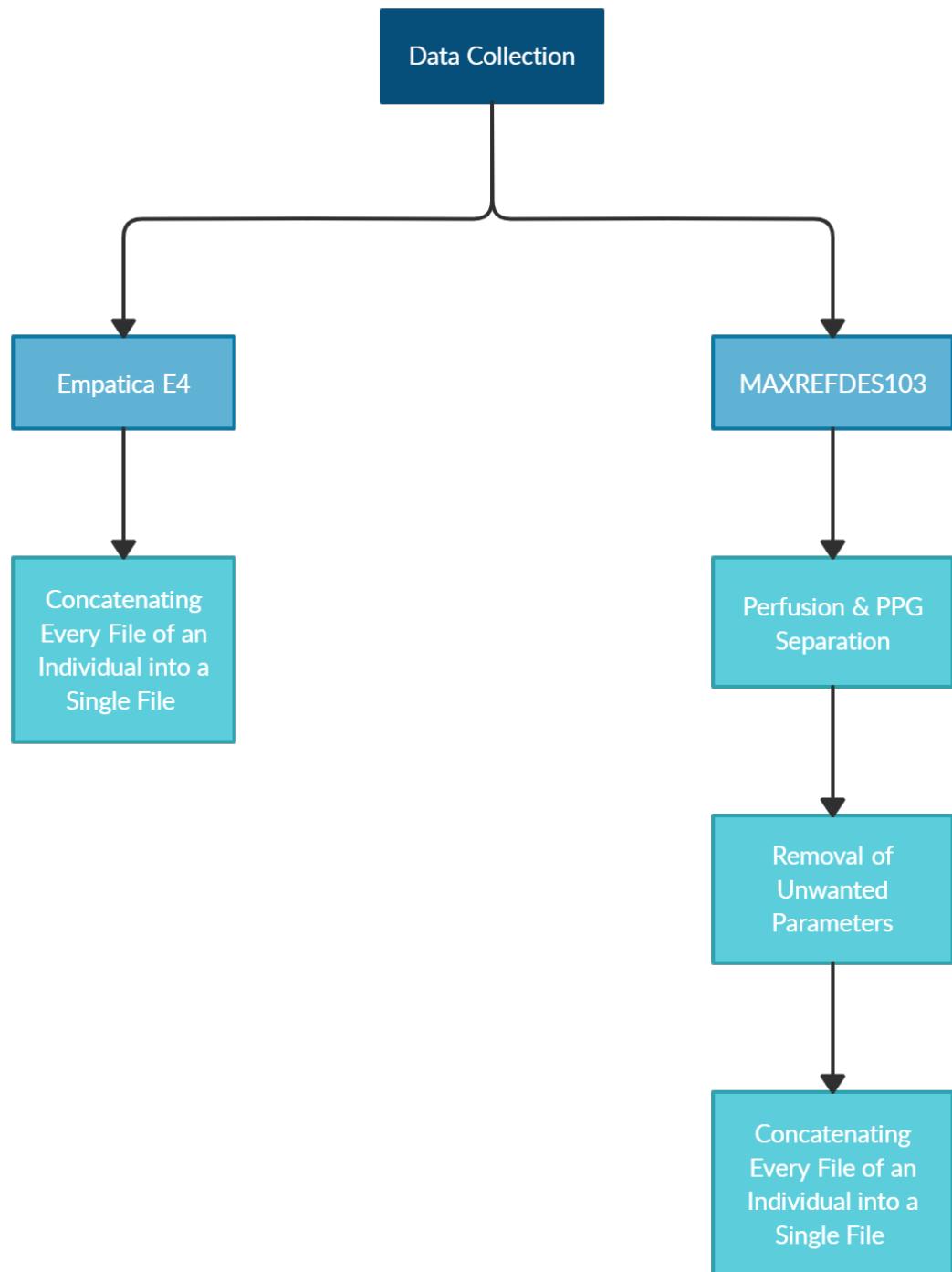


FIGURE 4.3: Step wise data cleansing process

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 4.4 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.

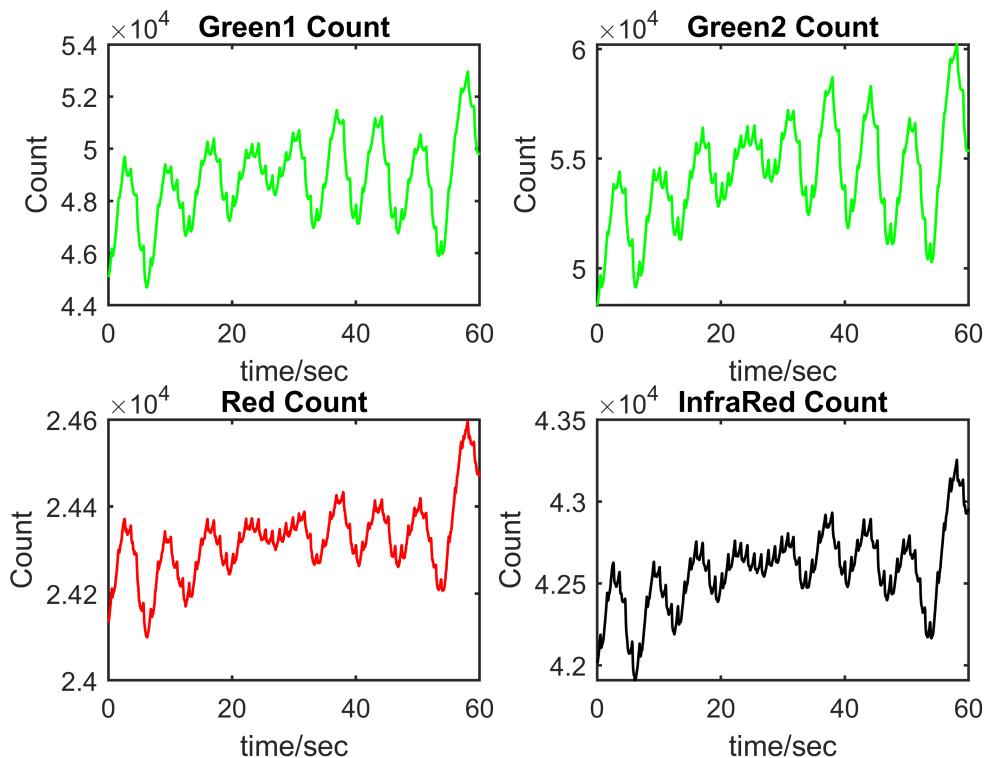


FIGURE 4.4: The Raw PPG signal obtained using Green, Red and Infra-Red light can be observed here. Green 1 and Green 2 data is obtained using a single source green LED light source but detected from two photodiode channels

The PPG signal in figure has been obtained using three different light sources from the MAXREFDES103 namely green, red and Infrared (IR). However, green<sub>1</sub> and green<sub>2</sub> 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 green<sub>1</sub> and green<sub>2</sub> 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 [25]. Comparing the DC components for the green plots, it can be observed that for green<sub>2</sub> count, the DC component is much higher as compared to the green<sub>1</sub> 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 4.5, 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

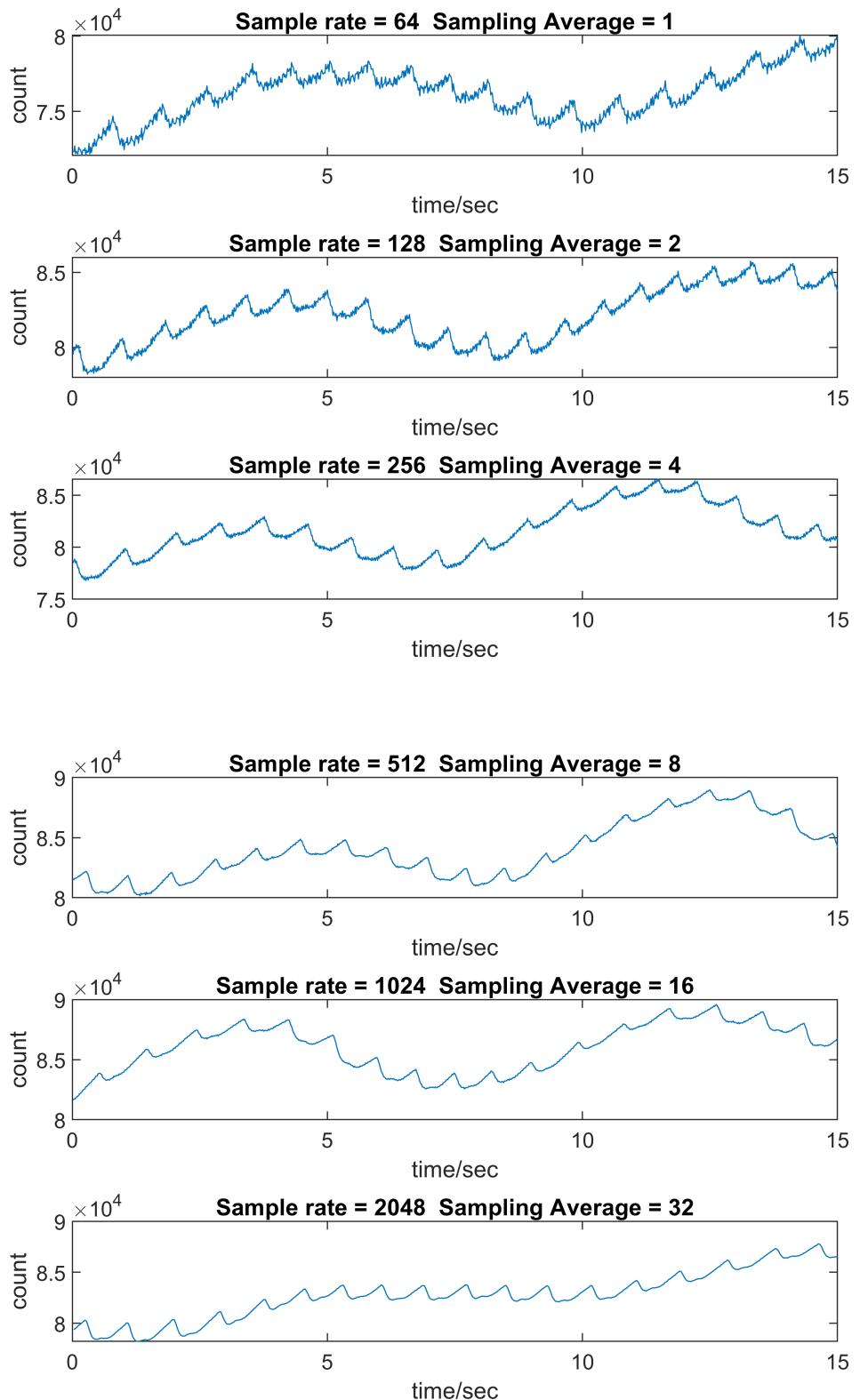


FIGURE 4.5: Plots of Raw PPG signal obtained under different configuration using MAXREFDES103 as mentioned in table 4.1.

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 4.5 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 MAXREFDES103 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 4.6 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:

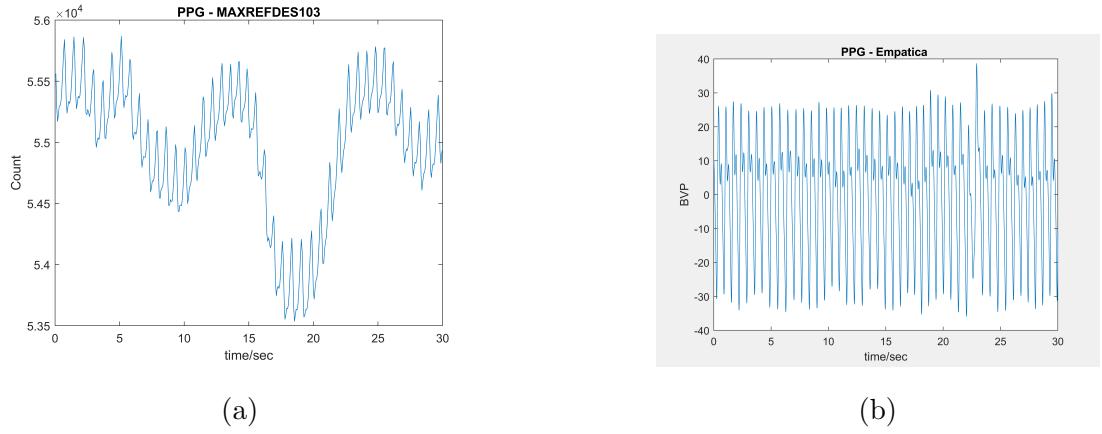


FIGURE 4.6: Plot of PPG signal obtained under a 30 second window from the same individual with (a) MAXREFDES103 and (b) E4 .

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

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

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 4.6, 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 2<sup>nd</sup> 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 4.7.

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 [26].

Figure 4.8 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.

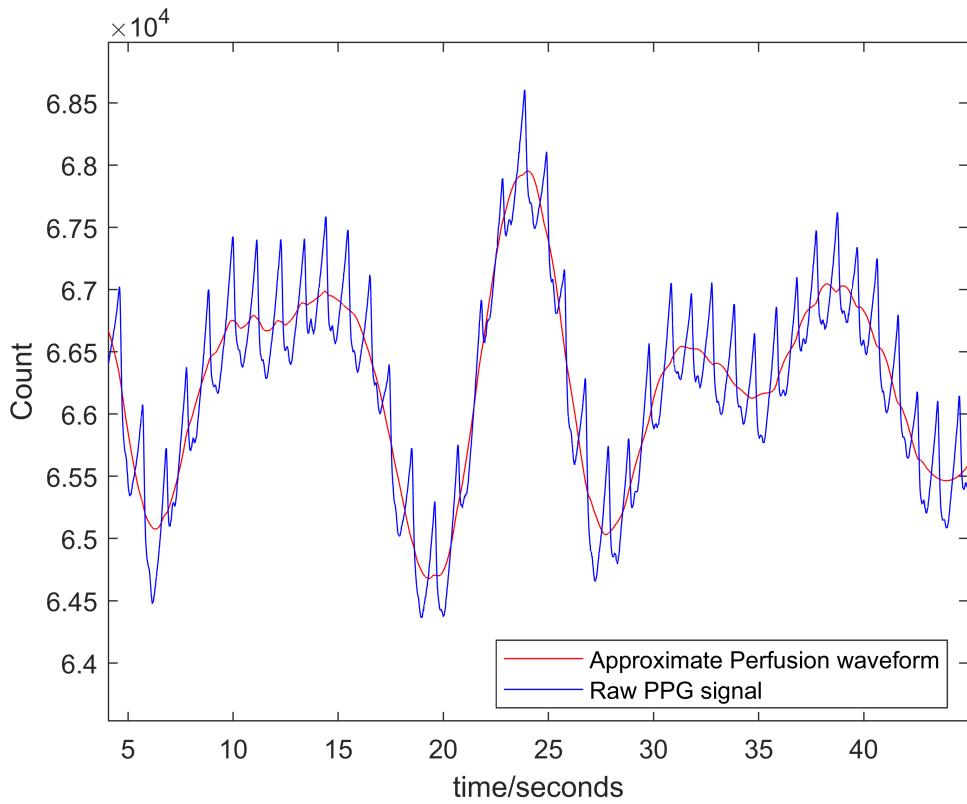


FIGURE 4.7: Raw PPG Signal (Blue) along with the approximated Perfusion Waveform (Red)

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.

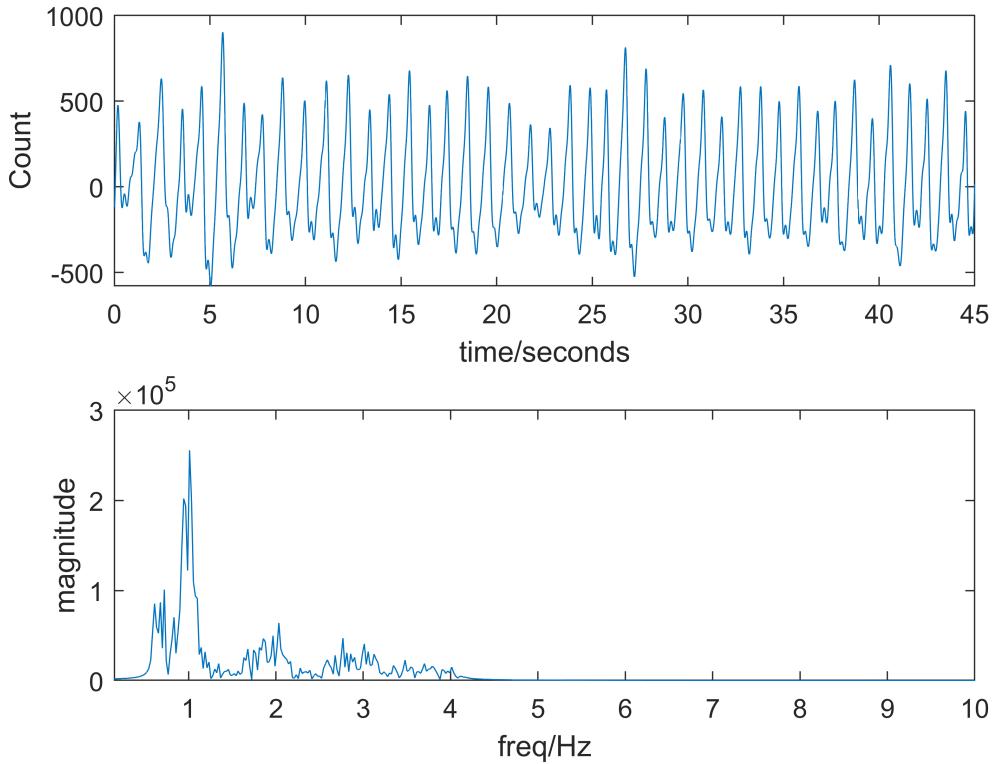


FIGURE 4.8: PPG Waveform obtained along with its Frequency Plot

### 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 4.9 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  $\leq 0.5$  Hz. The peaks at approximately 1 Hz and 2 Hz correspond to the rhythmic motion of the heart.

Firstly, by subtracting the median of the data set from all sample points, we are able to remove the DC component from our raw PPG signal. 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 2<sup>nd</sup> order polynomial at a frame length of 257 samples. The results

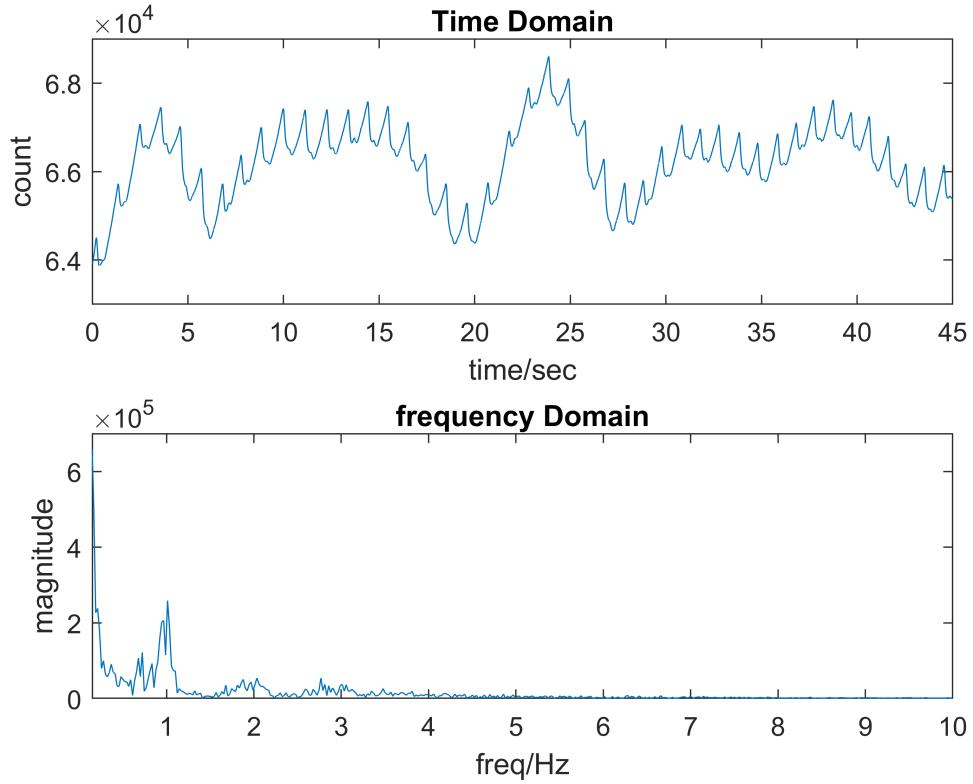


FIGURE 4.9: Raw PPG Signal and its Frequency Plot from MAXREFDES103

are shown in figure 4.10

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 MAXREFDES103 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.

## 4.4 Motion Artifacts Removal

Being aware of the limitations of computer processing power in wrist worn devices with PPG sensors, it might not be possible to fully recover the PPG signal

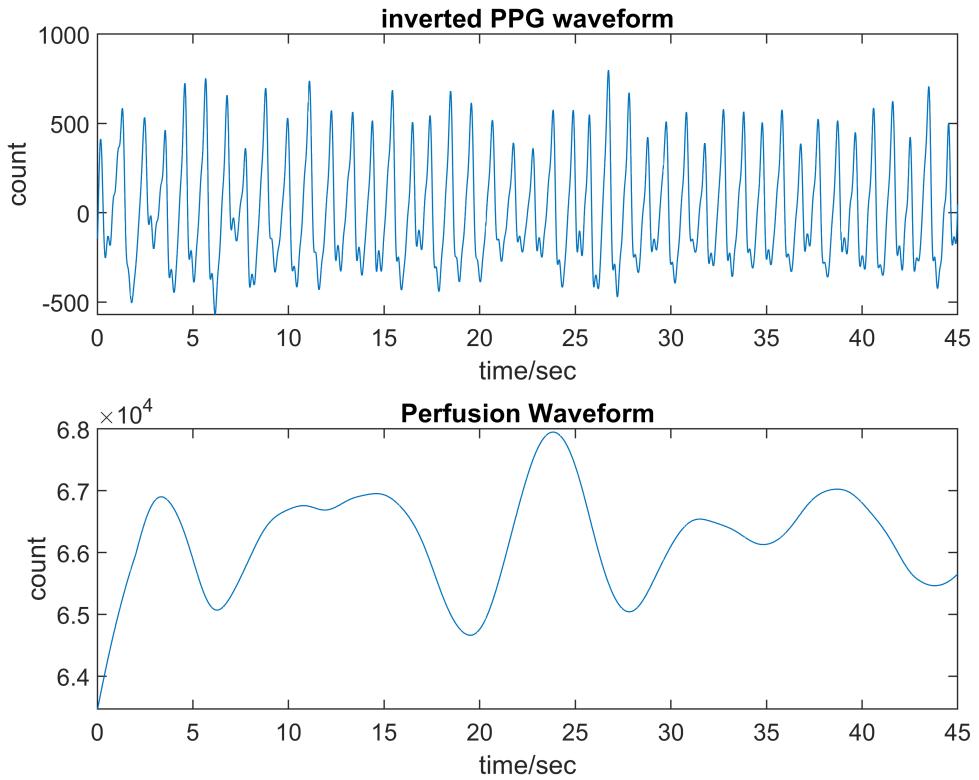


FIGURE 4.10: PPG Waveform and Perfusion Waveform obtained by filtering using a bandpass filter

efficiently over live simulation. This problem motivated us to create a general framework for the detection and removal of motion artifacts when required hence ensuring that minimum CPU power was being utilized leaving our algorithm more computationally efficient. Fig 4.11 summarizes our approach.

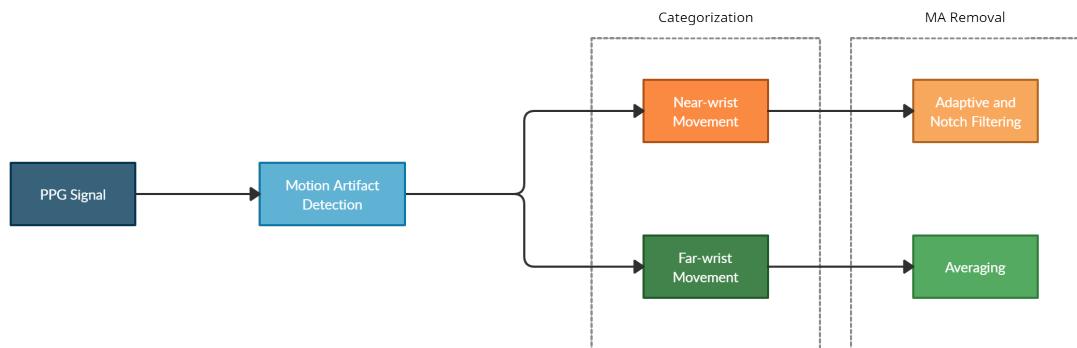


FIGURE 4.11: Motion Artifact removal implementation

#### 4.4.1 Detection of Motion Artifacts

Our initial step towards the removal of motion artifacts requires an accurate detection of time interval where motion artifacts present themselves. Although, we could use the accelerometer data to mark time instances where motion was induced, we realized that accelerometer data might not be readily available for every device. Moreover, for previous mentioned reasons in terms of computer efficiency and accelerometers correlation with motion artifacts in the frequency domain, we aimed at reducing our algorithms dependency on the accelerometer data as much as possible. For this purpose, we used two statistical variables namely Kurtosis and Skewness which formed a vital backbone in the detection of motion artifacts in the PPG signal.

- Skewness

A statistical measure of the distributional asymmetry of a probability distribution and according to [27], it was found that there was an association between skewness and corrupted PPG signals. Skewness can be expressed as:

$$Skewness = \frac{1}{N} \sum_{i=1}^N \left[ x_i - \frac{\mu}{\sigma} \right]^3 \quad (4.5)$$

where  $\mu$  is defined as the mean and  $\sigma$  is defined as the standard deviation of the PPG signal  $x$ , respectively. Moreover, the number of samples in the PPG signal is indicated by  $N$ .

- Kurtosis

A statistical measure for the degree of peakedness of a distribution. In comparison to a normal distribution, it defines the distribution of observed data around the mean [28]. According to [27], this measure captures the random variation of data from the mean and is defined as:

$$Kurtosis = \frac{1}{N} \sum_{i=1}^N \left[ x_i - \frac{\mu}{\sigma} \right]^4 \quad (4.6)$$

where  $\mu$  is defined as the mean and  $\sigma$  is defined as the standard deviation of the PPG signal  $x$ , respectively. Moreover, the number of samples in the PPG signal is indicated by  $N$ .

The idea is to create a distribution of Skewness and Kurtosis of a PPG signal over the period and analyze any deviation that originate from the normal that might

hint towards the presence of motion artifact. In order to do so, the initial step for our framework is to record a reference PPG signal for the user at complete rest position ensuring that the PPG signal contains no motion-induced noise for approximately 30 seconds. This initial step can be regarded as calibration for a device once the user puts it on for the very first time. Next, we extract the features from this reference PPG signal using the method of overlapping window. We iterate over the signal in windows such that the  $i^{th}$  window overlaps with the  $(i+1)^{th}$  window. Figure 4.12 shows this approach in a visualization.

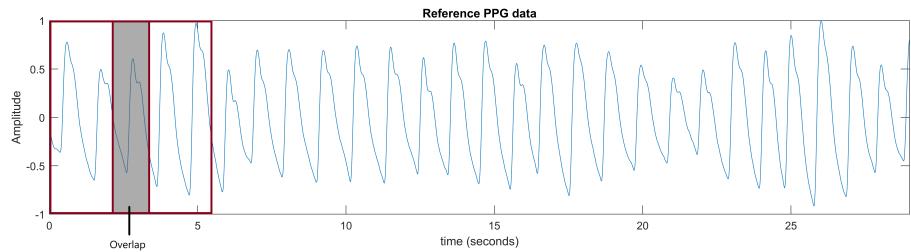


FIGURE 4.12: Iteration through the PPG signal using overlapping windows.  
The shaded area shows the percentage of overlap

The PPG data in each window is then used to compute the corresponding Kurtosis, Skewness and Initial time interval which form part of our feature row for each window. The feature row can be summarized as follows:

$$[Kurtosis_i \quad Skewness_i \quad Time_i]$$

These feature rows were computed for each window as we iterated through the reference PPG signal and were appended into our feature matrix. The final feature matrix that we obtained was of dimension  $M \times 3$  where  $M$  denotes the number of windows and 3 corresponds to our features. A more detailed visualization of our feature matrix for a reference PPG signal is as follows:

$$\begin{bmatrix} Kurtosis_1 & Skewness_1 & Time_1 \\ \vdots & \vdots & \vdots \\ Kurtosis_M & Skewness_M & Time_M \end{bmatrix}$$

Through experimentation, we found that the optimal values for our feature extraction using overlapping windows were as follows:

- Window size= 48 samples
- Overlapping percentage= 90

Once the reference feature matrix was obtained, we obtained the reference bounds that contain approximately 95% of the values of Kurtosis and Skewness obtained. The bounds were obtained using the equation:

$$Bound = \mu_{feature} \pm 2\sigma_{feature} \quad (4.7)$$

where feature is either kurtosis or skewness,  $\mu_{feature}$  represents the feature's mean and  $\sigma_{feature}$  is the feature's standard deviation.

The calculation of these bounds is vital in the detection of time intervals where motion artifacts were induced as we will see in the succeeding sub chapters. Fig 4.13 shows a plot of the Kurtosis and Skewness spread of a healthy individual at rest over the time of 25 seconds.

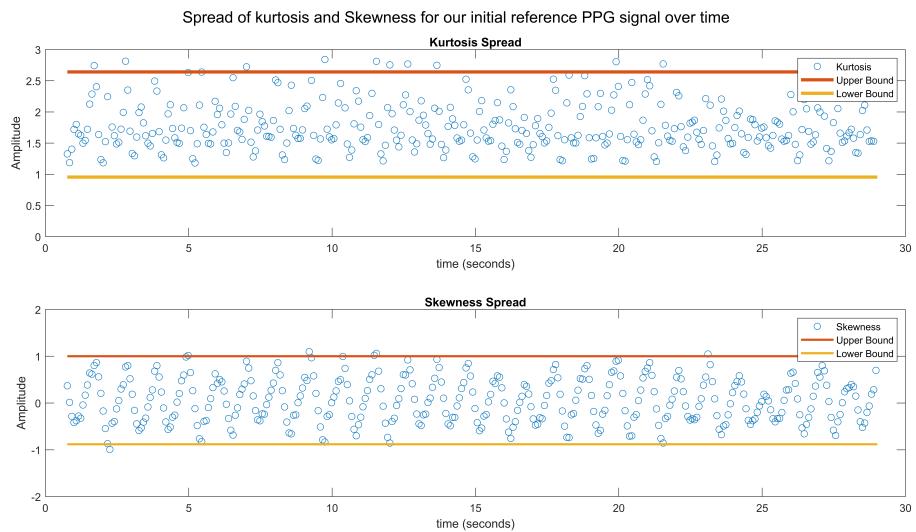
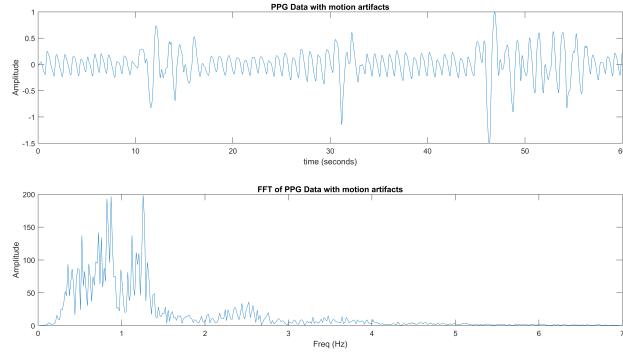


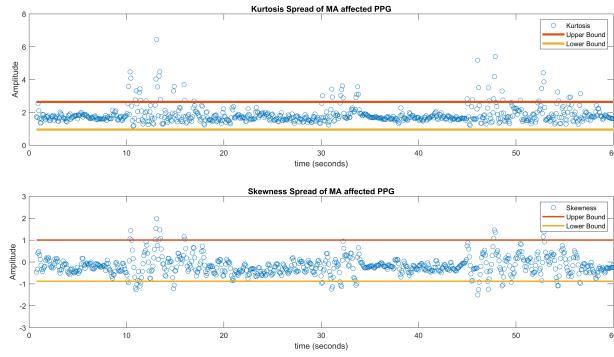
FIGURE 4.13: Spread of Kurtosis and Skewness for a healthy subject at rest over a period of 25 seconds

Once the calibration is done, now we need to obtain the features of our corrupted PPG signal using the same method and parameters as described previously for the same subject. Once the feature matrix is obtained, the previously found reference bounds are plotted with the new feature matrix. Fig 4.14 shows the PPG signal containing motion artifacts and the features obtained along with the reference bounds for a healthy individual. A more diverse spread for both Kurtosis and Skewness can be observed which corresponds to the additional noise that has been introduced in the PPG signal as part of the movement.

Next step involves the detection of time windows where motion artifacts affect the signal. Since each feature i.e., Kurtosis or Skewness is of size  $M$ , where  $M$



(a) PPG signal with motion artifacts and its corresponding frequency spectrum



(b) Kurtosis and Skewness for the PPG signal in (a)

FIGURE 4.14

denotes the number of windows that iterated over the signal, we compare the  $M$  values for each feature with their corresponding reference bound and mark each instance if  $Kurtosis_i$  and  $Skewness_i$  both fall outside the reference bound interval. The algorithm iterates over the window of 2 seconds and concatenates those windows which has marked instances of kurtosis and Skewness. This means that the window size that we obtain will be a multiple of 2 seconds. The window array that is computed is visualized as follows:

$$[Window_{init} \quad Window_{fin}]$$

Where the first column marks the initial time where motion artifacts are detected for the first window and the 2nd column marks the final time instance. Similarly, many of such time windows are detected and appended into our final detected windows matrix which is visualized as below:

Where  $k$  denotes the number of windows detected where motion artifacts were present. Fig4.15 shows the marked intervals where our algorithm detected motion artifacts.

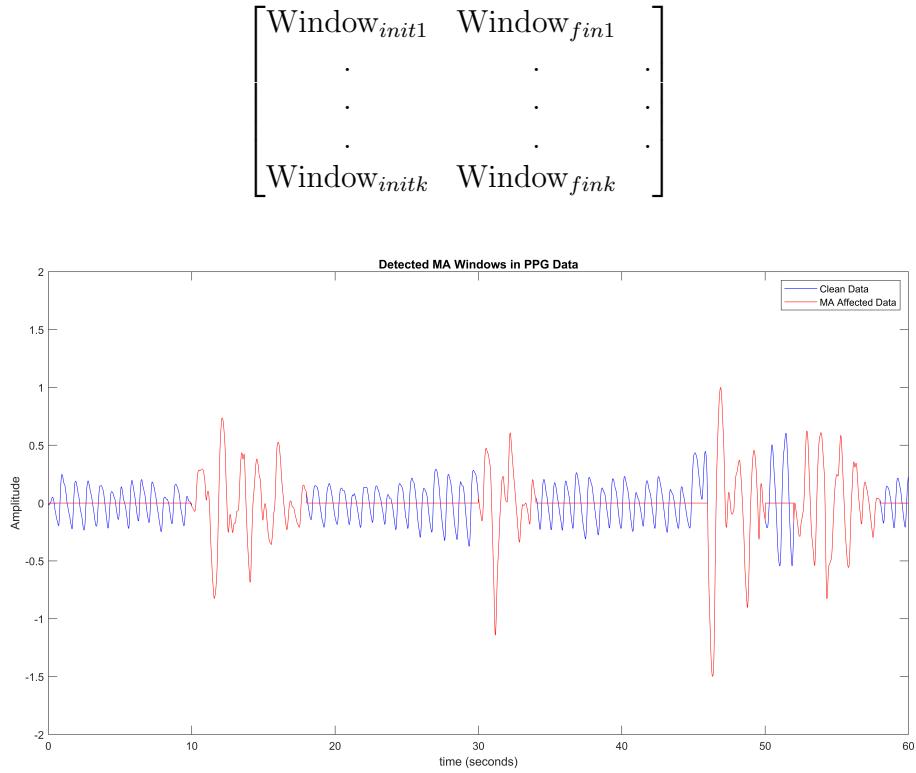


FIGURE 4.15: PPG signal with motion artifacts windows detected

Once the windows have been detected, we classify the MA corrupted PPG signal in these windows into far wrist and near wrist motion as per the window length. If the window length is less than or equal to 4 seconds, we classify it as far wrist action and near wrist otherwise. This is because far wrist i.e., the motion induced by the movement of body parts far from the wrist where the sensor lies e.g., shoulders, elbows etc. Such motion does not strongly affect the blood flow in the wrist as much hence leading to shorter time interval where such motions induce noise. Near wrist motion i.e., originates from the muscle movement near the wrist such as typing, writing, forearm movement etc. tends to deteriorate our signal at a much longer time interval with greater intensity. As a result, near wrist action induced motion artifacts will require more signal processing hence more computational power whereas the far wrist action induced motion artifacts will require less processing and computational power to achieve a clean signal. This classification corresponds to our aim of making an efficient framework that utilizes its computational power as per its requirements.

#### 4.4.2 Removal of Motion Artifacts

Once the windows corresponding to PPG signal affected by motion artifacts has been detected, we clean the signal based on two classifications of the respective

PPG segment obtained from the given window namely far wrist and near wrist.

#### 4.4.2.1 Far Wrist

As mentioned previously, far wrist action motion artifacts do not corrupt our PPG signal with such intensity. Therefore, a low computational method is more preferred for such type of motion. The approach that we adopted first generates a synthetic PPG segment by averaging the shape of a PPG waveform using previously available data and then replacing the corrupted PPG segment with this synthetically generated PPG segment. When it comes to generating a synthetic PPG waveform to replace the corrupted data at time  $t$ , we collect previously available data of  $t-5$  till  $t$  seconds if available otherwise a random sample of PPG waveform of  $t=5$  seconds from the reference data used for calibration is taken. Each individual PPG waveform from the obtained previous data is then separated using the double derivative. The double derivative aids us in finding the local minima and their corresponding index which can then be used separate the individual PPG waveforms. These individual PPG waveforms are then averaged to form an averaged shape of a PPG waveform of the subject. Fig 4.16 shows the averaged PPG signal obtained for the cleansing of PPG signal at the time interval of 30 to 34 seconds as shown in Fig 4.15.

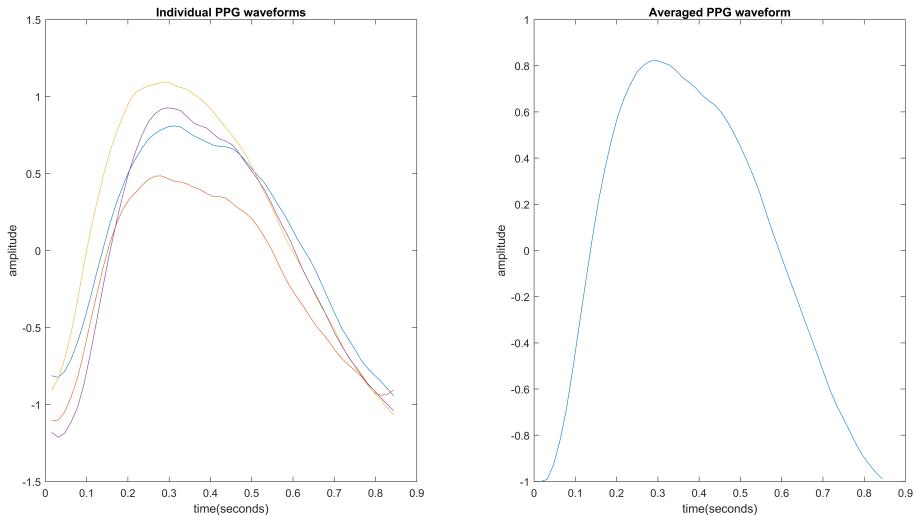


FIGURE 4.16: Generation of PPG waveform by averaging

This averaged shape of the PPG waveform is then concatenated together until the length of the signal reaches the length of the window whose corrupted data is to be replaced. Post processing is done to remove any Dc component or trend that may have been introduced in the synthetic signal as part of the concatenation process. Fig 4.17 shows the final synthetically generated PPG waveform that is to replace

the corrupted PPG waveform for the signal shown in Fig 4.15 in between  $t=30$  and  $t=34$  seconds.

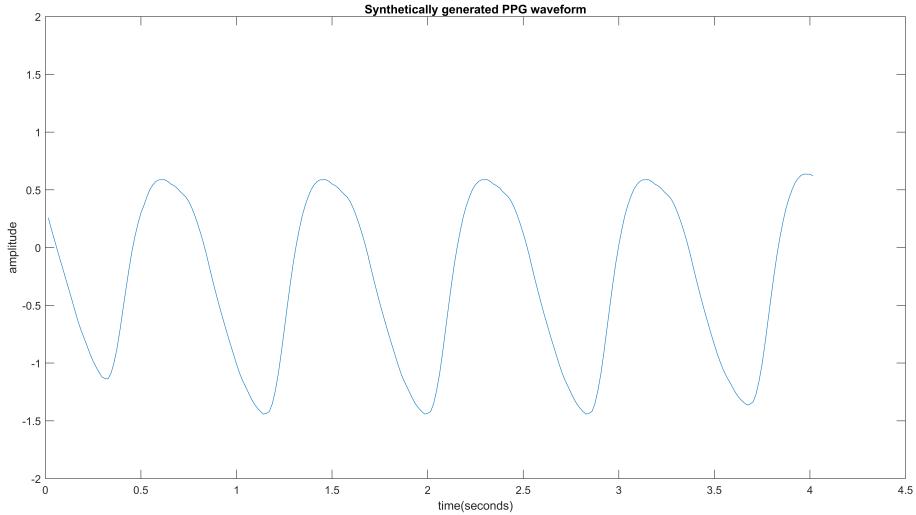


FIGURE 4.17: Synthetically generated PPG waveform

#### 4.4.2.2 Near Wrist

The near wrist motion more strongly affects the PPG signal and for greater time interval hence replacing the corrupted signal with the synthetic signal is not the most feasible option. Therefore, in order to remove near wrist motion artifacts, a combination of RLS adaptive filters and notch filters are implemented.

First, the raw PPG signal passes through the RLS adaptive filtering system. The RLS filter is a type of adaptive filter that recursively minimizes the least square cost function weights by reducing the mean square error between the desired and input signals. The cost function of the RLS adaptive filter is as follows:

$$C(w_N) = \sum_{i=0}^N \lambda^{N-i} e^2(i) \quad (4.8)$$

where,  $e(i)$  is the error between the desired signal and estimated desired signal,  $\lambda$  is the forgetting factor,  $N$  is the window length and  $w_N$  represents the filter weights. For the given input signal  $s(n)$ , which is the linear combination of the desired signal  $d(n)$  and the noisy signal  $N(n)$ ,  $s(n) = d(n) + N(n)$ , an adaptive filter automatically updates the weights of the filter until the error between the estimated desired signal  $d_{est}(n)$  and the reference signal, which in our case is the desired signal  $d(n)$ , converges. The RLS algorithm is chosen as it provides fast convergence.

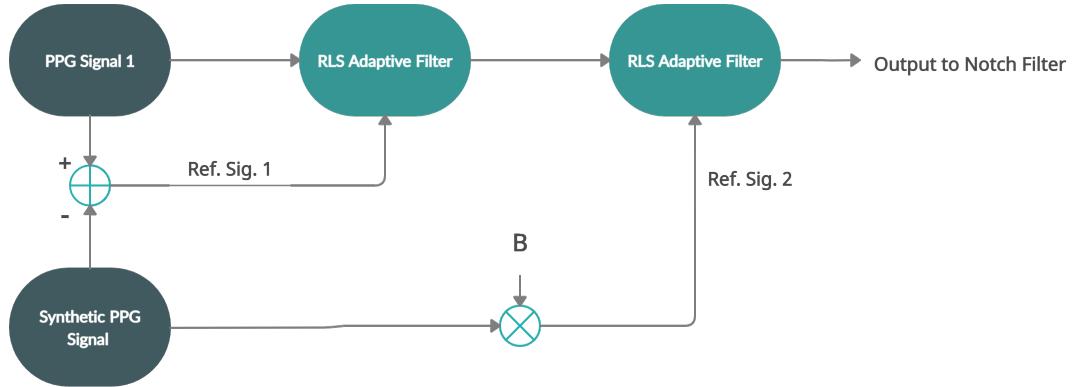


FIGURE 4.18: Adaptive Filter

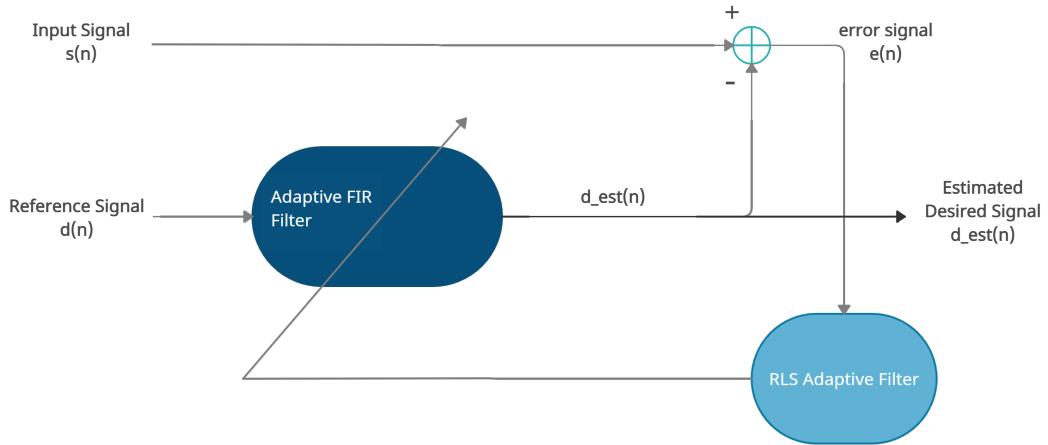


FIGURE 4.19: RLS Adaptive Filter

Our approach is influenced by the approach used in [13]. As can be observed from figure 4.19, a raw PPG signal and a synthetic PPG signal based on the average of the previous 5 seconds of the same PPG signal, serve as the input signals. The synthetic signal is then subtracted from the raw PPG data to remove tissue effect induced motion artifacts. The difference between the two signals serves as the reference signal  $d_n(n)$  for the first pair of RLS adaptive filters, enhancing the raw PPG signal. Next the output of one of the filters is multiplied by  $\beta$  and then subtracted from the other filter's output. This weight subtraction is used to remove any motion artifact influenced venous blood movement. Based on the approach of [29], the optimal values for  $\beta$  were determined to be  $\beta = 1.05$  for  $X2 < X1$  and  $\beta = 0.95$  for  $X1 < X2$ . The difference from the weighted subtraction is then used as a reference signal for the second pair of RLS adaptive filters, which then further enhance the initial raw PPG signals. This output is then passed on to the notch filter for further processing.

Due to the quasi-periodic nature of the PPG signal, the notch filter can be used to remove any unwanted frequencies from the PPG signal using the estimated heart rate frequency (HRF) along with its second harmonic frequency [30]. Equation (4.9) represents a digital notch filter's transfer function [30]:

$$H(z) = \frac{(r^2 - (1 + r^2)\cos(w_n)z^{-1} + z^{-2})}{(1 - (1 + r^2)\cos(w_n)z^{-1} + r^2z^{-2})} \quad (4.9)$$

where,

$$w_n = \frac{2\pi f_n}{F_s} \quad (4.10)$$

$f_n$  is the notch filtering frequency and  $F_s$  is the sampling rate and  $r$  is the parameter to control the bandwidth. Based on the approach used in [30], first the HRF is determined from the output of the RLS adaptive filtering. The HRF is calculated by finding the average difference in peaks within a minimum separation of 0.7 seconds. Once the HRF is found, the notch filter removes the frequencies around the central HRF,  $f_{\text{HRF}}$  at a bandwidth of 0.4Hz, leaving behind the noisy frequencies. The signal then passes through the same procedure again with another notch filter centering on the second harmonic of HRF, i.e.  $2f_{\text{HRF}}$  due to the quasi-periodicity of the PPG signal, thereby removing the PPG signal and leaving behind the unwanted noise from any remaining motion artifacts. The output from both these filters is then subtracted from the input of the first notch filter, thus, removing any motion artifacts that remain in the PPG signal. The figure 4.20 shows the step-by-step process of the notch filter.

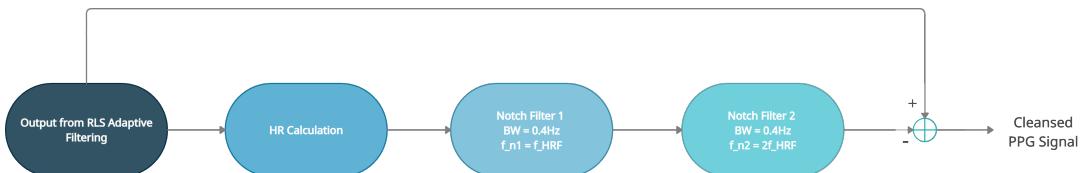


FIGURE 4.20: Notch Filter

The output signal was then normalized as we iterated through each window of detected motion artifacts. The signal we obtain is the cleansed PPG signal. The above methodology was applied to the signal shown in fig, where 4 windows were detected two of which were far wrist and 2 near wrist. The resultant PPG waveform is shown in fig 4.21.

Fig 4.22 shows the spread of Kurtosis and Skewness both before and after the implementation of our framework on the PPG signal in fig 4.14. It can be clearly

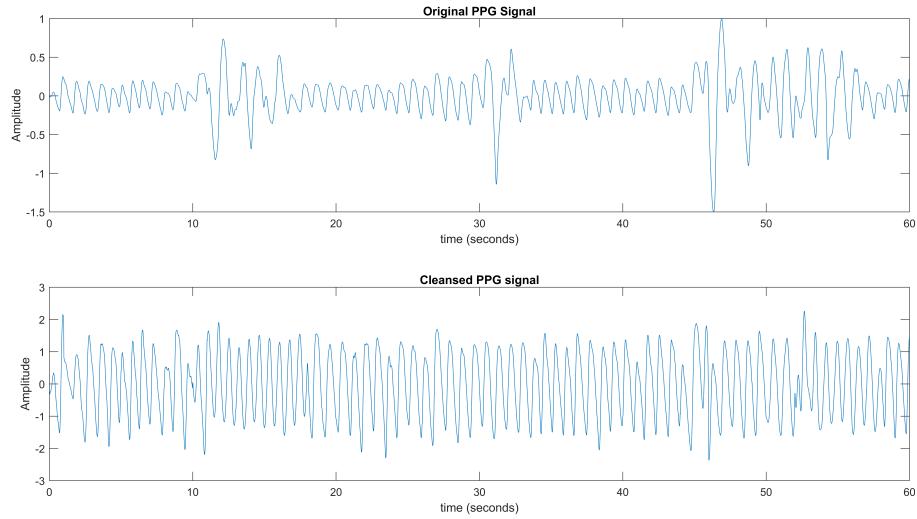


FIGURE 4.21: PPG signal before vs after cleaning

observed that the spread of Kurtosis and Skewness around the time intervals where motion artifacts were detected reduced.

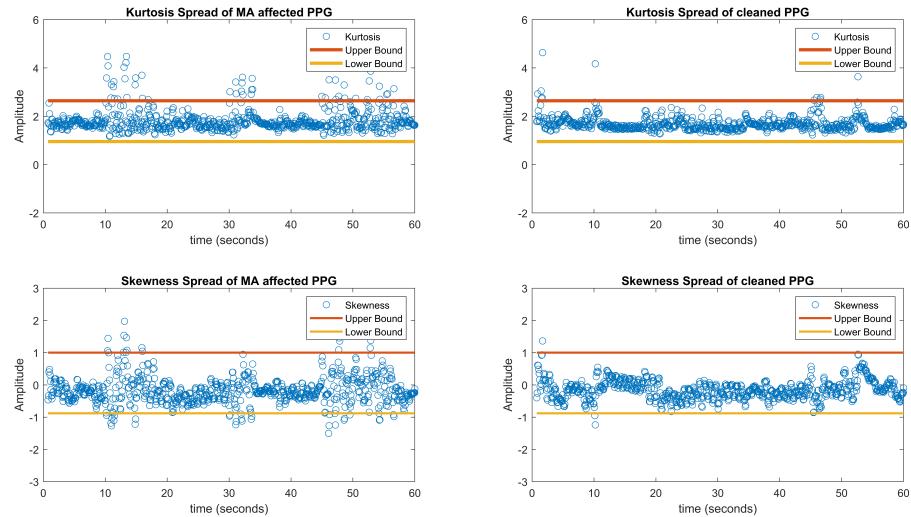


FIGURE 4.22: Spread of kurtosis and Skewness before vs after

## 4.5 Experimentation and Results

### 4.5.1 PPG Signal Cleansing

Experimentation and Results: We validated our framework on the datasets taken during the second phase of data collection. The PPG signal data of MAXREFDES103 was used for each subject. The following type of motion affected PPG signals were used:

- **Excercise 1** : Walking
- **Excercise 2** : Moving the dominant arm up and down within a controlled pace (Slight arm movement)
- **Excercise 3** : Typing with the dominant hand
- **Excercise 4** : Writing with the dominant hand

As mentioned previously, our framework requires an initial reference PPG signal with no motion artifacts for calibration. For this purpose, we had recorded PPG signal for each subject in rest position. The framework was run on 8 subjects each with the above 4 movements and the output PPG signal was recorded. The input PPG signal (corrupted signal) and the output PPG signal (cleansed signal) were both utilized in the extraction of features namely kurtosis and skewness for each subject for each exercise. Hence for each exercise, a total of 8 feature matrices was obtained for each corresponding input PPG signal and 8 feature matrices for each corresponding output PPG signal from each subject. Next, for each exercise, kurtosis for each of the 8 subjects were concatenated together to form a 1d matrix. Similar processing was done for Skewness as well for each exercise. As a result, our final outputs were 4 Kurtosis and 4 Skewness 1d matrices for 4 exercises mentioned above for each before and after the implementation of the framework. The distribution of these matrices is shown in fig 4.23 for comparison.

To summarize, the spread of both kurtosis and skewness distribution decreased for each exercise after the implementation of our framework. This indicates that we were able to remove the variation that would otherwise point to the presence of motion artifacts in the PPG signal. The difference in spread of Kurtosis can be more prominently observed in fig 4.23.

#### 4.5.1.1 Performance Index

To express how well did our framework perform on the signal, we created a performance index (PI) which compares the Upper bound of both kurtosis and skewness for before and after the application of framework with the reference bound calculated during the calibration stage. The performance index can be expressed as:

$$PI_{ijk} = \frac{UP_{ijk} - UP_{reference_i}}{UP_{reference_i}} \quad (4.11)$$

Where:

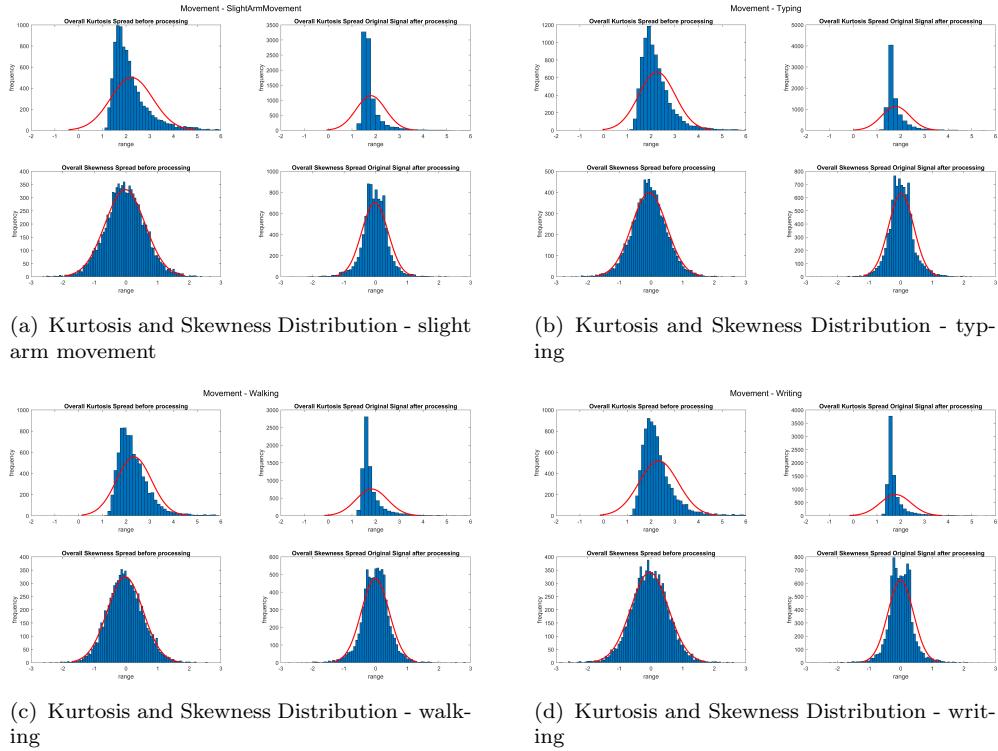


FIGURE 4.23

- $i$  = Subject number
- $j$  = Exercise number
- $k$  = Before (b) or after (a) cleansing the signal using the described framework
- $UP_{ijk} = \mu + 2 \sigma$  of our corrupted signal Kurtosis/Skewness Spread.
- $UP_{reference_i} = \mu + 2 \sigma$  for our reference data Kurtosis / Skewness spread.

PI is hence the percentage difference between the upper bound of a sample PPG signal and the corresponding reference PPG signal. This PI is then calculated for each subject for each exercise mentioned above both before and after the application of the framework for both Kurtosis and Skewness spread. Ideally, the PI for after the application of framework must be smaller and less than 1 as compared to the PI before for each exercise. Table 4.2 and Table 4.3 summarizes the PI obtained for all the data used for experimentation purpose.

Exercise	$1_b$	$1_a$	$2_b$	$2_a$	$3_b$	$3_a$	$4_b$	$4_a$
Subject 1	0.62	0.3	1.02	0.55	0.81	0.57	0.96	0.3
Subject 2	0.62	0.4	0.6	0.31	0.49	0.17	0.57	0.38

Subject 3	0	0	1.09	0.31	0.87	0.56	1.44	0.46
Subject 4	0.46	0.44	0.84	0.45	0.86	0.33	0.74	0.24
Subject 5	0.85	1.07	1.32	0.93	0.8	0.74	1.13	1.18
Subject 6	0.57	0.48	0.84	0.27	1.06	0.28	0.95	0.33
Subject 7	0.89	0.13	0.51	0.26	0.45	0.23	0.45	0.23
Subject 8	1.27	0.52	1.15	0.23	0.99	0.49	0.95	0.46

TABLE 4.2: PI matrix for Kurtosis. Syntax  $j_b$  refers to  $j^{th}$  exercise PI before cleansing and Syntax  $j_a$  refers to  $j^{th}$  exercise PI after cleansing

Exercise	$1_b$	$1_a$	$2_b$	$2_a$	$3_b$	$3_a$	$4_b$	$4_a$
Subject 1	1.32	1.03	2.10	1.26	1.63	1.58	1.77	0.86
Subject 2	2.10	1.44	1.72	1.08	1.24	0.80	2.07	1.16
Subject 3	0.00	0.00	5.40	2.21	4.83	2.15	4.49	2.09
Subject 4	0.17	0.36	1.09	0.27	1.14	0.25	0.96	0.09
Subject 5	8.25	9.33	11.97	8.76	8.32	11.10	10.89	11.79
Subject 6	6.47	5.72	9.43	4.74	6.89	4.02	8.75	5.24
Subject 7	1.04	0.37	0.64	0.24	0.42	0.20	0.42	0.20
Subject 8	8.87	2.95	8.01	2.95	6.49	3.22	6.24	2.68

TABLE 4.3: PI matrix for Skewness. Syntax  $j_b$  refers to  $j^{th}$  exercise PI before cleansing and Syntax  $j_a$  refers to  $j^{th}$  exercise PI after cleansing

### 4.5.2 Parameter Estimation

#### 4.5.2.1 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 [31]. 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 4.24 shows the respiratory signal of a healthy individual over a period of 70 seconds at rest position.

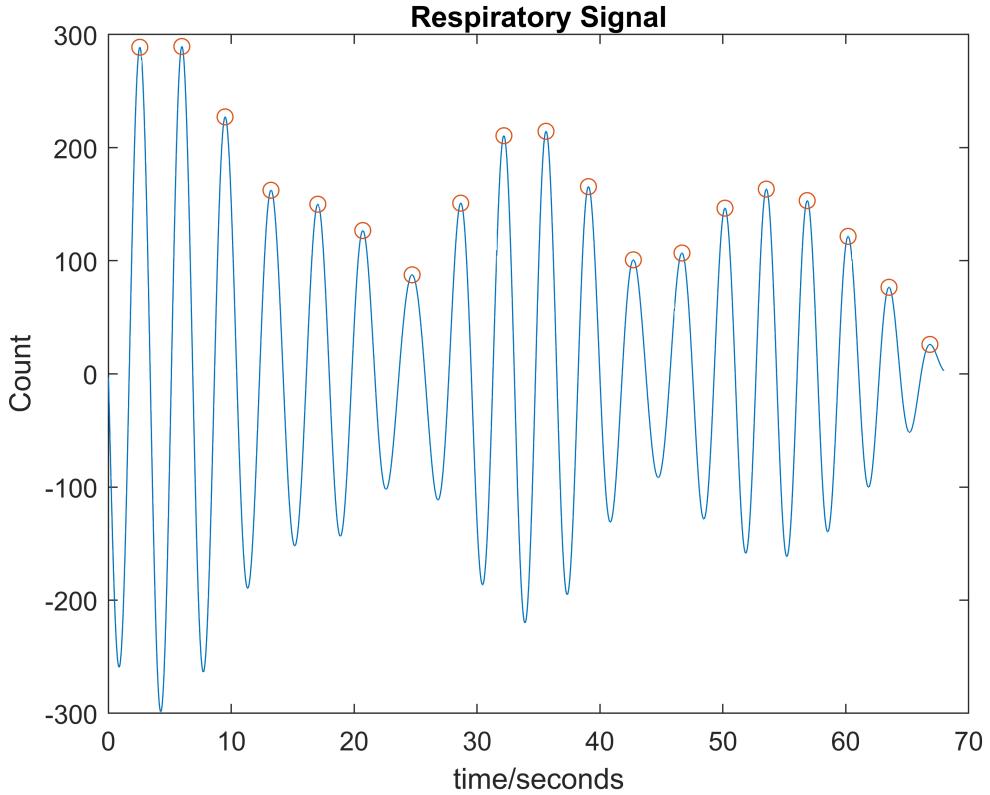


FIGURE 4.24: Respiratory Signal at 0.3 Hz with Local Maxima obtained

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

$$Breath\ rate\ per\ second = \frac{sample\ rate}{mean\ breath\ rate\ in\ samples} \quad (4.12)$$

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  $\pm 1$  breath per minute.

## 4.6 Graphical User Interface (GUI)

A Graphical interface was developed using MATLAB GUI for users to conveniently interact with our developed algorithms without much hassle. User Interface sets a platform of the entire lengthy and tedious process we went through under a single tab making it user friendly. In this section of the report we will be having a complete overview of our GUI. Along with a step-wise guide to remove motion artifacts from any PPG signal using this Graphical User interface only, we will also look at breath rate estimation in this GUI. At the back-end of this User interface, we are using the same aforementioned processes to first separate the PPG and perfusion signals from each other and then to remove the motion artifacts from the PPG signal along with breath rate estimation from the perfusion signal.

The welcome window of the Graphical User interface shown in Figure 4.25 is the introduction of our project and the team members. The "Start" push button takes us to the next window.

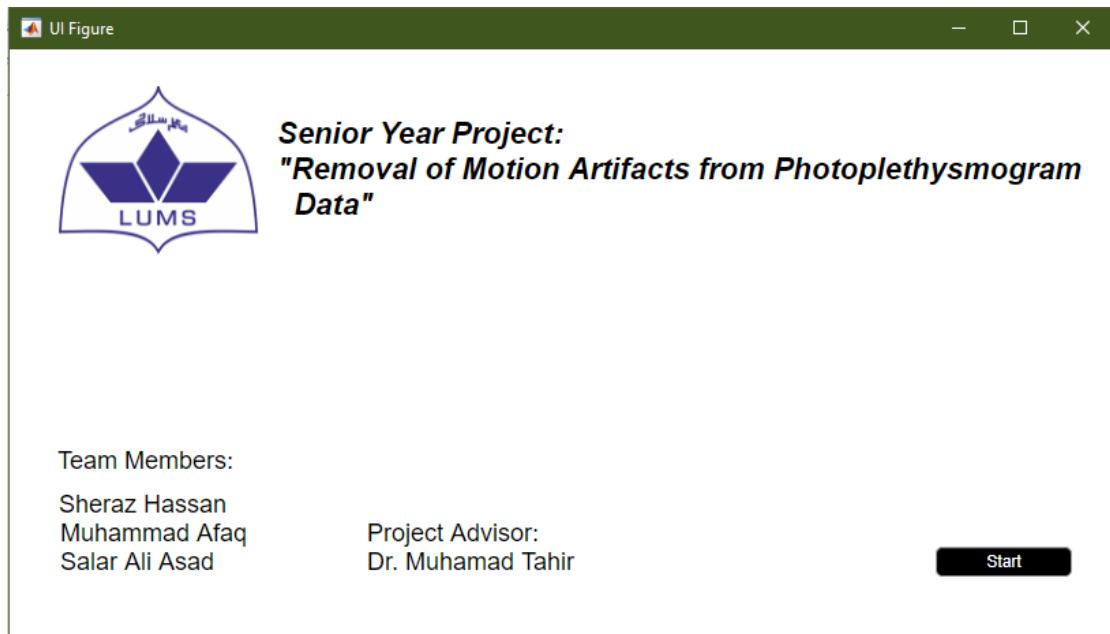


FIGURE 4.25: Start window of GUI

Welcome window is followed by the window shown in Figure 4.26 and its the tab where the actual process begins. Here the user will be asked to enter the file name (which the user wants to process) and press the upload button (the file should be in the same folder as of GUI). Once the Signal is visible on the ‘Raw PPG’ plot, the user can use the slider to opt the time period he/she wants to work with and press the “Start processin” push button which is labeled as (1) in 4.26. Once done only the signal from selected time period will be displayed on the plots below and only this same data will be carried to next step of processing. Now, after the separation of PPG and Perfusion, the user can choose to either process the perfusion and get breath rate estimation by clicking on the push button labeled as (3) or process the PPG signal to remove the motion artifacts by clicking on the push button labeled as (2).If either of these push buttons are pressed without uploading or processing the data, the GUI will send an error back to the user.

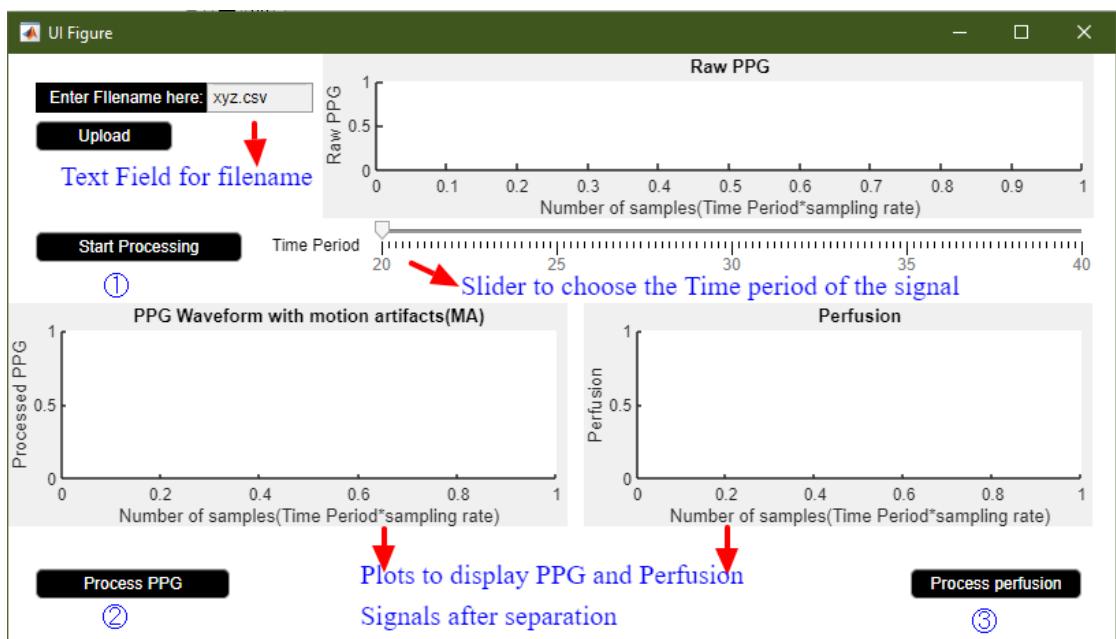


FIGURE 4.26: Upload window of GUI

Push button (2) that is “Process PPG” takes the user to the window where the motion artifacts are removed and displayed. The first plot displays the clean PPG signal and the detected motion artifacts in blue and red respectively,followed by the second plot which is the FFT graph of this signal. The next plot displays the PPG signal which has been processed and cleansed of all the motion artifacts. The blue part is the original signal whereas the red one is the replaced signal. This plot of cleansed PPG is also followed by its FFT graph. There are push Buttons

at the end of the window labeled as (4) in the figure 4.27. Both PPG signal with motion artifacts and cleansed PPG signal has its own push button. Once pushed, graphs of the Skewness and Kurtosis spread of the signals pop up. As discussed in the previous section, the adaptive filter requires a reference signal to reduce error and remove motion artifacts. Hence for this window to operate, the user will be required to add a reference file named as “Reference.csv” in the GUI folder.

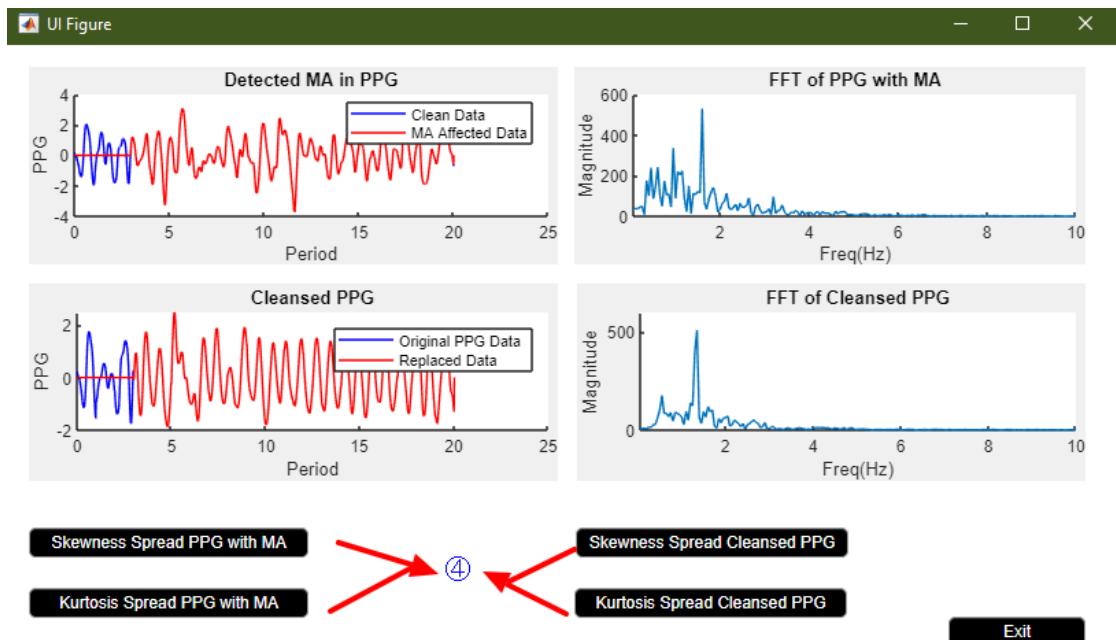


FIGURE 4.27: PPG window in GUI

If the user selects Push button (3) in the second window 4.26, He/she will end up in the parameter estimation window of the GUI. In this window, shown in figure 4.28, the user can find the breath rate that is estimated through the perfusion signal. At the top of the window, the plot shows the perfusion graph which is the same from the previous window as selected through the slider. That plot is followed by another plot which shows the Respiratory signal at 0.3Hz through which the estimation is carried out and displayed automatically in the text field below. Pressing the exit button will bring the user back to the second window where the slider can be changed or a new file can be uploaded to work with it.

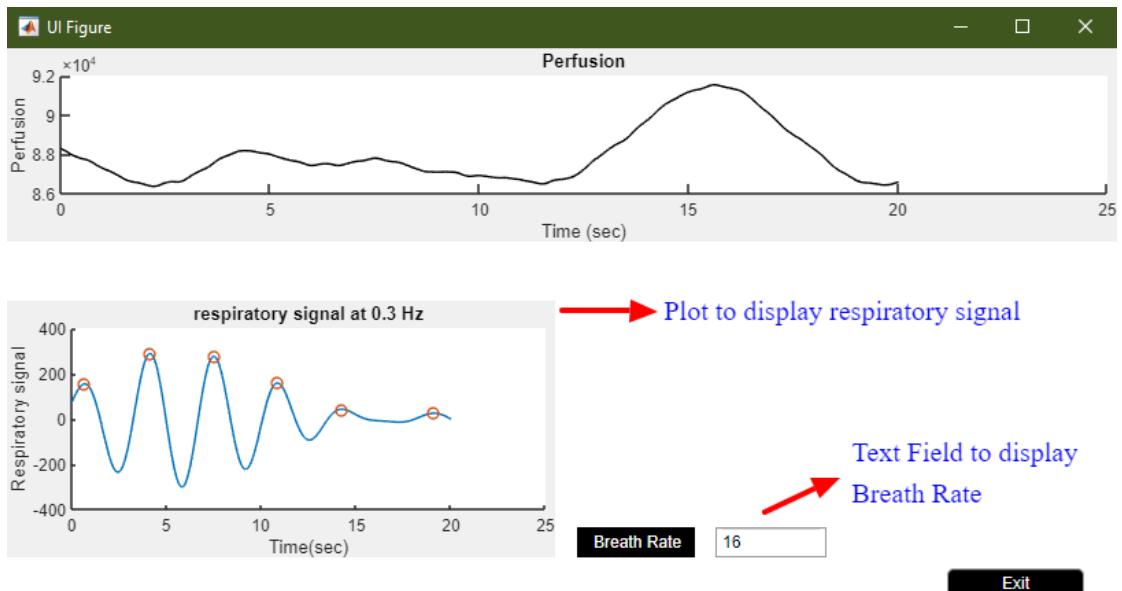


FIGURE 4.28: Perfusion window in GUI

# Chapter 5

## Cost Analysis, Future Work, and Conclusion

### 5.1 Result and Cost Analysis

In SPROJ, we have separated the perfusion waveform and the PPG waveform from the raw PPG signal using polynomial fitting and filtering whilst comparing the PPG data collected from high-end and low-end PPG devices. We also established a relation between the PPG data and its corresponding accelerometer data under the influence of motion artifacts by analyzing the frequencies of motion. Although filtering out these frequencies of motion does clean our data to some extent, it does not entirely remove the noise from the PPG waveform.

In SPROJ II, we have furthered our progress by developing an approach for detecting, classifying and finally removing motion artifacts. We start with identifying the areas containing motion artifacts and then categorize them as either far-wrist or near-wrist movement. For far-wrist motion, the PPG signal is processed using the averaging method. For near-wrist motion, the PPG signal is processed by first passing the signal through an RLS adaptive filtering system and then through two cascading notch filters where unwanted frequencies are subtracted from the PPG signal.

Regarding the cost of this project, it was relatively inexpensive. While the Empatica E4 was used as a reference to validate our results, our work primarily utilized the MAXREFDES103, which is significantly cheaper than the Empatica E4, costing approximately \$220 as compared to the approximately \$1200 Empatica E4. This helps fulfill our aim to develop an approach that was compatible with more inexpensive devices, making the utilization of this approach more affordable yet

at the same time practical.

### 5.1.1 Motion Artifact Removal

As highlighted before, this process takes place in three stages: detection, classification and removal. This approach is practical on PPG signals with low to moderate levels of motion artifacts. However, it struggles when motion artifacts are severe, particularly when the entire signal contains motion artifacts instead of just specific portions of the signal. These limitations are present due to the assumption that the synthetic reference signal is the average of the previous 5 seconds before the motion artifact. As a result, motion artifacts that are severe and affect the entire signal's length perform poorly with our approach as the synthetic reference itself contains motion artifacts. However, while this limitation exists, our approach is novel compared to other existing methods because it aimed to optimize efficiency and reduce computational complexity. By categorizing motion artifacts based on their severity, those motion artifacts that are less intense will be processed through a much simpler algorithm with reliable accuracy, thereby saving cost and time. More intense motion artifacts will be processed through a much more complex algorithm that, although will take more time, is more robust and accurate. Our technique, unlike existing methods which use only one approach to remove all motion artifacts by distinguishing the nature of the motion artifacts, allows for more time and cost-effective approach by optimizing the computational complexity of the algorithm based on the intensity of the motion artifact.

### 5.1.2 Health Parameter Approximation

As mentioned before, the perfusion variation is influenced directly by characteristics, including stress levels, sleep quality, hormonal levels, and  $\text{SpO}_2$ . While we have approximated the breathing rate, this approach also can monitor other health parameters, allowing for continuous monitoring of a patient without impeding on their day-to-day activities, thereby giving doctors a much better insight into the health condition of the patient.

## 5.2 Future work

COVID-19 situation limited our project's growth immensely and we believe this project holds immense potential for the future that is yet to be explored. Below are some possibilities that future groups can undertake to improve upon the results that we compiled:

1. Collection of more data. The COVID-19 situation limited our interactions

with others which restricted the participants' diversity and motions for the data we collected. Collecting more data will help make the approach more flexible and robust to a broader range of movements.

2. The use of a synthetic signal dependent on the raw PPG signal itself limits the approach's effectiveness with intense motion artifacts. Therefore, developing a method that does not rely on a synthetic signal will be more robust.
3. Develop an online system. The system we developed detected motion artifacts offline and not in real-time. In the future, creating a method to remove motion artifacts in real-time will prove to be much more helpful in the real world.
4. Implementation of hardware. Our work was primarily software-based. However, to make a product applicable in the real world, it is necessary to implement both accurate and inexpensive hardware.
5. Detection of cardiac conditions and other health parameters. One crucial direction that can serve as the next step from motion artifact removal is detecting cardiac diseases. This approach can allow for early detection and timely intervention to minimize the severity of the cardiac conditions and continuous real-time monitoring.

### 5.3 Conclusion

This project highlights the untapped potential of PPG signals in real-world applications, particularly in the medical field. While further work towards developing a more robust and accurate approach is necessary before it can be applied in the real world, the results we have obtained show some promising first few steps in the right direction, with the hope that future development will allow this approach to bear fruit, benefiting and possibly even saving millions of lives worldwide through accurate and reliable detection and diagnosis of cardiac conditions and other 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] A. T. M. N. K. K.;, “Multiwavelength pulse oximetry: theory for the future.” [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/18048900/>
- [12] A. T.;, “Pulse oximetry: its invention, theory, and future.” [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/14625714/>
- [13] Y. R. M. S. I.;, “A motion-tolerant adaptive algorithm for wearable photoplethysmographic biosensors.” [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/24608066/>
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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>

- [18] A. Tharwat, “Independent component analysis: an introduction,” *Applied Computing and Informatics*, 08 2018.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] P. Regalia, *Adaptive IIR filtering in signal processing and control*. Routledge, 2018.
- [24] 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.
- [25] 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.
- [26] 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.

- [27] N. Selvaraj, Y. Mendelson, K. H. Shelley, D. G. Silverman, and K. H. Chon, “Statistical approach for the detection of motion/noise artifacts in Photoplethysmogram,” *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2011, pp. 4972–4975, 2011.
- [28] M. Elgendi, “Optimal Signal Quality Index for Photoplethysmogram Signals,” *Bioengineering (Basel)*, vol. 3, no. 4, Sep 2016.
- [29] “Real-time robust heart rate estimation from wrist-type ppg signals using multiple reference adaptive noise cancellation.” [Online]. Available: <https://ieeexplore.ieee.org/document/7755741>
- [30] M. Wang, Z. Li, Q. Zhang, and G. Wang, “Removal of motion artifacts in photoplethysmograph sensors during intensive exercise for accurate heart rate calculation based on frequency estimation and notch filtering,” Jul 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/15/3312>
- [31] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. F. Pimentel, A. E. W. Johnson, J. Alastruey, 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.