Heart Rate and Heart Rate Variability Estimation in the Presence of Motion Artifacts

by

Jelena Nikolic-Popovic, M.Eng.

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Master of Applied Science

in Biomedical Engineering
Ottawa-Carleton Institute for Biomedical Engineering
Department of Systems and Computer Engineering

Carleton University
Ottawa, Ontario

© 2016, Jelena Nikolic-Popovic

Abstract

Vital signs such as heart rate and heart rate variability can be acquired using a variety of equipment, such as ECG, pulse oximeter and even video camera. ECG using wet electrodes is considered the gold standard, but it is not suitable for long-term patient monitoring. Dry electrodes (e.g. a wearable chest-strap) could solve this problem, but the motion of the sensor relative to the skin affects measurements. Non-contact modalities (e.g. heart rate detected from a video of patient's face) could offer further advancements in patient care, but again motion artifacts, caused by changing illumination conditions, affect measurements. Mainstream processing techniques typically assume ideal conditions and fail under realistic conditions. This thesis pinpoints the failure mechanisms of a few commonly used heart rate estimation methods under realistic conditions and proposes mitigation techniques, hoping to contribute to the effort of increasing adoption rate of modern and convenient sensor technologies for patient care.

Acknowledgements

Firstly, I would like to thank my supervisor, Dr. Rafik Goubran, for his wholehearted support and infinite patience, his guidance, the knowledge that he generously shared, and his truly motivational words when it got tough.

Additionally, I am very grateful to the members of my thesis committee – Prof. Martin Bouchard, prof. James Green and prof. Andy Adler for their insightful comments and encouragement. I am also very grateful to my fellow students - Bruce Wallace, Stephanie Bennett and Madeline Harlow for helping navigate through logistical and academic hurdles.

I thank my parents, for instilling in me the importance of education. Finally, I am infinitely grateful to my husband Vlad and my girls Nina and Sara, who were my guiding lights from one day to the next, who inspired me to set high goals, and not stop until I reach them.

Table of Contents

Abstract.....	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Illustrations.....	vii
1 Chapter: Introduction	1
1.1 Motivation	1
1.2 Problem Statement	2
1.3 Objectives and Scope	4
1.4 Thesis Contributions	5
1.5 Thesis Organization	7
2 Chapter: ECG Processing with Motion Artifacts	8
2.1 ECG Measurement	8
2.2 ECG Signal Processing	12
2.3 Motion Artifacts in Wearable ECG Measurements	15
2.4 Literature Review of ECG Processing with Motion Artifacts	16
3 Chapter: Facial Video Processing with Motion Artifacts	19
3.1 Blood Volume Pulse Measurement.....	19
3.2 Video Signal Processing	24
3.3 Motion Artifacts in Video-Based Measurements	32
3.4 Literature Review of Video Processing with Motion Artifacts.....	33
4 Chapter: Experimental Setup and Data Collection	37
4.1 Wearable Sensors	37
4.2 ECG Data	39

4.3	Video Data	40
5	Chapter: Measuring Vital Signs Using Wearable Sensors	44
5.1	Method	45
5.2	Measurement Results	46
5.3	Discussion	47
6	Chapter: Motion Artifacts and Classification using HRV Features	54
6.1	Physionet Databases.....	54
6.2	Corruption of R-R Intervals	56
6.3	Corruption of the Raw ECG Signal.....	65
6.4	Traditional Reliability Metric.....	71
6.5	Proposed Reliability Metric	72
7	Chapter: Motion Artifact and Video-Based Heart Rate Measurement	73
7.1	Method	73
7.2	Extracted BVP Signal without Motion.....	75
7.3	Extracted BVP Signal with Motion.....	79
7.4	Estimated Heart Rate with Motion.....	84
7.5	Proposed Mitigation of Motion Artifacts	88
8	Chapter: Conclusions	98
8.1	Thesis Conclusions.....	98
8.2	Future Research.....	100
	References.....	102

List of Tables

Table 1: Classification performance	61
Table 2: Per-subject excerpt classification results for CHF patient database	62
Table 3 : Fraction of RR intervals considered normal sinus beats	68
Table 4: HRV feature mean value as a function of NN/RR ratio	70
Table 5: Average heart rate from video vs reference (BPM).....	94
Table 6: Average heart rate from video vs reference (BPM) after discarding data	97

List of Illustrations

Figure 1: Signal processing for vital sign extraction	4
Figure 2: The ECG waveform for a single heart beat (source: Wikipedia)	8
Figure 3: Inter-beat intervals (IBI).....	9
Figure 4: Tachogram.....	10
Figure 5: Gel electrodes and wearable textile electrodes.....	11
Figure 6: Analog front end for ECG (ti.com/sbau171d.pdf)	12
Figure 7: Signal processing steps for extracting HRV from ECG	12
Figure 8: Adaptive filtering for motion artifact reduction.....	17
Figure 9: Blind source separation using independent component analysis	18
Figure 10: A pulse oximeter	20
Figure 11: Measuring the PPG signal using a dedicated light source.....	21
Figure 12: Acquisition of the digital PPG waveform	22
Figure 13: Basic signal processing for extracting heart rate from a video of a person.....	24
Figure 14: Advanced signal processing steps in extracting heart rate	24
Figure 15: ROI selections: face box; forehead, cheeks (one region), cheeks (two regions)	25
Figure 16: Spatial and temporal filtering of pixel intensity	28
Figure 17: BVP signal overwhelmed by motion.....	29
Figure 18: Time and spectral domain BVP signal	31
Figure 19: Rigid, large scale head motions.....	34
Figure 20: Wearable sensor for measuring heart rate and breathing rate	38

Figure 21: Wearable sensor for measuring skin conductivity	38
Figure 22 AFE4490 EVM setup with a PC	41
Figure 23 Acquired PPG data with finger inserted (top) and then pulled out (bottom) ...	42
Figure 24: Two examples of collected raw sensor data (HR and BR from Zephyr [61] SC, Activity from Q-sensor [62])	47
Figure 25: Statistics derived from raw data.	49
Figure 26: Impact of segmentation on HR statistics: segment aligned with accelerometer activity (top), segment shorter than activity (bottom)	50
Figure 27: Cardio exercise: low effort (blue) and high effort (red)	52
Figure 28: Isometric exercise: low effort (blue) and high effort (red)	53
Figure 29: Three types of RR interval sequence corruption: (a) 10% of beats shift by 5%; (b) 10% of beats shift by 20%; (c) 50% of beats shift by 5%	58
Figure 30: Impact of motion artifacts on key HRV features: misclassified false negative subjects in noisy conditions (top), and correctly classified true positive subjects in all conditions (bottom)	64
Figure 31: System under analysis.	66
Figure 32: An example of a noise-corrupted 5-min ECG segment: 0dB, 25% (top); 10dB, 25% (middle); 20dB, 75% (bottom)	67
Figure 33: A realistic motion scenario	74
Figure 34: Effect of spatial filter parameters	76
Figure 35: Skin pixel: time-domain BVP waveforms	77
Figure 36: Non-skin (background) pixel: time-domain BVP waveforms	77
Figure 37: Peak detection for subject 1 (top) and subject 2 (bottom)	79

Figure 38: Impact of deliberate motion in time domain.	80
Figure 39: BVP waveform for varying spatial filter parameters (up-down motion).	81
Figure 40: BVP waveform (left-right motion) and tracked X/Y coordinates.....	83
Figure 41: Tachogram from video with motion between beat #10 and beat #20.	84
Figure 42: BVP waveform (side to side motion).....	86
Figure 43: BVP waveform (up-and-down motion).....	86
Figure 44: Short term FFT (up-and-down motion), before and during motion.....	87
Figure 45: Short term FFT (side to side motion), during motion.....	88
Figure 46: Heart rate extraction with illumination and motion context.....	89
Figure 47 Video frames showing segments with center, right and left position, for indoor, directional light (top) and outdoor, uniform light (bottom).....	90
Figure 48: Spatially and temporally filtered BVP waveform, and motion and illumination context, for video with directional light.....	92
Figure 49: Spatially and temporally filtered BVP waveform, and motion and illumination context, for video with uniform light.....	93
Figure 50: Peak detector algorithm applied to a 10-sec segment of the reference (PPG) waveform (left) and video BVP waveform (right)	94
Figure 51: Tachogram of (a) BVP waveform from video and (b) reference waveform for video with uniform illumination.....	95
Figure 52: Tachogram of (a) BVP waveform from video and (b) reference waveform for video with directional illumination.....	95

1 Chapter:

Introduction

1.1 Motivation

Rapid technological advancements in the areas of sensor design, embedded processing, big data and deep learning algorithms, made possible by advancements in underlying technologies such as semiconductors and information technology, are quickly adopted in the consumer product space. Examples of such technologies are wearable sensors, such as a chest strap for heart rate monitoring, and the ubiquitous video camera, found not only in smartphones, laptops and cars, but also available in a wearable form-factor in head-mounted computing devices pioneered in the Google Glass product. The one thing these devices have in common is the capability to continuously record and transmit data to server farms for big data mining.

Such technologies, however trivial and meaningless their everyday consumer use may appear, do have a potential to make an impact on humanity in significant ways, for example by improving people's outcomes in the medical sense.

For example, using a video camera to record a neonate's heart rate and breathing rate instead of the traditional ECG electrodes and chest straps can help prevent and control infection rate in neonatal intensive care unit (NICU) by decreasing the number of physical interactions between medical staff and the patient, reducing handwashing requirements and simplifying patient workflow for overworked nurses.

Using wearable or non-contact sensors to detect heart rate and heart rate variability in a remote healthcare setup can help with diagnosis and early detection of conditions of patients in a home care setting for chronically ill or aging population, thereby reducing number of visits to the emergency room and improving outcomes.

Monitoring the driver's heart rate and breathing rate can alert a smart, autonomous car, that the driver is going through a medical crisis and unable to control the vehicle, at which point the car can make the decision to take over the control of the vehicle and drive it to a safe parking spot.

Finally, monitoring heart rate variability in a human-computer interaction (HCI) setting such as during a gaming or a video chat session can detect emotional crises and alert mental healthcare workers of a youth in need.

1.2 Problem Statement

What is missing to bridge the gap between consumer and healthcare uses of wearable and non-contact sensors is improved accuracy and robustness of signal processing algorithms. In most of the research focusing on monitoring, detection and classification of heart rate and heart rate variability, the signal processing algorithms are applied to data acquired under restrictive, ideal conditions. Applying the same algorithms

to data originating from consumer grade sensors and containing motion artifacts will produce unreliable results.

For example, for heart rate and heart rate variability analysis and classification, research is typically done using large publicly available databases such as Physionet [1], where the data is acquired using clinical grade ECG equipment. If data is instead collected using a wearable chest strap with dry electrodes, it may be potentially corrupted by motion artifacts and the algorithm developed for clinical-grade wet electrodes will not produce accurate results.

Similarly, for video-based heart rate detection, initial research, starting in 2007, has focused on subjects who were completely still, sitting in in a lab under ceiling-mounted neon light bulbs. Subsequent research has allowed some subject motion and used computer vision algorithms for facial feature tracking to compensate for motion. However, even with tracking, the heart rate detection algorithms break down under lighting conditions where motion results in large variations in illumination of the skin regions of interest.

The work presented in this thesis investigates the impact of realistic data acquisition conditions, involving subject motion, on the ability of the mainstream signal processing algorithms to accurately detect the heart rate and other vital signs. The goal is to identify vulnerabilities that need to be addressed in order to develop motion artifact-robust versions of those algorithms, as illustrated in Figure 1.

Figure 1: Signal processing for vital sign extraction

1.3 Objectives and Scope

The specific objectives related to analyzing and improving the impact of realistic motion scenarios on heart rate and heart rate variability detection are as follows:

- Start with basic signal processing algorithm for heart rate and heart rate variability detection and/or classification using a clinical grade sensor;
- Devise a data acquisition system for obtaining both a signal corrupted by motion artifacts representative of a wearable or a non-contact sensor, and a ground truth measurement, whereby the two data sources are synchronized with each other.
- Define a protocol for controlling the level of motion artifact during data acquisition;
- Quantify the impact of motion artifacts on each stage of signal processing, and

- Propose an approach to adapt the algorithms to realistic scenarios with the objective to improve accuracy.

1.4 Thesis Contributions

The research contributions made during the development of this thesis are as follows:

1. Multiple wearable sensors and physiological signals were investigated during exercise resulting in findings related to inter-subject, intra-subject variability, effects of subject motion, as well as challenges associated with usage of wearable sensors for vital sign measurements. This contribution resulted in the following publication: J. Nikolic-Popovic and R. Goubran, "Measuring heart rate, breathing rate and skin conductance during exercise," Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on, Bari, 2011, pp. 507-511, [2].
2. The impact of motion artifacts on heart rate variability feature extraction and classification, were analyzed by corrupting R-R intervals, to allow for comparison to the existing "clean signal" databases and classification results. This contribution resulted in the following publication: J. Nikolic-Popovic and R. Goubran, "Impact of motion artifacts on Heart Rate Variability measurements and classification performance," Medical Measurements and Applications Proceedings (MeMeA), 2013 IEEE International Symposium on, Gatineau, QC, 2013, [3].
3. The impact of motion artifacts on features of heart rate variability used in classification of chronic heart failure (CHF) patients was further analyzed under

varying and controlled motion artifact conditions, and the most vulnerable HRV features were identified. This contribution resulted in the following publication: J. Nikolic-Popovic and R. Goubran, "Towards increased usability of noisy ECG signals in HRV-based classifiers," Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on, Lisboa, 2014, [4].

4. Specific to video-based heart rate measurement, using a popular algorithm for extracting heart rate from video based on spatial and temporal filtering as a basis, it was quantified how key parameters used in the algorithm affect its performance in situations when the subject is not sitting still. Issues were identified and general approaches proposed. This contribution resulted in the following publication: J. Nikolic-Popovic and R. Goubran, "Impact of motion artifacts on video-based non-intrusive heart rate measurement," Medical Measurements and Applications (MeMeA), 2016 IEEE International Symposium on, Benevento, 2016, [5].
5. To enable detailed performance evaluation of video based camera detection under varying conditions, by also providing a ground truth heart rate reference, a measurement system was developed consisting of a pulse-oximetry clip and an analog-front end (AFE4490), to be synchronized with facial video and jointly processed. In addition, a flexible MATLAB framework for heart rate extraction was developed to support multiple video formats, color spaces, spatial filters, ROI regions, and heart rate extraction in the temporal or spectral domain.
6. Detection and mitigation methods for adverse motion and illumination conditions for video-based heart rate processing were proposed.

1.5 Thesis Organization

The thesis is organized in eight chapters. In Chapter 1, the motivation behind the research is described, along with the envisioned objectives and achieved contributions. Chapter 2 gives necessary background in biomedical measurements and signal processing techniques for estimating heart rate variability from ECG, and it also includes a literature review related to motion artifacts. Chapter 3 has the same scope, but for measuring heart rate using a video camera. In Chapter 4, data acquisition methods are described. In Chapter 5, preliminary experiments and results using a number of wearable sensors are presented, with the goal to gain insight into measurement challenges across different modalities, measurands, subjects and sessions. Chapter 6 presents experiments and results related to the impact of motion artifacts on the performance of a HRV-based classification algorithm, designed for clean ECG signal, under varying levels of motion artifact. Chapter 7 discussed the impact of motion artifacts on the performance of video based heart rate estimation. Finally, in Chapter 8, conclusions are listed and suggestions made for further research.

2 Chapter:

ECG Processing with Motion Artifacts

2.1 ECG Measurement

The electrocardiogram (ECG) is a signal which captures the electrical activity of the heart over time, as the heart contracts and relaxes. A typical waveform is shown in Figure 2. For a single heartbeat, the waveform consists of four segments: (1) P wave (represents atrial depolarization); (2) QRS complex (represents ventricular polarization); (3) T wave (ventricular repolarization); (4) U wave (papillary muscle repolarization).

Figure 2: The ECG waveform for a single heart beat (source: Wikipedia)

The heart rate is measured by measuring the interval between two heart beats, also called Inter-Beat Interval (IBI). Inter-Beat Intervals are most commonly measured using the R peak as the reference point, as shown in Figure 3. With that, the terms IBI and RR-intervals are used interchangeably.

Figure 3: Inter-beat intervals (IBI)

From IBIs, one can compute:

- (1) Instantaneous heart rate, as the inverse of the IBI;
- (2) Average heart rate, as a number of IBIs over some period of time, divided by the sum of their durations;
- (3) Heart Rate Variability, as the instantaneous heart rate time series. Heart Rate Variability is represented using a tachogram, showing the IBI as a function of heart beat number, as shown in Figure 4.

Figure 4: Tachogram

The ECG is acquired via electrodes placed on subject's skin. The medical gold standard consist of multiple electrodes (up to 12), and using wet (gel) electrodes to improve electrode-skin contact, i.e. reduce impedance. Recently, with the advancements in wearable technologies and the desire for long term monitoring, capacitive (dry) electrodes have become more popular. Research into electrical properties of different types of electrodes has generally concluded that the level of noise and the artifacts are greater on the dry/textile electrodes than on gel electrodes [6], and that dry/textile electrodes in general have a higher dependency on the electrode-skin pressure [7] which could introduce greater intra-subject variability. However, the advantages of gel electrodes diminish in long-term monitoring scenarios when the gel invariably dries out. Some of the more popular types of electrodes are shown in Figure 5.

Figure 5: Gel electrodes and wearable textile electrodes

In modern equipment, the electrodes capture an analog signal which is then filtered, amplified in the analog domain and digitized. A typical analog front end for ECG acquisition is shown in Figure 6. The analog waveform is digitized using a sampling rate of 128 samples/sec or higher, and precision of 16-24 bits.

Figure 6: Analog front end for ECG (ti.com/sbau171d.pdf)

2.2 ECG Signal Processing

The basic signal processing steps are shown in Figure 7.

Figure 7: Signal processing steps for extracting HRV from ECG

The digitized ECG signal is first filtered to remove basic artifacts like drift, baseline wander, and 60Hz power supply noise, prior to detection of QRS complexes, each of which indicates one heartbeat.. The time series of inter-beat intervals can be further processed for the purpose of analyzing heart rate variability and using it for example for classification of diseased vs normal subjects. In the following paragraphs, each of these steps is discussed specifically with motion artifacts in mind, to give the necessary background for the results discussed in later chapters.

2.2.1 Filtering

Some of the noise sources in the ECG can be eliminated from the ECG waveform using simple filtering techniques. For example, 60 Hz power line interference can be eliminated with a 60Hz notch filter, and baseline wander and respiration can be eliminated by removing the low frequency components using a high pass or band pass filter.

The more troublesome types of artifacts are those that overlap with the ECG frequency range of interest in heart rate detection (0.8-4 Hz for the heart rate range of 50bpm-250bpm). Those artifacts are caused by muscle contraction and skin-electrode contact motion. This is still an active area of research, experimenting with techniques such as independent component analysis, wavelets, empirical mode decomposition, and even neural networks. Many of the techniques have the disadvantage of being only suitable for off-line processing rather than real-time monitoring, due to latency (in case the algorithm requires a block of data rather than operating sample-by-sample), or complexity (for example in case of neural networks which have high processing requirements and may not be meeting real-time requirements in embedded systems), and as such will not be considered in this work.

2.2.2 QRS detection

The most important, and sometimes the only, processing step in heart rate estimation is the detection of the QRS complex shown in Figure 2. This seems fairly simple to do in a clean ECG waveform such as that shown in Figure 3, but in noisy ECGs, this can be a challenging task. The result could be missed or extra beats, leading to

erroneous heart rate result, potentially causing errors in diagnosing arrhythmias and greatly affecting any subsequent processing of heart rate variability.

The traditional approaches to QRS detection are based on signal threshold, slope (first derivative), and additional heuristics regarding where in the time waveform a heartbeat can be expected.

In this work, we use two QRS detection algorithms which are readily available in the Physionet framework [1]: (1) SQRS, which explicitly estimates signal quality and based on it, dynamically adjusts slope and timing criteria, and (2) WQRS, based on using a nonlinear scaling factor for ECG curve length.

2.2.3 Feature Extraction and Classification

Medically important features (measures) of heart rate variability have been defined by a medical task force [8] and can be categorized as follows:

- (1) Time Domain (Statistical measures) – Require a pre-processing step to clean up ECG abnormalities, due to sensitivity of the R-R interval data. The pre-processing step turns RR interval sequence into NN (Normal-to-Normal) interval sequence.

The following statistical measures reflect gradual changes:

- a. AVNN – average NN interval
- b. SDNN - Standard Deviation of NN Intervals
- c. RMSSD - Root Mean Square of Successive Differences
- d. SDANN - Standard Deviation of Averaged NN intervals
- e. pNN50 - % of intervals that differ from the preceding one by more than 50msec

(2) Frequency domain (Spectral) measures: Frequency ranges of interest are divided as follows::

- a. HF (0.15-0.40 Hz) - respiratory sinus arrhythmias
- b. LF (0.04-0.15Hz) - rhythmicity of systolic arterial pressure, observed under conditions of sympathetic activation
- c. VLF(0.0033-0.04Hz) - for long term recordings
- d. ULF(below 0.0033Hz)

(3) Geometric measures: Based on histograms of intervals or differences between successive intervals, or Pointcare's plot (interval duration plotted against previous interval duration).

(4) Non-linear methods.

The choice of features used for classification depends on the specific classification task, and the associated discriminatory power of features. For example, it was found that for classification for chronic heart failure patients vs normal subjects, the features that have the most discriminatory power are a combination of statistics and frequency-domain features, such as AVNN, LF/HF, as well as some non-standard features [9].

2.3 Motion Artifacts in Wearable ECG

Measurements

The source of motion artifacts in wearable equipment using dry electrodes is the varying impedance caused by displacement of the electrode relative to the skin. Consider a wearable chest strap worn while walking or running: with each step, there are slight

motions of the torso that cause slight displacement of the strap relative to the skin and degrade the electrical signal picked up by the electrode. This effect becomes even more pronounced when moving to truly “wearable” sensors, such as textile electrodes embedded in the garment.

In clinical grade, wet electrodes, the variations in impedance are minimized by using a gel layer between skin and the electrode. This maintains constant impedance in spite of the motion of the skin and the electrode relative to each other. As the gel dries out with time however, the impedance degrades, reduces signal quality and causes heart beats to be missed, or extra heart beats to be falsely detected [10]. With dry electrodes, the problem is aggravated by the fact that the impedance is already significantly higher than with gel electrodes since there is an air gap between the skin and the electrode, and it varies significantly with motion [11]. The impedance is also a function of many environmental parameters such as applied pressure and padding, temperature and humidity [12], which further complicates repeatability of measurement conditions in a long-term monitoring setting.

2.4 Literature Review of ECG Processing with Motion Artifacts

The first wave of research into noise-tolerant and robust QRS detection has focused on the noise sources in medical-grade ECG with wet electrodes, such as baseline drift, muscle activation, respiration, power source noise, and the ability to detect arrhythmias (extra beat and ectopic beats). This “first wave” of algorithms was based on simple

filtering (low pass, notch or bandpass), amplitude thresholds and slope properties of the QRS complex [13]. One of the more advanced algorithms is WQRS, which converts the ECG signal into a curve length signal using a transform with a non-linear scaling factor [14]. The algorithms were primarily tested by synthetically adding a noise signal to a noise-free ECG signal.

The research into robust and noise tolerant QRS detection has been revived in the last decade due to new developments in the field of wearable ECG measurement systems, ranging from capacitive, dry electrodes, to smart clothing. The main approaches can be classified into: (1) adaptive filtering; (2) blind source separation (BSS) with independent component analysis (ICA), and, less often, wavelet filtering and neural networks.

In adaptive filtering, the basic idea is that a reference signal is used to adjust filter coefficients, as shown in Figure 8.

Figure 8: Adaptive filtering for motion artifact reduction.

For the purpose of motion artifact reduction, a reference signal is often obtained using additional sensors, such as accelerometers [15], skin stretch sensors [16], additional electrodes [17], electrode-tissue impedance (ETI) signal [18].

The generic principle behind using ICA is shown in Figure 9. The purpose of ICA is to recover individual sources from a set of observations which are assumed to be a linear mixture of sources. In this application, it is assumed that there are two sources, ECG and the motion artifact, and ICA is performed to separate the two components and recover ECG [19]. Ideally, the number of sensors (observations) should be the same as the number of sources, so this method will work if there are at least two ECG leads.

Figure 9: Blind source separation using independent component analysis

With wearable measurement systems, obtaining additional channels is less expensive than was the case with traditional ECG systems with wet electrodes, both in terms of the time needed to setup the system on a patient, as well as the actual cost. For example, in a wheelchair based system, electrodes are placed both on the seat and the backrest, providing additional channels [20]. In [21], seven channels are collected using textile electrodes to provide greater spatial diversity.

Additional signal processing approaches include (1) auto-correlation, which relies on similarity of QRS complexes [22]; (2) modeling the IBI duration as a time-varying history-dependent inverse Gaussian distribution [23]; (3) using ensemble empirical mode decomposition to remove motion-artifact induced noise [24]. Methods for monitoring signal quality and reliability of measurement have been discussed in [25], [26] and [27].

3 Chapter:

Facial Video Processing with Motion Artifacts

3.1 Blood Volume Pulse Measurement

3.1.1 Blood Volume Pulse and Photo-Plethysmography

The physiological signal of interest is the cardiovascular pulse wave, also known as blood volume pulse (BVP), as it propagates into the capillary bed underneath the skin. With every heartbeat, blood is pumped through even the smallest blood vessels found in the capillary bed of tissues, and affects their volume in a pulsatile manner, therefore getting the name Blood Volume Pulse. Since the blood volume changes are synchronous to the heart beat, measuring the BVP signal gives a proxy (indirect) measurement of the heart rate.

Measuring changes in volume is called plethysmography, and performing the measurement using optical modality is called Photo-Plethysmography (PPG). Changes in

the light absorption in tissue are correlated to the changes in the volume of the tissue due to the cardiovascular pulse wave and can be used as a proxy measurement of the heart rate.

The most common device which uses the principle of photo-plethysmography is the pulse oximeter which non-invasively measures oxygen saturation. A typical pulse oximetry device is shown in Figure 10.

Figure 10: A pulse oximeter

This type of device has been clinically used for decades, and it is popular due to its non-intrusive nature (measurements are made at skin surface), and low cost.

Its principle of operation is as follows: skin is illuminated using a dedicated light source, and the amount of light reflected is measured using a photo detector. The resulting signal is the PPG signal, which is the inverted waveform of the light absorbed by the illuminated area, also called “catchment area” [28]. Traditionally, both the light source and the photodetector are placed on the skin (typical sites are finger, earlobe, baby toe), so the measurement falls in the “non-intrusive” and “non-contact” category. The measurement principle is illustrated in Figure 11.

Figure 11: Measuring the PPG signal using a dedicated light source

The light source is a LED emitting a single wavelength, typically in the red and/or infrared region of the spectrum. The specific wavelength depends on the exact application of the PPG signal, but some considerations are [28]: (1) minimizing absorption of the light by water content of the tissue (water has a window of wavelengths in the visible red and near infrared portion of the spectrum where it passes light through more easily); (2) absorption properties of oxyhaemoglobin (indicated oxygenated blood) and reduced haemoglobin (use a single, isobestic, wavelength to measure haemoglobin, and two distinct wavelengths to measure how much of haemoglobin is loaded with oxygen), and (3) tissue penetration depth of different wavelengths.

In modern devices for measuring PPG waveforms, the optical signal acquired by the photo detector is converted into an electrical signal, sampled and digitized for further processing. An example of such system is shown in Figure 12.

Figure 12: Acquisition of the digital PPG waveform

The time-varying (AC) component of the PPG waveform is synchronous with the heart beat and can be used as a source for heart rate measurement. This is not without problems however. The accuracy of the heart rate information obtained from the PPG can be greatly affected by movement artifacts of the probe relative to the skin, during cardiac arrhythmia, and respiratory effects. Signal processing algorithms have been developed to combat these issues, with good results relative to the ground-truth, ECG obtained heart rate measurement [28]. Not only heart rate, but also the oxygenation levels measured using the PPG modality are affected by motion artifacts [29].

3.1.2 Measuring the PPG Signal in a Non-Contact Manner

As described in the previous section, the traditional equipment for retrieving the PPG signal is of the non-invasive, contact type. In many scenarios, non-contact

measurement would be preferred and could improve patient comfort and safety. The following two examples illustrate such situations: (1) long term monitoring is needed, where prolonged contact with equipment would result in skin sores and general discomfort to the patient; (2) making contact with the patient to position a measuring device could jeopardize its safety – this is especially the case in NICU setup where patients are vulnerable and any contact with the patient should be minimized to prevent and control infections.

A small number of researchers, such as in [30] and [31] have considered non-contact photo-plethysmography using dedicated (and multiple) single-wavelength light sources. This type of measurement falls in the category of non-invasive, non-obtrusive (i.e. the patient is aware of the measurement being taken), and non-contact.

The natural evolution of the research is toward a non-invasive, non-obtrusive, non-contact measurement. This could be achieved if dedicated light sources and obvious photo detectors were to be eliminated, i.e. leading towards using ambient light as light source, and a standard video camera. In many situations, the patient would not be aware of the camera, making the measurement truly non-obtrusive. In other scenarios, although the patient may be aware of the camera, its use for physiological measurements may not be evident again making the measurement non-obtrusive. This would be useful in situations where the act of taking measurement may affect the measurand itself (such as in white coat hypertension [32]).

The first published research in this area is less than 10 years old. Two seemingly independent research teams, [33] and [34], in 2007 and 2008, respectively, have established that it is possible to extract a BVP waveform from video of a person's face,

from changes in intensity of certain skin pixels over time. To be able to find such a small signal in a very noisy environment, after the video is cropped to a region of interest (ROI), signal processing techniques such as spatial averaging and bandpass temporal filtering have been applied to the video, after which it was possible to observe the heart rate as a spectral peak (i.e. in frequency domain). The basic steps are shown in **Error!**

Reference source not found..

Figure 13: Basic signal processing for extracting heart rate from a video of a person

3.2 Video Signal Processing

In this section, the signal processing chain as it applies to challenging video is examined in more detail. The expanded set of steps consist of ROI selection, ROI tracking, spatial processing, channel selection and combining, temporal processing, and heart rate extraction. This is shown in Figure 14.

Figure 14: Advanced signal processing steps in extracting heart rate

3.2.1 ROI selection

The first step in video processing is to find the region of interest (ROI), i.e. pixels in the video frame which will be used to find the PPG signal. From studies of still subjects [34] the top candidates are the forehead and the cheek regions. The most frequently applied approach to finding the ROI is to use OpenCV face detection [35], [36], [37], [38], [39], [40], [41], [42] which generates a face box. To extract the more relevant portions of the face box, one can simply choose to use some portion of the width and height [35], [41], or apply a skin detection algorithm to find and apply a “skin mask” to the ROI [42]. The second most popular selection is the forehead rectangle [43], [44], in some cases subdivided into multiple regions [45]. Since the forehead could be occluded due to hairstyle, the cheek area is also of interest, either as one single region between eyes and lower cheek regions [46], or as two separate regions (left and right). The location of the above mentioned ROI types on a person’s face are shown in Figure 15.

Figure 15: ROI selections: face box; forehead, cheeks (one region), cheeks (two regions)

3.2.2 ROI tracking

Once the ROI is identified, it needs to be tracked on a frame by frame basis to minimize motion artifacts. The common approach is to use OpenCV face detection or

feature detection algorithms. Multiple implementation options are possible, some of which (KLT, [47]), do better with motion than others (Viola-Jones, [48]). Some advanced algorithms detect many facial features (a.k.a. landmarks) which can be used to define and track ROI very precisely [49]. For the purpose of this thesis, visible facial markers will be used and the tracking problem will not be considered since it appears that suitable solutions already exist.

3.2.3 Color Space, Channel Selection and Combining

Modern digital cameras used in current video research have CMOS imagers which produce video consisting of three channels, typically Red, Green and Blue (RGB). In the original research from [34], it was reported that the green (G) channel has the best PPG signal, but that the signal was recognizable in the other two channels as well. This was explained by the haemoglobin/oxyhaemoglobin absorption being the highest in the particular region of the visible light spectrum [28]. This finding led to two main directions in channel selection for video-based heart rate extraction: (1) use green channel only [36], [37], [50], and (2) use all three channels, and combine them using signal processing. The earliest reference to combine multiple camera channels is [35], where blind source separation (BSS) of the three channels into three independent components, using independent component analysis (ICA), resulted in one component having the best quality signal, better than the green channel alone. Other approaches to combining channels include linear discriminant analysis (LDA) [39], independent vector analysis (IVA) [46], Kalman filtering [51], adaptive difference [52], and using Pearson correlation coefficient [53]. Some research has reported inconsistent results though for motion

scenarios, where the component at the output of the ICA algorithm which had the best signal varied from one scenario to the next and it was not possible to determine a priori, or that the green channel alone contained the better signal than any of the ICA components.

Alternative color spaces exist as well. The RGB camera output can be converted into luminance/chrominance component using color space conversion (CSC) algorithms, resulting in LUV or YCbCr formats [54]. In [42] for example, the U component (red to green) was used and it was found to be less motion sensitive than the green channel only.

3.2.4 Spatial and Temporal Filtering

A single pixel, at a fixed location in a video frame, suffers from camera noise artifacts, and the PPG signal is not visible. To combat the camera noise effects, one can compute average pixel intensity in the channel of choice, over some predetermined number of pixels. This averaging operation is effective in removing the camera noise.

In addition, a temporal band-pass filter in the spectrum of interest is applied to improve signal quality. The frequency range corresponding to the heart rate range of 45bpm to 240bpm is [0.75Hz 4Hz].

The effects of spatial and temporal filtering are illustrated in Figure 16.

Figure 16: Spatial and temporal filtering of pixel intensity

One important parameter in spatial filtering is the size of the area to be averaged. Values found in literature range from a very small area, for example 40x40 pixels [55], to an entire face box [35]. The parameters that impact the size of the averaging area are: (1) camera resolution, (2) the distance between the subject and the camera, (both of which impact the size of the ROI in terms of pixels), and (3) the quality of the camera. The greater the camera noise, the more pixels are needed to be averaged. However, in motion scenarios, a larger pixel area may imply a greater range of changes in illumination. One possible solution is to divide the ROI into multiple sub-regions, average over the sub-region and then apply a combining/selection algorithm to decide which region to use for further processing – this was done in [45] using 6 regions in the forehead and in [46] using 7 regions for the area of the face between eyes and upper lip. In both those cases, the sub-regions are defined from facial landmarks which are tracked on a frame-by-frame basis. The size of ROIs may therefore vary dynamically from one frame to another, as the subject moves, which further complicates the sub-region combining/selection algorithm.

For temporal filtering, the important parameter is the location of the band pass. In motion scenarios, the frequency of motion often overlaps with the desired frequency range for heart rate, and in case of large illumination changes, the variations in pixel intensity can overwhelm the PPG signal. This effect can be seen in Figure 17. One solution to combat that is to use an adaptive, narrower band pass filter, where the center frequency for one time segment is obtained from knowledge of the heart rate in the previous segment, assuming that the variations in heart rate from one segment to the next are relatively small [43], [52].

Figure 17: BVP signal overwhelmed by motion

3.2.5 Heart Rate Extraction in Spectral and Temporal Domain

At this point in the signal processing chain, a single time-domain waveform, representing a facial PPG signal is available and the task is to extract the frequency of the dominant periodic signal which is synchronous with the heart rate. Heart rate extraction

can be done in the time domain or the frequency domain, and each approach has its advantages and disadvantages.

In time-domain, the locations of local maxima are found using peak detection algorithms. Since the sampling rate for the time domain waveform is equivalent to the video frame rate (typically 15-30 fps), some researches choose to interpolate it to ~1KHz using cubic spline interpolation [36]. This is especially important if the goal is to extract heart-rate variability (HRV) information from IBI (inter-beat intervals).

The advantages of the time domain approach is that it can be easily implemented in real-time. The disadvantage is that it could be susceptible to noise of nearby spectral content. In addition, the local maxima finding algorithm often uses a threshold value which, when varied, produces vastly different results in terms of false positives and missed peaks.

In frequency domain, heart rate is simply found as the largest spectral peak. The waveform is converted into the frequency domain using the FFT-based Welch method [56] for computation of power spectral density (PSD). The waveform is first divided into time-domain windows, which could be overlapping (sliding window approach) or non-overlapping. The size of the window is a tradeoff between time domain and spectral domain resolution. For example, to compute an average heart rate over a period of 20-30 sec, which, at ~30 frames per second, represents ~1000 samples, one can perform a 1024-point FFT [35]. An example is shown in Figure 18: (a) after spatial filtering (b) after temporal filtering, (c) FFT after temporal filtering, with the heart rate being the largest spectral peak. If the heart rate varies during this relatively large time window, there may

be undistinguishable, closely located frequency peaks which would reduce the accuracy of heart rate detection in the spectral domain.

Figure 18: Time and spectral domain BVP signal

If the goal is to update the heart rate more frequently, using a smaller number of time series samples, the spectral resolution suffers, since the number of bins depends on the number of samples available. One way to increase spectral resolution is by zero-padding the waveform to 3-4 times its length and then performing a larger size FFT. For example, if an 8-sec video (~250 samples) is zero-padded to 4 times its length, a 1024-point FFT can be used to obtain the frequency-domain representation.

A common method of extracting more frequent readings from a single time-domain waveform is to use overlapping sliding windows, and perform an FFT for each window [39], [41], [35]. For example, in [39], 8-second windows with 6-second overlap are used, and a zero-padded FFT is performed once every 2 seconds giving an updated heart-rate reading.

3.3 Motion Artifacts in Video-Based Measurements

Moving from the contact to the non-contact modality for acquiring the BVP signal does not result in elimination of motion artifact issues. Although there is no physical contact via skin, as is the case with the pulse oximeter finger clip, the subject moves relative to the probe which, in this case, is camera. There are two types of movement possible: (1) probe relative to a stationary subject (for example, camera shake), and (2) subject relative to a stationary probe. The first problem is relatively easy to solve, using tripod (a priori), or posteriori, using widely available camera-shake compensation algorithm, such as the one familiar to YouTube users [57].

The second type of motion artifact (moving subject recorded by a stationary camera) is a far more difficult one to resolve, and it has the following two aspects: (1) subject motion results in the movement of the ROI from one video frame to the next, and (2) subject motion results in changes in illumination from one frame to the next.

As can be seen, most research to date has focused on resolving the “moving ROI” problem while assuming uniform and stationary light conditions. The problem of non-inform, and time-varying illumination has not been addressed, and it is a missing piece for turning video-based heart rate detection into clinically usable technology.

3.4 Literature Review of Video Processing with Motion Artifacts

Many of the research papers discussing video-based heart rate detection claim the algorithms and results to be “motion-tolerant” or “motion-robust”. However, the definition of “motion” greatly varies from one publication to the next. Roughly, motion can be classified as:

- (1) large, rigid motion, where the head is turned in one of 6 possible directions (yawing, rolling, pitching, swaying, surging, heaving) as described in [53] and shown in Figure 19.
- (2) Small, non-rigid motion, which occurs during talking or making various facial expressions. These are typically considered in human-computer interaction scenarios.

In some cases, even small involuntary motion produced when the subject’s head was not placed in a chinrest was considered source of motion artifacts [58]. In another case, a very specific periodic motion which occurs while subject is using exercise equipment such as bike or stepper is considered [41].

Figure 19: Rigid, large scale head motions

A similar situation applies also to considerations of different illumination conditions. In some cases, the illumination source is constant, but its intensity varies, for example by dimming the fluorescent ceiling lights in the lab [59]. In other cases, the direction of light is changing by placing a chair next to a window in 4 different orientations [53]. Only a few contributions consider motion-induced illumination artifacts, i.e. shadows due to subject motion while the ambient light is uneven [55], [53].

Due to the above described inconsistency in motion scenarios, comparison of research results is difficult and it will not be the focus of this section. Rather, the focus is on categorizing the approaches.

In most cases, the chosen approach to compensate for motion is to implement a computer vision algorithm to track subject's face or a set of facial features, therefore allowing the ROI to move within the video frame from one frame to another. The initial work used the OpenCV Viola-Jones algorithm for face detection, [48], [37]. It was noted however that this algorithm does not do well with motion, so additional algorithms are

used to detect features [43], [60] or facial landmarks [46], [50], and track them over time using KLT (Kanada Lucas Tomasi) algorithm from MATLAB Computer Vision System Toolbox [36] .

Alternative motion tracking algorithms include modeling to learn shape (Active Shape Model) [38], optical flow [51], or using a Markov model to model motion [45].

In some cases, in addition to tracking, further techniques have been proposed to mitigate motion artifacts:

- Adaptive band-pass filter. Often times, especially for large rigid motion, the motion overlaps with the heart rate frequency band and the temporal bandpass filter which covers the entire possible band (typically from $\sim 0.5\text{Hz}$ to 2Hz) is not effective. If however the video is processed one short segment (5-10sec) at a time and the heart rate from the previous segment is known, a more narrow bandpass filter, with adaptive center frequency determined from the previous segment, can be employed and can eliminate some of motion artifacts [38], [43], [40].
- Kalman filter [37], [51], where the PPG time series corrupted by motion artifact is modeled using state-space modeling, and the system parameters of the model are subsequently used to build a Kalman filter;
- Multiple ROIs: random patches [55], forehead and cheeks [50], forehead subdivided into 6 regions [45];
- Multiple channels: red/green [52], multiple imagers [58];
- Compensating for illumination changes using average intensity, which relies on the constant illumination of the background (a.k.a. illumination rectification) [60];

Finally, if too much motion is detected, which may imply that the resulting illumination variations simply overwhelm the PPG signal, “bad” segments of video can simply be discarded [40].

4 Chapter:

Experimental Setup and Data Collection

4.1 Wearable Sensors

The equipment for the research into vital sign measurement during exercise was chosen based on the following requirements:

- is wearable;
- has multi-hour on-board logging capability;
- has data exporting capability (for post processing)
- collects physiological signals along with information about motion;
- is readily available for purchase.

These requirements led to the choice of the Physiological Status Monitoring (PSM) solution from Zephyr Technologies [61] shown in Figure 20 and Q-sensor from Affectiva [62], shown in Figure 21.

Figure 20: Wearable sensor for measuring heart rate and breathing rate

Figure 21: Wearable sensor for measuring skin conductivity

The chest belt sensor shown in Figure 20 measures the following physiological signals:

- Heart Rate (HR)
- Breathing Rate (BR)
- Skin Temperature (ST)

In addition, the device measures the following activity data:

- Activity level, in Vector Magnitude Units (VMU), which is a sum of movement in all three planes, integrated over 1 second interval, and measured by the three-axis accelerometer;
- Posture (attitude of the device relative to vertical, in degrees).

The activity data will prove useful for segmentation of physical data as well as time alignment of multiple devices. All data is updated once per second.

The chest belt can store up to 480 hours of data. Data is transferred to a PC using USB interface.

The bracelet sensor shown in Figure 21 measures:

- Skin conductance (SC)
- Skin temperature (ST)

It also has an on-board three-axis accelerometer whose data is provided in a raw format. The bracelet's storage capacity is sufficient for logging months of data at a time. Data is transferred to a PC using USB interface.

4.2 ECG Data

The research into heart rate variability and its susceptibility to motion artifacts is based on digitized, long-term (~24hr) ECG recordings stored in Physiobank archives [1] whose collection of physiological databases is widely used in medical and engineering literature and is continuously being expanded.

The specific databases used are the MIT-BIH Normal Sinus Rhythm Database (NSRDB) and the BIDMC Congestive Heart Failure Database (CHFDB). NSRDB database consists of recordings of 18 subjects, sampled at 128Hz, while the CHFDB consists of recordings of 15 subjects, sampled at 250Hz. Both databases are annotated, i.e. along with the full recordings are provided locations of the beats (QRS complex), validated manually and therefore assumed to be gold standard.

A third Physiobank database is used to provide the noise signal used to corrupt the above mentioned clinical-grade ECG recordings and simulate effects of wearable electrodes. The MIT-BIH Noise Stress Test Database was initially created to test arrhythmia detectors [63], and is therefore relevant for any system or algorithm which relies on accurate beat detection such as HRV analysis systems considered in this work. The database consists of half-hour recordings of three types of noise: baseline wander, electrode motion and muscle activity. In this work, we are primarily concerned with electrode motion noise, since it is considered to be the most troublesome type of noise for beat detection.

Physiobank also provides a program used to combine clean ECG signal with a noise signal, and allows independent control of noise duration and the level of noise.

4.3 Video Data

4.3.1 Equipment

The objective of the experiments is to acquire video of the subject along with ground truth heart rate measurement.

The video is recorded using a Sony Alpha 7 camera [64], using a tripod to minimize camera motion. Camera frame rate is 30 frames per second. The video is recorded in MP4 format and converted to AVI format on the PC for video processing. Advanced features which may impact the appearance of illumination and motion are turned off. While a lower end camera such as a standard PC-based web cam would also

have adequate resolution, it would introduce other artifacts such as variable (less reliable) frame rate, which is why it was decided against in the current research.

Ground truth heart rate is acquired using the Texas Instruments AFE4490 EVM [65]. This is an evaluation module for an optical biosensing (i.e. PPG) analog front end semiconductor chip, which comes with a DB9 oximetry cable with IR and Red LEDs, and is connected to a PC using USB connection. The data acquisition is controlled using a GUI where data is displayed in real time and can also be recorded in a Excel spreadsheet format for further analysis. The AFE4490 EVM setup is shown in Figure 22.

Figure 22 AFE4490 EVM setup with a PC

The two data sources are synchronized using a marker which is visible on both waveforms: the action of pulling out the finger out of the DB9 oximeter sensor clip. This action is performed at the end of the recording. In the video data file, one can pinpoint the

exact frame where the pull out occurred. On the AFE4490 data capture, pulling out of the finger results in the sudden jump in the reading (in Volts) from a level of $\sim 0.1\text{-}0.2\text{V}$ as seen in Figure 23 (top), to the supply voltage level of 1.2V , as shown in Figure 23 (bottom, around sample 170600).

Figure 23 Acquired PPG data with finger inserted (top) and then pulled out (bottom)

4.3.2 Protocol

The goal of the experiment is to collect video and ground truth data under realistic and variable but controlled motion and illumination conditions. Reference video segments without motion and with ideal light are recorded as well. Since the research presented in this paper is exploratory and focusing on intra-subject variability while changing environmental conditions, it was deemed sufficient to limit the experimentation to a single subject. It is understood that this restriction will, to some extent, limit the

breath of conclusions that can be drawn – factors such as skin tone, age, make-up, hair styles may all influence the results.

The type of motion we are interested in are those that could occur naturally in human-computer interactions, or during driving, such as moving back/front, leaning left/right, or heaving up/down; and head shake, nod or tilt. The two extreme cases considered are one-shot motion (the subject moves to a position, stays there for some time, and then moves back), and periodic motion (the subject performs back and forth motion a certain number of times). Non-rigid, smaller scale motion, i.e. those occurring when subject changes facial expressions, such as frowning, talking, smiling, will not be considered as the related illumination changes are less drastic. Another possible source of illumination changes is the change in the lighting condition itself, rather than the motion of the subject. The solution to this scenario is a subset of the solution for the moving subject and static light, since the shape, size and location of the facial ROI is static, which simplifies the ROI tracking and extraction.

As a light source, we explore directional daylight which is typical of home settings and during driving, as opposed to interior ceiling lights, which are expected to have a smaller degree of illumination related motion artifacts.

5 Chapter:

Measuring Vital Signs Using Wearable Sensors

The work presented in this chapter focuses on wearable devices, which could be of importance not only to a health-conscious consumer, but also in the healthcare system, for monitoring of patients whose prescribed treatment includes some form of physical activity. While previous related work typically uses an accelerometer based approach, it is desirable to acquire a more complete picture of impact of exercise on the subject by adding physiological signals such as heart rate, breathing rate and skin conductivity and correlating them with motion signals. Ultimately, such framework can be used to assess progress over time for different types of exercise regimens. The objective in this chapter is to evaluate discriminatory power of features extracted from wearable sensors in remote monitoring of exercise.

5.1 Method

Data from the equipment described in Section 4.1 was collected from eight healthy volunteers, four male and four female, all in the same age group (35-40 years) and in a very good physical condition, during a test protocol which lasted about 20 minutes per person. The test protocol consisted of following epochs:

1. 3 minute relaxation (basal state)
2. 5 minutes of isometric exercise: The subjects were asked to perform a set of five isometric exercises lasting one minute each (squat, left-leg lunge, right-leg lunge, plank position and a seated abdominal move). They were asked to minimize movement while holding each position and also to adjust the position such as to experience strong engagement of muscles.
3. 3-7 minute rest (until heart rate and breathing returned to normal)
4. 3 minute step test: The subjects were asked to step up and down a 30cm-tall bench, while keeping a steady rhythm (1 second up, 1 second down).
5. 3-5 minute rest (until heart rate and breathing returned to normal)

The transitions between protocol epochs were designed in such a way that the data corresponding to different epochs, viewed in time domain, can be easily segmented based on the readings from the motion sensors (accelerometers and the posture sensor) which were present on-board both sensor systems.

5.2 Measurement Results

After each subject executed the above described protocol, data files were downloaded from sensors to a PC. As each device has an on-board real-time clock (RTC), the data files were time stamped, which facilitated integrating and aligning data from two sensors using Matlab. Some manual fine tuning (using accelerometer data from both devices as a reference) was also necessary because only one of the devices was able to synchronize its on-board clock to the PC. The examples of aligned heart rate, breathing rate and skin conductance are shown in Figure 24, along with activity data for ease of interpretation. The highlighted segments represent the isometric exercise epoch and the step test epoch, respectively. The isometric epoch is characterized by activity spikes during posture changes and minimal activity levels during each posture. The step-test epoch is characterized by a constant elevated activity level.

Figure 24: Two examples of collected raw sensor data (HR and BR from Zephyr [61] SC, Activity from Q-sensor [62])

5.3 Discussion

As mentioned at the beginning of this chapter, the objective is to evaluate discriminatory power of features extracted from wearable sensors in remote monitoring of exercise. To this end, statistics of data acquired from each physiological sensor is analyzed across the five exercise epochs, and inter-subject and intra-subject variability is discussed.

5.3.1 Inter-Subject Variability

Some initial observations can be made when comparing measurements from the eight volunteers:

- The step-test epoch is clearly identifiable from available activity data. The fact that this epoch corresponds to an intense physical activity is also confirmed by a rise in heart rate. The shape of the HR curve for all subjects is quite similar. Breathing rate also increased in most cases, although not as sharply.
- The isometric test epoch can also be relatively easily identified in the graph, given prior knowledge of the test protocol, namely that the epoch consist of five postures which are held steady for 1 minute each. The five “spikes” on the activity data indicate posture changes. The execution of posture transitions is subject-dependent, so for some subjects the spikes are clean whereas for others they are “noisy” indicating that the subject’s is not able to hold a steady, firm position during the isometric exercise.
- The isometric epoch is also characterized by an increase in the heart rate, although the shape of the HR curve varies for each subject. This was expected as some subjects tried harder than others. This could also be the reason why the breathing rate changes during this epoch vary from one person to the other, as some subjects deal with physical effort by breathing deep and slow while others do not have such control of their breath.
- Finally, the skin conductance levels seem to be quite subject-dependent. Other than a general observation that female subjects have lower levels of SC than male subjects, and that the skin conductivity increases with physical effort, no quantitative conclusions can be made just from the raw data.

Simple statistical analysis was performed for all three measurands, all six subjects and all five epochs (rest, isometric, rest, step, rest). Minima, maxima and means were

calculated in each case. The data is displayed in Figure 25. Minimum, maximum and mean for each subject and each epoch is represented using an error bar.

Figure 25: Statistics derived from raw data.

The mean heart rate, as seen in Figure 25, appears to be the best feature as it changes significantly between rest epochs and exercise epochs, both for isometric and cardio exercise. The min/max ranges are similar for all of our six subjects, although this might not be the case for subjects of other age groups and health conditions. However, the heart rate alone can not be used to differentiate between cardio and non-cardio exercise. Other physiological and/or activity measurements will be needed for this. In addition, the discriminability depends on how well the epoch is defined. In Figure 26, we show that the HR statistics have a much better discriminability if the statistics are calculated over an interval which is by about 30 seconds shorter than the actual activity interval (bottom), then if they are calculated over an interval which coincides with high measured activity levels from accelerometers (top). Even though the average values are

similar between the two segmentation strategies, the latter one has a tighter range, as seen from the lengths of the error bars.

Figure 26: Impact of segmentation on HR statistics: segment aligned with accelerometer activity (top), segment shorter than activity (bottom)

5.3.2 Intra-Subject Variability

In addition to considering inter-subject variability for measurands, it is also of interest to examine intra-personal variability of measurands across multiple exercise sessions. The findings could be applied in the context of comparing non-controlled exercise sessions to a known, good, controlled exercise session.

To obtain data discussed in this section, subjects were instructed to perform the protocol described in the previous section two times, once at a high intensity and then repeat at a much lower intensity. Results are shown for one particular subject in Figure 27 and Figure 28, for isometric and cardio exercise, respectively.

It can be seen that the two different types of exercise have a different impact on physiological signals. For the cardio exercise shown in Figure 27 in the time span from 90s to 240s, there is a correlation between activity level, heart rate, heart rate recovery time and to some extent breathing rate and skin conductance. Once such correlation is established for a particular subject and a particular exercise regiment, if it changes for future sessions, it could signify a change in the subject's physical condition which could in turn trigger re-assessment of subject's treatment program. Day-to-day variations caused by normal daily fluctuations in the subject's level of commitment can be taken into account by examining activity levels – on days when the person feels committed, the overall activity level in VMU will be higher.

For isometric exercise, the conclusion about the impact of exercise is harder to assess. There is no significant difference in the physiological signals in Figure 28 between the high effort and low effort. The activity level spikes are related to posture changes from one exercise to the next and do not indicate level of effort. A closer look at the activity data in-between the spikes may contain data about whether the person is stable or shaking due to higher muscle effort.

5.3.3 Conclusion

The results can be summarized as follows:

- Accelerometer data discriminates between correct and incorrect exercising;
- Heart rate discriminates between effective and ineffective exercising Heart rate feature extraction should be performed after segmentation which is based on accelerometer data

- Breathing rate and skin conductivity require calibration per-subject.

Finally, experience with multiple volunteers shows that the chest strap is not user-friendly, it causes discomfort or has motion artifacts, so its during exercise should be reconsidered. The benefit of the chest strap is that it captures subject's ECG via a dry electrode. The ECG would be used for example to monitor the progression of the disease, such as chronic heart failure (CHF). The method, using heart rate variability (HRV) features will be discussed Chapter 6. As this measurement does not need to be obtained during exercise, the use of the chest strap or other sensor for ECG measurement could be reserved for periods of rest. To monitor heart rate during, or just after exercise, other, less intrusive modalities should be explored, such as a pulse oximeter clip, or video camera, as will be discussed in Chapter 7.

Figure 27: Cardio exercise: low effort (blue) and high effort (red)

Figure 28: Isometric exercise: low effort (blue) and high effort (red)

6 Chapter:

Motion Artifacts and Classification

using HRV Features

Analysis of Heart Rate Variability (HRV) is an active area of research in the engineering and the medical communities, especially into what are significant HRV features and how they can be used, for example to diagnose various diseases. In this chapter, HRV analysis is applied to signals from wearable sensors, whose outputs could exhibit varying levels of noise caused by motion artifacts. The main objective of this chapter is to quantify the impact of imperfect such HRV measurements on a classifier which uses time-domain and spectral –domain HRV features, and propose mitigation techniques which minimize the discarding of data.

6.1 Physionet Databases

The starting point for the analysis are reference databases obtained from Physionet [1], whose collection of physiological databases is widely used in medical and engineering literature and is continuously being expanded.

The databases are based on long term Holter ECG recordings. The exact time instances in the ECG marking heart beats (called beat annotations) are obtained by automated analysis with manual review and correction, and are therefore considered to be highly reliable. These beat markings are used to create RR interval sequence from ECG.

The specific databases used are:

- (1) Normal Sinus Rhythm RR Interval Database (54 healthy subjects). Each subject's recording is approximately 24 hours in duration, resulting in a sequence of about 100,000 RR intervals.
- (2) Congestive Heart Failure RR Interval Database (29 CHF patients). Each subject's recording is approximately 24 hours in duration, resulting in a sequence of about 100,000 RR intervals.
- (3) MIT-BIH Normal Sinus Rhythm Database (NSRDB), consisting of recordings of 18 subjects, sampled at 128Hz;
- (4) BIDMC Congestive Heart Failure Database (CHFDB)., consisting of recordings of 15 subjects, sampled at 250Hz;
- (5) Finally, the MIT-BIH Noise Stress Test Database is used to provide the noise signal used to corrupt the above mentioned ECG recordings. This database was initially created to test arrhythmia detectors [63], and is therefore relevant for any system or algorithm which relies on accurate beat detection such as HRV analysis systems considered in this work. The database consists of 3 half-hour recordings of three types of noise: baseline wander, electrode motion and muscle activity. In this work, we are primarily concerned with electrode motion noise, since it is considered to be the most troublesome type of noise for beat detection.

6.2 Corruption of R-R Intervals

6.2.1 Method

The reference RR intervals come from reliable ECG recordings obtained in a controlled environment using medical grade equipment. Here, we are interested in analyzing the effect of various electrode motion artifacts which could be present in an ECG obtained using consumer-grade ECG acquisition, with dry electrodes, under conditions when the subject is moving around freely. Such artifacts in the original ECG signal would potentially lead to incorrect determination of the exact heart beat instance. The impact of various artifacts on detection of heart rate has been discussed in [66], [22].

The type of noise added to clean signals in [22] comes from an on-line database [67], described in [63]. The noise signals originate from measurements taken from human subjects, with electrodes positioned on the body in such a way that the actual ECG signal is absent and the measured signal is strictly various types of artifacts (such as electrode motion and muscle noise). The level of noise can be varied when combined with a clean signal, as was done in [22] and documented in terms of SNR [dB]. As expected, the lower the SNR, the higher the error in detecting the R peaks and therefore R-R intervals. The amount of error in detecting the R-peak, using conventional threshold based detection, in terms of milliseconds, can be as high as 5-20% of the R-to-R interval for high levels of artifact noise.

Instead of adding noise to the ECG waveform prior to peak detection, the same artifact can be simulated by shifting the reference RR intervals. The main reason for this approach is to be able to extend research to Physionet databases which only include

R-R intervals rather than the full ECG waveform. The shift is modeled as a signed Gaussian variable with a zero mean and an adequate standard deviation, up to 20% of the mean RR interval, to match the most extreme case discussed in [67]. The method of shifting RR intervals is also used in [23]. Although it is not directly possible to correlate the amount of R-R shift to the SNR of the ECG signal, had the ECG signal been corrupted directly, the worst-case R-R shifting scenarios considered in this work correspond to worst-case noisy ECG considered in [67].

To analyze the impact of varying levels of motion artifacts, we vary:

- (1) The amount of time shift,
- (2) The percentage of beats shifted.

The three scenarios analyzed in this work are shown in Figure 29. The red points are the shifted RR intervals. Note that every occurrence in shift affects two RR intervals – if one moves forward in time (interval increases), the next one moves back in time (the next interval decreases).

Figure 29: Three types of RR interval sequence corruption: (a) 10% of beats shift by 5%; (b) 10% of beats shift by 20%; (c) 50% of beats shift by 5%

In Figure 29(a), the corruption is a low noise condition, both in terms of the amount of shift (5% of the mean RR interval), and in terms of the percentage of beats

affected (10%). The other two scenarios, Figure 29(b) and Figure 29(c) are high noise conditions: the corruption in Figure 29(b) has high amount of shift (20% of the mean RR interval) for a low percentage of beats affected (10%), and the corruption in Figure 29 (c) has a low amount of shift (5% of the mean RR interval), but a high percentage of beats affected. The two high noise scenarios have similar noise energy but corrupt the RR interval sequence very differently.

6.2.2 HRV Features and Classification

The feature extraction and classification algorithms closely follow the methods documented in [9]. Standard time domain and spectral HRV measures are used as features based on which the classifier makes decisions. The classifier is a two-stage pruned decision-tree whose design is obtained using the entire normal and CHF dataset (a total of 83 subjects), with leave-one-out validation methodology, where the classifier is trained on the whole dataset excluding one subject, then tested on the one subject, and the process is repeated for all subjects.

The RR intervals for each subject are first segmented into 5-min excerpts, for which HRV features are calculated using HRV toolkit open source software [68]. The software automatically generates the standard time domain and spectral features specified in [69].

Each 5-min excerpt is then classified as “normal” (class 0) or “CHF” (class 1) based on the short-term HRV feature vector for the excerpt. The classifier design is given in [9] – it is a relatively simple decision tree which consists of 9 nodes and uses 4 of the standard HRV features:

- **RMSSD**: The square root of the mean of the sum of the squares of differences between adjacent RR intervals;
- **HF**: Total spectral power in 0.15-0.4 Hz range;
- **TOTPWR**: Total spectral power in 0-0.4Hz range;
- **LF/HF**: Ratio of low to high frequency power (LF is spectral power in the 0.04 – 0.15Hz range).

Each subject is then classified as “normal” or “CHF” based on the percentage of the number of excerpts classified as “normal” or “CHF”. The percentage of excerpts in each class is then compared to a fixed threshold value, also given in [9].

6.2.3 Results

First the impact of motion artifacts on classifier performance is discussed, followed by a deeper analysis of specific HRV measures which play a role in classifier performance.

The classification performance is shown in Table 1, for clean data and the three types of artifacts shown in Figure 29. The performance is described in terms of

- Confusion matrix parameters (TP-True Positive, CHF classified as CHF, FP-False Positive; Normal classified as CHF, TN-True Negative: Normal classified as Normal, FN-False Negative: CHF classified as normal), and
- Accuracy: $(TP+TN)/(TP+FP+TN+FN)$

	TP	FP	TN	FN	Accuracy
Clean	23	0	54	6	92.8%
Corruption (a)	19	0	54	10	87.6%
Corruption (b)	8	0	54	21	74.7%
Corruption (c)	7	0	54	22	73.5%

Table 1: Classification performance

The most visible impact of increased noise artifacts is the increase in FN (CHF patients classified as Normal). On the other hand, noise artifacts have no impact at all on the classification of normal subjects.

In terms of the impact of the type of noise, low noise from Figure 29(a) causes a slight increase in FN and decrease in accuracy, and extreme levels of noise, Figure 29(b) and Figure 29(c), cause reversal of TP and FN relative to the clean data case: while for clean data, most CHF subjects are classified as CHF, with high noise levels, most CHF subjects are classified as Normal. This result is somewhat surprising given than one might expect a poor classifier to roughly classify with 50-50 accuracy both Positives and Negatives – in this case, there is a bias towards the Normal class. This bias can be potentially explained as being a consequence of class imbalance: a weak classifier will tend to over-predict the dominant class (in this case, Normal), to maximize overall accuracy. Further insight into the causes of such bias can be obtained from the view of how individual 5-min excerpts contribute to the final class for each subject. The detailed classification of results in terms of excerpts for each of the 29 CHF subjects is shown in Table 2 .

CHF Database	Clean		Corruption (a)		Corruption (b)		Corruption (c)	
Subject	% FN	C	%FN	C	%FN	C	%FN	C
1	84.78	0	94.93	0	100.00	0	100.00	0
2	66.53	1	68.95	1	79.03	0	84.68	0
3	61.07	1	41.98	1	91.22	0	98.85	0
4	49.19	1	83.06	0	97.18	0	98.39	0
5	6.16	1	3.99	1	14.49	1	16.67	1
6	15.90	1	0.00	1	2.51	1	2.51	1
7	45.78	1	55.02	1	72.18	0	83.06	0
8	56.94	1	38.19	1	86.46	0	97.22	0
9	14.98	1	0.00	1	2.65	1	3.10	1
10	17.88	1	5.84	1	17.52	1	17.52	1
11	99.30	0	100.00	0	100.00	0	100.00	0
12	5.74	1	11.76	1	28.92	1	30.20	1
13	0.00	1	99.31	0	98.61	0	98.96	0
14	30.69	1	20.30	1	55.00	1	78.22	0
15	32.61	1	1.45	1	10.87	1	10.87	1
16	60.61	1	73.86	0	98.48	0	99.24	0
17	98.58	0	97.51	0	99.64	0	99.64	0
18	59.09	1	87.12	0	98.48	0	98.48	0
19	99.24	0	99.62	0	100.00	0	100.00	0
20	50.00	1	35.94	1	69.40	0	70.46	0
21	95.64	0	99.64	0	100.00	0	100.00	0
22	68.36	1	68.00	1	97.80	0	99.64	0
23	58.91	1	60.22	1	93.07	0	98.55	0
24	47.39	1	48.43	1	72.47	0	74.22	0
25	68.04	1	41.55	1	76.26	0	84.47	0
26	25.82	1	4.00	1	21.82	1	28.00	1
27	45.09	1	52.36	1	94.55	0	97.45	0
28	36.73	1	42.55	1	80.00	0	86.55	0
29	98.26	0	94.77	0	96.18	0	96.17	0
TP		23		19		8		7
FN		6		10		21		22

Table 2: Per-subject excerpt classification results for CHF patient database

The “%FN” column represents the percentage of 5-min excerpts which were classified as Normal. Following algorithm described in [9], class “C” of a subject is simply obtained by comparing % of FN excerpts with a fixed threshold: any subject with less than 68% of excerpts classified as Normal is overall classified as CHF.

From Table 2, considering the impact of varying levels of noise artifacts, subjects can be split into three main types:

(1) “Always FN” type (shown in the dark shade in Table 2), subjects which are misclassified regardless of noise. % of FN excerpts is significantly above the threshold, for all noise levels.

(2) “TP $\rightarrow$ FN” type, subject correctly classified for clean data but misclassified for noisy data, shown in Table 2 in a lighter shade. This is the majority of subjects. The number of the subjects of this type directly translates into classification errors due to noise. It is interesting to note that for most such subjects, % of FN excerpts is just below the threshold for clean data (i.e. they could have been considered borderline even with clean data).

(3) “Always TP” type, subjects correctly classified regardless of noise levels (not shaded in Table 2). These types of subjects are not susceptible to noise. % of FN excerpts is low for clean data, and, interestingly, it could rise or fall for noisy data.

For further in-depth analysis of how HRV measures obtained from signals with varying noise levels contribute to the classification results, of interest is to evaluate statistical properties of features used in the classifier, as a function of noise levels. Figure 30 shows the means of four features used in this classifier, and it shows how they vary with noise for the latter two types of subjects: miss-classified in noisy conditions (“TP $\rightarrow$ FN”) subjects in Figure 30 (top) and correctly classified in all conditions (“always TP”) subjects in Figure 30 (bottom). X-axis labels represent the four different noise conditions under study: clean data followed by the corruption in Figure 29(a), (b), (c).

Figure 30: Impact of motion artifacts on key HRV features: misclassified false negative subjects in noisy conditions (top), and correctly classified true positive subjects in all conditions (bottom)

It can be seen from Figure 30 that as noise increases, HF increases, LF/HF decreases, RMSSD increases. Taking into account the original decision tree design from [9], these trends favor the right-hand side of the tree which decides in favor of the Normal class, as corroborated by results in Table 1.

Also from Figure 30, TOTPWR does vary as a function of motion artifacts. However, the absolute level of TOTPWR is the main difference between a “TP→ FN” type subject (misclassified due to noise) and an “Always TP” (never misclassified) type of subject: if the total power of the RR interval sequence is low enough, regardless of other features, the classifier performs well.

In order to improve performance, the classifier originally designed for clean data would have to be adapted especially taking into account HRV feature properties for “TP→ FN” type subjects: those with relatively high power, but LF/HF and HF features unreliable due to noise. Any classifier redesign would have to change the role of the LF/HF and HF and possibly introduce other, more robust, features.

6.3 Corruption of the Raw ECG Signal

6.3.1 Method

The processing stages applied to the Physionet databases are shown in Figure 31. The majority of the processing blocks are implemented using PhysioToolkit functions, which provide a large number of programmable options which affect the performance of the overall system. These options are shown in Figure 31 as “Control” blocks. Analysis parameters are marked as control blocks (noise gain, noise duration, QRS algorithm choice and filtering for HRV features):

- Gain Control: Noise recordings are multiplied by a gain which is a function of desired SNR.
- Duration Control: The noise is only added to the ECG for a specific fraction of the 5-min segment.
- Algorithm Control: Two QRS detection algorithms have been tested: SQRS [13] and WQRS [63]. Each has a threshold parameter which can further be controlled.

- Filtering Control: The HRV feature extraction step is performed using Physionet's HRV Toolkit [68] and its “get_hrv” script which has programmable input parameters such as thresholds for outlier filtering.

Figure 31: System under analysis.

6.3.2 Results - Impact on HRV features

Nine combinations of noise levels (resulting in 20dB, 10dB and 0dB SNR) and noise duration (25%, 50% and 75% of the 5-min segment) were simulated. A sample 5-min ECG segment for select noise corruption scenarios is shown in Figure 32. Reference (clean) ECG (top), 25% corrupted at 10dB (middle), 75% corrupted at 20dB (bottom).

Figure 32: An example of a noise-corrupted 5-min ECG segment: 0dB, 25% (top); 10dB, 25% (middle); 20dB, 75% (bottom)

Various possible noise scenarios have implications on the following quantities which are of interest in subsequent processing:

- NN/RR ratio. This parameter is calculated from the RR interval sequence produced by the QRS detector by taking a ratio of those RR intervals which are considered “Normal sinus rhythm” (falling between 0.4 and 2sec), and all generated RR intervals. It provides a measure of the amount of generated missed or extra beats in an ECG segment and it can be used to discard segments if they are deemed to be unreliable.
- HRV features, both time domain and frequency domain.
- Classification performance.

The impact of noise on these quantities is discussed in the following paragraphs.

6.3.2.1 NN/RR Ratio

NN/RR ratio is used in some classification schemes as a reliability measure based on which a decision is made whether the ECG segment should be included in processing or discarded [9]. The suggested threshold in [68] is 0.8, meaning that if less than 80% of all RR intervals fall in the normal sinus beat range (between 0.4 and 2sec), the entire 5-min segment is discarded from analysis. We have analyzed the behavior of the NN/RR metric for the above mentioned 9 patterns of noise corruption and across two different QRS detection algorithms. The results are summarized in Table 3.

Noise Level	Noise Duration (% of 5-min segment)	NN/RR Ratio	
		SQRS	WQRS
No Noise	0	0.96	0.96
20dB	25	0.95	0.96
	50	0.94	0.96
	75	0.89	0.96
10dB	25	0.88	0.96
	50	0.70	0.93
	75	0.35	0.91
0dB	25	0.68	0.50
	50	0.42	0.25
	75	0.20	0.10

Table 3 : Fraction of RR intervals considered normal sinus beats

From Table 3, if NN/RR is to be used as a reliability metric with the default threshold of 0.8 and SQRS as the beat-detection algorithm, the segment would be discarded for 5 out of 9 noise scenarios (10dB 50% and 75%, and all three 0dB cases).

The situation is somewhat better for WQRS algorithm. Again, using NN/RR with the default threshold of 0.8, the segment would be discarded only 3 out of 9 noise corruption patterns would (all at 0dB). However, we note that even for WQRS, an entire segment is discarded in the situation where only 25% of data is corrupted by high noise although the vast majority of the segment (75%) is noise-free.

In both cases, using a static threshold of 0.8 for NN/RR seems to be a fairly drastic measure and could result in a significant and unnecessary loss of data, considering that a large portion of the 5-min segment is clean.

6.3.2.2 HRV Feature Contamination

Rather than using NN/RR ratio as a proxy for reliability of a segment, we now focus directly on the HRV features which are used by a subsequent classification algorithm, as shown in Figure 31. We focus on the 4 HRV features used for CHF classification [9]:

- RMSSD: The square root of the mean of the sum of the squares of differences between adjacent RR intervals.
- TOTPWR: Total spectral power in 0-0.4Hz range.
- HF: Total spectral power in 0.15-0.4 Hz range.
- LF/HF: Ratio of low to high frequency power (LF is spectral power in the 0.04 – 0.15Hz range).

We note that the first one is a time-domain feature and the remaining three are frequency-domain features.

For the above mentioned HRV featured, we analyze the correlation between NN/RR ratio and actual contamination of the HRV feature. Firstly, all segments are divided into three groups: high NN/RR (greater than 0.8), medium NN/RR (between 0.5 and 0.8) and low NN/RR (less than 0.5). Subsequently, the mean value of each HRV feature is computed for each of the three groups. The results are summarized in Table 4, for WQRS detector.

Feature	NN/RR		
	> 0.8	0.5 - 0.8	< 0.5
RMSSD	42.4	34.6	42.2
TOTPWR	5448	8713	14747
HF	828	1131	1665
LF/HF	2.11	2.21	5.81

Table 4: HRV feature mean value as a function of NN/RR ratio

As can be seen from the results in Table 4, there is no direct correlation between HRV feature means and NN/RR ratios that applies for all features. For example, LF/HF, an important feature used for classification, only increases by about 5% when NN/RR drops below 0.8 but stays above 0.5. This means that, for this particular feature, the data segment does not have to be discarded for segments where NN/RR is below 0.8.

Similarly, RMMSD value does not have a clear trend as NN/RR drops and as such, with proper split thresholds in the classifier, this feature could be used even if NN/RR is below 0.8.

For the remaining two features, TOTPWR and HF, both frequency domain features, there is a high correlation between the level of contamination of their averages and NN/RR ratio, so it is clear that they cannot be used as NN/RR drops below a certain threshold.

6.4 Traditional Reliability Metric

To observe the impact of NN/RR ratios and HRV features originating from noisy ECG signal on the final outcome of the system (classifier output), the CHF detection algorithm from [9] was applied to the HRV features obtained from ECG signals with 50% noise corruption at 0dB, 10dB and 20dB. The results validate conclusions reached in Section 6.2 where a less sophisticated ECG corruption scheme was used.

The classifier tends to classify CHF patients as “Normal” resulting in a high percentage of false negatives. This is due to the tendency of noise corrupted signal to have increased TOTPWR and decreased LF/HF.

If, however, only segments with $NN/RR > 0.8$ are considered for classification, the performance of the classifier does not suffer. However, only about 15% of the segments satisfy this criterion, therefore drastically reducing the number of datasets available for training and testing the classifier.

6.5 Proposed Reliability Metric

We conclude that the rule to discard any ECG segments for which less than 80% of detected heart beats are considered normal, results in unnecessary loss of data which would still result in uncompromised system performance at classifier level.

These results suggest that better results could be achieved through using an adaptive system which would , based on the noise corruption profile for a particular segment, adjust the following system parameters:

- Threshold used in the QRS detector.
- NN/RR threshold above which segments are not discarded.
- HRV feature values which split the classifier decision tree.

These suggestions are feasible assuming proper training data is available to develop the algorithms and the classifier. In this work, a clean signal was artificially corrupted, which is one of the possible approaches. Ideally, the training data should come from a long term ECG recording obtained using dry, wearable electrodes instead of the traditional gel electrodes used in Holter monitors, and would be available for public use in a database, similar to Physionet [1]. Availability of such data would not only facilitate work on robust algorithms under realistic signal acquisition conditions, but would also enable development of noise corruptions models. Such models could be used to corrupt any “clean” ECG databases, and exponentially increase availability of realistic ECG data for development of robust algorithms.

7 Chapter:

Motion Artifact and Video-Based

Heart Rate Measurement

For a video-based heart rate measurement system to be feasible, it needs to work well in realistic scenarios where the subject does not sit completely still in front of a camera. Motion artifacts, if not taken into account when designing the system, yield inaccurate results and potentially create false alarms. The objective of this chapter is to analyze impact of motion artifacts on accuracy of HR detection in a video of subject's face, and subsequently propose mitigation techniques to recover accuracy.

7.1 Method

A series of short, 30-sec videos was recorded under idealistic (still, i.e. without deliberate motion) and realistic (with motion) scenarios as shown in Figure 33. The motion scenarios to be considered are those of large scale (such as a head turn), as well as small scale (such as head nodding and camera shaking). Different lighting scenarios to be considered are for example natural (day) light, LED light and incandescent light,

illuminating the subject in a symmetrical (from the ceiling) or non-symmetrical (from the side) manner.

Figure 33: A realistic motion scenario

The signal processing algorithm follows for the most part that described in [70] and available as Matlab source code from [71], for non-commercial research purposes. The front-end signal processing is part of the framework known as Eulerian Video Magnification, which has been successfully used in extracting heart rate from thermal video [72] and wrist [73], as well as to extract other physiological signals such as breathing [74] and pulse-transit time [75].

The input to the system is a sequence of video frames. Each pixel is subjected to first spatial and then temporal filtering. We use as reference the same processing blocks as used by the authors of [70] for their face video, also available at [71]. Specifically, we do not use any processing designed to amplify motion – our focus is on color only.

Spatial filtering is performed using Gaussian pyramid decomposition. The processing is done in the NTSC (YIQ) color space rather than RGB, where Y represents luminance and I and Q represent chrominance components. This process essentially consists by blurring and down-sampling the image for a prescribed number of levels. The

blurring is achieved using a customizable filter design and it operates independently in each the x- and the y-direction. At each level, the image is down-sampled by a factor of 2 in each direction. Therefore, the programmable parameters in this stage are (1) blurring (spatial) filter and (2) number of levels. The default parameters used for the face video in [71] are a binomial filter of size 5, and 4 levels of down-sampling.

Temporal filtering is performed using an ideal bandpass filter. This process is also performed in the NTSC (YIQ) color space. For each pixel and each color channel, an FFT is performed and a mask is applied which zeros out any FFT components outside of the frequency band of interest. The signal is then returned into the time domain using an IFFT. The default frequency band used for the face video in [71] is [50, 60] beats per minute (bpm).

Following filtering, heart rate is extracted from video by choosing a pixel location (coordinates), plotting its value over the duration of the video or a portion of the video, and locating its peaks in either frequency domain (if the heart rate is constant for the duration of the observed video segment), or time domain, by finding the time instances associated with each individual peak.

7.2 Extracted BVP Signal without Motion

Initially, we have validated our Matlab environment using the reference source code, data files, and results given in [71], using “face.mp4” video file. The resolution of the reference video is 528x592 (width x height), which is significantly lower than our videos which are 1920x1080 (width x height).

As a next step, we have processed our videos varying parameters of both the spatial and the temporal filter.

For spatial filtering, we considered two different values for each of the two spatial filter parameters: Gaussian pyramid levels (4 or 5), and spatial pooling (binomial filter of size 5 or size 20). The effect of each of the parameters can be seen in Figure 34.

Figure 34: Effect of spatial filter parameters

For temporal filtering, instead of using a very narrow filter of 50bpm to 60bpm used for processing the reference signal “face.mp4”, we use a bandpass filter from 50bpm to 120bpm, which covers a wider range of possible heart rates. The waveforms used to extract the heart rate are pixel intensities over the duration of video. Sample waveforms are illustrated in Figure 35 and Figure 36. Figure 35 represents a skin pixel (from a forehead location) which indeed contains the heart rate signal, and Figure 36 represents a background pixel which does not contain the heart rate signal. In both figures, top and middle are waveforms after spatial and after temporal filtering, respectively, and bottom

represents the FFT of the signal after temporal processing, Temporal filtering is done using a 50-120bpm filter. In both figures, the luminance component is used as pixel intensity.

Figure 35: Skin pixel: time-domain BVP waveforms

Figure 36: Non-skin (background) pixel: time-domain BVP waveforms

As can be seen from Figure 35 , the periodicity of the heart rate signal is visible even in prior to temporal filtering, albeit with some drift and low-frequency noise. After

filtering, using local maxima in Figure 35(b), it would be possible to obtain the heart rate signal with relatively little jitter. In the spectral domain, there is a clear peak frequency corresponding to the heart rate at ~60bpm, since the heart rate was constant for the duration of the video and the FFT. Additional spectral content exists in the region from 50-120 bpm coinciding with the lower and upper cutoff frequencies for the temporal filter, which is due to illumination changes and slight motion.

From Figure 36 (non-skin pixel), even though the signal magnitude is comparable to skin pixel, there are no clear peaks in the FFT. However, once filtered using a band-pass filter targeting the normal heart range, one could believe to have “found” the heart rate. This is even more pronounced if a very narrow filter is used, like done for the reference video “face.mp4”. Even non-skin pixels may appear to contain the “heart rate” signal. Therefore, it is important that the region of interest containing skin pixels is properly determined prior to feeding the pixels into the signal processing algorithm.

To validate our results, the pixel intensity waveform for a forehead pixel has been passed through a generic peak detection algorithm, from which IBI (inter-beat intervals) have been computed. The pixel intensity waveform is shown in Figure 37, , along with the detected peaks (in red circles), for two subjects. The average extracted heart rate over the 20 sec video is obtained from the average inter-beat interval, and compared against pulse oximeter readings available from the recorded video. The extracted average heart rate agrees with the pulse oximeter reading with 1-2bpm accuracy. For Subject 1, the average extracted HR is 72.1bpm while the actual heart rate measured by the pulse

oximeter is 72.2bpm. For Subject 2, the average extracted HR = 65.8bpm while the actual heart rate is 65.2 bpm.

Figure 37: Peak detection for subject 1 (top) and subject 2 (bottom)

7.3 Extracted BVP Signal with Motion

We now turn to the scenario where there is additional, deliberate motion introduced in the video, in order to study the resilience of this algorithm to motion artifacts. We are primarily concerned with small-scale motion which may not be easily trackable using computer vision algorithms.

Realistic examples of such motion include camera shaking or up/down or left/right head nodding. Such a scenario would not add any additional complications if the additional motion-related noise signal is in a frequency band different than the target frequency band, as shown in Figure 38 . The high-frequency noise signal visible after spatial but before temporal filtering in Figure 38(top) is due to camera shaking and it falls in a frequency band outside of the target 50-120 bpm band, and is rejected after temporal filtering in Figure 38(bottom).

Figure 38: Impact of deliberate motion in time domain.

However, if such motion falls in the target frequency band, temporal filtering is not helpful. We postulate that the parameters of spatial filtering have little impact on the “still” video but have an impact of video which includes small motion, and vary, independently and combined, both the size of the spatial filter ($n=5$ and $n=20$) and the number of levels of the Gaussian pyramid ($l=4$ and $l=5$).

The results are shown in Figure 39 for a video where the subject is nodding head up-and-down, in the time-domain after spatial and temporal filtering.

Figure 39: BVP waveform for varying spatial filter parameters (up-down motion).

In Figure 39 (subject moving head up and down), the motion in the video is labeled “Period of Motion”, but it is also visible on the time-domain waveform itself. The impact of the spatial filtering parameters can be observed. Considering the time domain waveform for spatial filter parameters $n=5$ and $l=4$, it is not possible to extract an accurate heart rate from local maxima, due to the presence of “dips” in the peak. However, for spatial filter parameters $n=20$ and $l=5$ (i.e. increasing both filter size and number of levels in Gaussian pyramid), the local maxima are more obvious. Similar observations have been made in spectral domain — only for spatial filter parameters $n=20$ and $l=5$ do we see a presence of one clear frequency peak and other spatial filter cases also have significant other frequency components.

Following this coarse spectral analysis of the extracted waveform and eliminating everything but the largest spatial filter settings, we now perform a more in depth analysis relationship between the extracted waveform and motion in the video. For this purpose, we use a facial marker to track a particular location on the face. This allows us to track the magnitude of the motion by tracking X and Y coordinates of this marker over time, as shown in Figure 40.

Figure 40: BVP waveform (left-right motion) and tracked X/Y coordinates

It can be seen that the motion results in increased amplitude in the extracted waveform, in spite of large spatial pooling. One explanation is that this is due to changes in illumination. Further experiment should be conducted to compare left-right motion with up-down motion without changing lighting conditions. Depending on the location of the light source, one direction of the motion should be less susceptible to motion artifacts than the other.

Further processing of the waveform shown in Figure 40(top) using peak locator algorithm yields the extracted average heart rate of ~ 58 bpm, which is lower than the reference heart rate obtained from the pulse oximeter (~ 65 bpm). The discrepancy comes from underestimating the heart rate during period of motion, as can be seen from the tachogram in Figure 41 (motion occurs approximately between Beat #10 and Beat #20). Again it can be seen that the extracted heart rate is dominated by motion rather than the blood pulse signal.

Figure 41: Tachogram from video with motion between beat #10 and beat #20.

In addition to spatial and temporal filtering, other techniques may be used to compensate for different types of motion. For example, for camera shaking, video stabilization could be used. To compensate for the larger motion of the subject, tracking algorithms could be used. However, these methods will still require additional techniques to be put in place due to motion-induced changes in light conditions.

7.4 Estimated Heart Rate with Motion

In the previous section, it has been established that the presence of motion may overwhelm the extracted PPG waveform, as shown in Figure 40, and that this potentially has effects on the estimated heart rate, from examining the tachogram in the time domain, Figure 41. In this section, the analysis is taken further, in the following three directions:

- (1) The impact of motion on the heart rate is quantified, where the heart rate is estimated based on spectral domain properties, and updated periodically, rather than considering an average heart rate estimate over the entire video segment

(2) Impact of varying illumination conditions is quantified: considered are two cases with motion, one where motion produced a high amount of shadow and another one where the amount of shadow is lower.

(3) Mitigation techniques are proposed.

The starting point for heart rate extraction is the previously discussed PPG waveform obtained by spatially averaging a patch of skin, tracked under motion using a facial marker. The waveform is shown in Figure 42 for a side-to-side head motion, and in Figure 43 for an up and down head motion. The spatial filter is a simple average of pixel intensities in the green component of the RGB color space, over a ROI which consists of 50x80 forehead pixels. As the head moves, the ROI is tracked so it is always the same skin patch that falls in the ROI.

As can be seen in the figures, the two types of motion both overwhelm the PPG signal, but to a varying degree: in the case of side-to-side motion, the impact is greater due to the direction of the light (daylight coming from a side window). In the case of up and down motion, the impact of motion is somewhat lower in amplitude, so that the PPG waveform is still recognizable during motion in Figure 43.

Figure 42: BVP waveform (side to side motion)

Figure 43: BVP waveform (up-and-down motion)

The common way of monitoring the heart rate over the course of the video is to perform a short-term FFT in a sliding window manner and detect the FFT peak that falls in the valid heart rate range. In Figure 44, the resulting short term FFT over a 8-second

sliding window for waveforms from Figure 43 is shown after 11sec and after 20 sec. The heart rate estimated from Figure 44(top) , based on the highest spectral peak, is 72bpm, which agrees well with the reference heart rate reading obtained with the finger clip sensor (73bpm), as it is to be expected during periods without motion. However, in Figure 44 (bottom), which represents the short term FFT obtained from the motion-corrupted segment of the waveform in Figure 42 (12-20seconds), there is no spectral peak at the expected 72bpm – instead, the actual heart rate is underestimated at 63.28bpm, and, in addition, there higher spectral peak is actual the one corresponding to the periodicity of the motion (at around 45bpm).

Figure 44: Short term FFT (up-and-down motion), before and during motion

For the waveform shown in Figure 42, where the motion signal appears to completely overwhelm the heart rate signal, the short-term FFT obtained during the period of the motion (11-19 seconds) is shown in Figure 45 does not feature any distinguishable spectral peaks in the expected heart range, and the conclusion is that the combination of motion and illumination artifacts makes the heart rate signal unrecoverable during this period, using the particular waveform from Figure 42.

Figure 45: Short term FFT (side to side motion), during motion

7.5 Proposed Mitigation of Motion Artifacts

The scenarios described in the previous section do demonstrate the ability of unfavorable illumination conditions, combined with subject motion, to completely overwhelm the PPG signal of interest. In this section, a less drastic scenario is considered where periods of stillness are interleaved by periods of large motion. The goal is to automatically distinguish between the two types of segments, and (1) in the case of still segments, provide a reliable measurement and (2) in the case of segments with motion, determine if the illumination conditions are favorable, allowing for a reliable measurement, or not, requiring the measurement to be discarded. The algorithm for heart rate extraction remains the basic one discussed in previous sections, but it is made robust in realistic scenarios by including environmental conditions (motion and illumination).

7.5.1 Method

The concept is shown in Figure 46 . The robustness of heart rate measurement from video is achieved by (1) tracking facial features to determine the amount of motion and always filter the same skin patch; (2) determining illumination changes with motion by continuously performing pixel averaging over the skin patch, and (3) combining the previous two approaches to (a) adjust filtering parameters as necessary and (b) determining whether the measurement is reliable or should be discarded. This processing should be performed on a beat-by-beat basis and in real-time.

Figure 46: Heart rate extraction with illumination and motion context

Two video sequences are analyzed with the same type of motion, which can be generally categorized as large scale transitional motion. Each video consists of 9 segments: (1) center position, (2) moving to right, (3) right position, (4) moving to center, (5) center position, (6) moving to left, (7) left position, (8) moving to center, and (9)

center position. The two videos record the same type of motion, but under different illumination conditions – one with directional daylight, achieved by recording the subject indoor, next to a window, and another one with uniform light, achieved by recording the subject outdoor but away from direct sunlight. Representative frames from both videos are shown in Figure 47. The red box shows the skin patch which is used for heart rate measurement.

Figure 47 Video frames showing segments with center, right and left position, for indoor, directional light (top) and outdoor, uniform light (bottom).

7.5.2 Spatial and Temporal Filtering

The extracted waveforms are shown in Figure 48 for the video with directional light, and Figure 49 for the video with uniform light. Waveforms (a) represent spatially filtered pixel intensity and waveforms (b) spatially and temporally filtered pixel intensity, where the pixel of interest is in the center of the red box show in Figure 47. The pixel of interest is chosen to be slightly above the middle point between two eyebrows, since this area is clearly visible throughout the subject motion. The pixel of interest is being tracked

as the subject is moving. For reference, the Y and X coordinates of this pixel are shown in Figure 48 and Figure 49 (c) and (d), respectively. Since the movement is side to side, only changes in the X coordinate are significant. The x-axis represents the frame number, or pixel sample – the 1800 samples represent a 60-sec video at a sampling rate of 30 frames per second. The impact of different illumination conditions is visible from the differences in Figure 48 and Figure 49 (a), which represent spatially averaged pixel intensity of the pixel of interest. As expected, the intensity varies more significantly in the directional light condition. The intensity is the lowest during the “right position” (samples ~700-950 and Figure 47 top-middle), where the subject is turned away from the window, and is the highest during the “left position” (samples ~1300-1500 and Figure 47 top-right)), where the subject is turned toward the window. Between these two extremes, the intensity differs by almost a factor of 4. In the case of uniform illumination, the intensity varies less than 15% (Figure 49(a)).

Once the temporal filter is applied, since it is a band-pass filter in the frequency range of interest for heart rate, the “DC”-component is removed and, as a consequence, the spatially and temporally filtered versions of the waveform no longer contain the information about the illumination. It is this waveform however that represents the physiological BVP waveform and that will be used to extract the heart rate.

Figure 48: Spatially and temporally filtered BVP waveform, and motion and illumination context, for video with directional light.

Figure 49: Spatially and temporally filtered BVP waveform, and motion and illumination context, for video with uniform light

7.5.3 Heart Rate Extraction

To extract the heart rate, peak detection is performed, followed by computation of inter-beat-intervals, on the BVP waveforms in Figure 48(b) and Figure 49(b). For comparison, the same operations are also performed on the PPG waveform obtained from the pulse oximetry front end shown in Figure 22. The output of the peak detector algorithm, for a segment of the PPG and BVP waveforms, are shown in Figure 50.

Figure 50: Peak detector algorithm applied to a 10-sec segment of the reference (PPG) waveform (left) and video BVP waveform (right)

As seen from Figure 50, although there is some baseline drift, the locations of peaks, each resulting from a heartbeat, are very clear, which results in a reliable reference measurement for inter-beat intervals. In contrast, the BVP waveform segment obtained from video, representing roughly the same segment, contains some areas of ambiguity (around segment samples 120-150), coinciding with the sudden motion of the subject.

The overall accuracy of the heart rate measured from video, as compared to the reference signal, is initially assessed using the average heart rate, which is computed as the inverse of the average inter-beat interval for the duration of the video. The average heart rate, in beats per minute (BPM) for the two videos, is shown in Table 5.

	Video (BVP) waveform	Reference (PPG) waveform	% error
Video with uniform illumination	74.5	77.1	3.3%
Video with directional illumination	86.1	80.9	6.5%

Table 5: Average heart rate from video vs reference (BPM)

From this data, it appears that the more variable illumination condition result in higher error in average HR. However, since the videos contain periods of motion interleaved with period of stillness, and the ambiguous peak detection coincides with

motion, the heart rate measurement should be analyzed on a beat-by-beat basis, focusing on the intervals affected by motion. A tool which can be used to this end is a tachogram, a graph which shows the inter-beat interval or its inverse (instantaneous heart rate) as a function of the beat number.

Figure 51: Tachogram of (a) BVP waveform from video and (b) reference waveform for video with uniform illumination

Figure 52: Tachogram of (a) BVP waveform from video and (b) reference waveform for video with directional illumination

Tachograms for video-based and reference heart rate, for videos with uniform and directional illumination, are shown in Figure 51 and Figure 52, respectively. The figures provide an insight into variability of the measured video-based instantaneous heart rate compared to the reference, however, even though the original waveforms were synchronized, due to the cumulative nature of the time-axis of the graph, the two measurements cannot be compared on a beat by beat basis. Still, valuable insight can be obtained by comparing the measured heart rate ranges and deviations.

The obvious conclusion from the two figures is that the video with directional illumination has more artifacts than the video with uniform illumination. In Figure 51, most of the motion artifacts are exhibited as dips in the tachogram, resulting in a few beats below the expected range but not completely outside of the possible range of human heart beats. In Figure 52, one can observe 6 unusual heart rate values, all related to transitional motion. In 5 out of 6 cases, the heart rate is higher than expected, stemming from low IBIs caused by extraneous detected peaks. These are caused by the transitional motion, exacerbated by changing illumination condition, like seen in Figure 48 (a), and possibly by imperfect marker tracking during transition. The cumulative effect is that the detected heart rate is not reliable during these periods and should be discarded.

7.5.4 Mitigation of motion artifacts

The unfavorable illumination conditions reduce the tolerance for errors in skin patch tracking and result in unreliable measurement of heart rate from video. The proposed mitigating action is to detect these unfavorable changes in illumination conditions resulting from motion, as are visible in the spatially averaged skin patch

intensity and shown in Figure 48 (a), correlated with coordinates of tracked facial markers, and discard a segment of heart rate measurements. On the other hand, if motion is detected from tracking facial markers, but there are no accompanying changes in illumination conditions, like seen in most of the transitions in Figure 49 (a), then data does not need to be discarded.

In the general case, thresholds would need to be defined to differentiate between these two extreme cases, either based on absolute values or gradients. In the specific case of the videos analyzed in this work, by locating the step transitions in Figure 48(a) and Figure 49(a) and eliminating two affected heartbeats, we have obtained results shown in Table 6.

	Video (BVP) waveform	Reference (PPG) waveform	% error	discarded/total beats
Video with uniform illumination	76.7	77.1	0.5	2/59
Video with directional illumination	80.7	80.9	0.24	12/92

Table 6: Average heart rate from video vs reference (BPM) after discarding data

By discarding only a small number of heart beats associated with motion transitions accompanied by illumination transitions, the accuracy of heart rate measurement is improved.

8 Chapter:

Conclusions

8.1 Thesis Conclusions

This thesis investigated heart rate and heart rate variability estimation under realistic data acquisition conditions, where the subject is moving relative to the sensors. The sensors which were investigated include wearable (dry or textile) electrodes and video cameras, both suitable for use cases where minimally intrusive, long term monitoring of patients is of interest. Some examples of such scenarios are baby monitoring in the neonatal intensive care unit, driver monitoring, and remote healthcare.

It was expected that the methods developed for ideal conditions will not perform as well in realistic conditions. The objective that was set for this thesis was to control the parameters of the degraded conditions, quantify the degradation, identify the mechanism in which the existing methods fail, in order to propose mitigation techniques. This was accomplished for both scenarios investigated: heart rate variability estimation using wearable electrodes, and heart rate estimation using a video camera.

For the heart rate variability estimation for the purpose of further feature extraction and classification, the primary problem identified is that the method designed for ideal condition would result in more than 85% of the data (i.e. 5-min ECG segments) being discarded as unreliable. This leaves only a small portion of the available data set for both training and testing of the classifier, which impacts the accuracy of classification. It was shown in the thesis that this criterion for discarding the data is too harsh and that some of the data segments declared as unreliable are in fact usable and do not degrade the performance. The proposed solution is to modify the reliability detection metric by making it adaptive, based on noise characterization, on a segment-by-segment basis. In addition, the reliability metric (NN/RR ratio), since it is susceptible to noise, should be featured less prominently in the classification algorithm, i.e. moved down in the decision tree structure.

For the heart rate estimation from video, it was found that, under unfavorable illumination condition, the subject motion can result in variations in illumination which completely overwhelm the signal of interest, namely the blood volume pulse. This happens even if the facial features are perfectly tracked and the same patch of skin is used for processing for every single frame. A solution could be simply to discard any portion of the video which shows significant motion of facial features, based on the motion vectors detected by the feature tracking algorithm. This could again result in unnecessary loss of data because (1) not all motion occurs under unfavorable illumination conditions, and (2) the threshold for how much motion is acceptable is also a function of illumination conditions. The proposed solution is to monitor both subject motion relative to the

camera, as well as the change in illumination to detect whether the estimated heart rate is reliable or not.

8.2 Future Research

Future research into detection and mitigation of motion artifacts in minimally invasive heart rate and heart rate variability monitoring would further refine the proposed techniques as follows:

- More parameters of motion artifacts could be defined and controlled for, in order to cover a wide range of scenarios. For example, ECG noise in wearable sensors should be also a function of the pressure at which the sensor is applied to the skin, sweating, etc. Similarly, for video recording of subject's face, all six directions of rigid motion should be investigated. The size of motion should be controlled, as well as ambient light directionality and intensity. Ideally, motion conditions and artifacts would be standardized for each sensor modality. This would eliminate many claims of "motion robust" algorithms, which often only focus on a limited set of adverse conditions.
- Mitigation techniques should be developed by using the least possible number of tunable parameters. Many algorithms rely on multiple thresholds, heuristics, exceptions, which inevitably leads to overfitting to a particular dataset and is bound to fail in other data sets.
- The techniques should also be developed with real-time monitoring in mind, since timely detection of anomalies could potentially save lives. Therefore, the developed techniques need to be implementable in real-time manner, with

minimal latency. This favors methods such as filtering over methods which require large windows of data to be available, such as independent component analysis or neural networks.

The work in the above mentioned direction, although challenging, would enable adoption of modern sensors in the field of healthcare and, as a result, reduce the amount of handling of each patient required by a healthcare professional, and alleviate resourcing issues that exist in today's healthcare system.

References

- [1] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet:Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. 215-220, 2000.
- [2] J. Nikolic-Popovic and R. Goubran, "Measuring heart rate, breathing rate and skin conductance during exercise," in *2011 IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA)*, Bari, 2011, pp. 507-511.
- [3] J. Nikolic-Popovic and R. Goubran, "Impact of motion artifacts on Heart Rate Variability measurements and classification performance," in *2013 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, Gatineau, 2013, pp. 156-159.
- [4] J. Nikolic-Popovic and R. Goubran, "Towards increased usability of noisy ECG signals in HRV-based classifiers," in *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Lisboa, 2014, pp.1-4.
- [5] J. Nikolic-Popovic and R. Goubran, "Impact of motion artifacts on video-based non-intrusive heart rate measurement," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Benevento, 2016, pp.1-4.
- [6] A. Searle and L. Kirkup, "A direct comparison of wet, dry and insulating bioelectric recording electrodes," *Physiological measurement*, vol. 21, no. 2, p. 271, 2000.
- [7] B. Taji and e. al., "Impact of skin–electrode interface on electrocardiogram

- measurements using conductive textile electrodes," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 6, pp. 1412-1422, 2014.
- [8] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability Standards of measurement, physiological interpretation, and clinical use.," *European Heart Journal*, pp. 354-381, 1996.
- [9] L. Pecchia, P. Melillo, M. Sansone and M. Bracale, "Discrimination Power of Short-Term Heart Rate Variability Measures for CHF Assessment," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 1, pp. 40-46, January 2011.
- [10] S. R. Wiese and e. al., "Electrocardiographic motion artifact versus electrode impedance," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 1, pp. 136-139, January 2005.
- [11] E. Sazonov and M. R. Neuman, *Wearable Sensors: Fundamentals, Implementation and Applications*, Elsevier, 2014.
- [12] A. Cömert, M. Honkala and J. Hyttinen, "Effect of pressure and padding on motion artifact of textile electrodes," *BioMedical Engineering OnLine*, vol. 12, no. 26, 2013.
- [13] G. Friesen, T. Jannett, M. Jadallah, S. Yates, S. Quint and H. Nagle, "A comparison of the noise sensitivity of nine QRS detection algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 1, pp. 85-98, January 1990.
- [14] W. Zong, G. Moody and D. Jiang, "A robust open-source algorithm to detect onset and duration of QRS complexes," *Computers in Cardiology*, pp. 737-740, September 2003.

- [15] S. W. Yoon, S. D. Min, Y. H. Yun, S. Lee and M. Lee, "Adaptive motion artifacts reduction using 3-axis accelerometer in e-textile ECG measurement system," *Journal of Medical Systems*, vol. Nov, no. 2, pp. 101-106, 2008.
- [16] W.-C. Lee, Y.-S. O. Yang, T.-C. Ke, C.-S. Wei, Lee and H.-C., "Adaptive Reduction of Motion Artifact in a Portable ECG System," in *Proc. IEEE Conf. Sensors 2010*, 2010, pp. 704-707.
- [17] T. Alkhidir, A. Sluzek and M. K. Yapici, "Simple method for adaptive filtering of motion artifacts in E-textile wearable ECG sensors," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, 2015, pp. 3807-3810.
- [18] H. Kim, S. Kim, N. V. Helleputte, T. Berset, D. Geng, I. Romero, J. Penders, C. V. Hoof and R. F. Yazicioglu, "Motion Artifact Removal using Cascade Adaptive Filtering for Ambulatory ECG Monitoring System," in *Proc. IEEE Biomedical Circuits & Systems Conf.*, 2012, pp. 160-163.
- [19] S. Tian, J. Han, J. Yang, L. Zhou and X. Zeng, "Motion artifact removal based on ICA for ambulatory ECG monitoring," in *2015 IEEE 11th International Conference on ASIC (ASICON)*, Chengdu, 2015, pp.1-4.
- [20] O. Postolache, P. Girao, J. Mendes, E. Pinheiro and G. Postolache, "Physiological Parameters Measurement Based on Wheelchair Embedded Sensors and Advanced Signal Processing," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 10, pp. 2564-2574, Oct 2010.
- [21] U. Wiklund, M. Karlsson, N. Ostlund, L. Berglin, K. Lindecrantz, S. Karlsson and L.

- Sandsjö, "Adaptive spatio-temporal filtering of disturbed ECGs: a multi-channel approach to heartbeat detection in smart clothing," *Medical & Biological Engineering & Computing*, no. 6, pp. 515-23, June 2007.
- [22] M. Nakano, T. Konishi, S. Izumi, H. Kawaguchi and M. Yoshimoto, "Instantaneous Heart Rate detection using short-time autocorrelation for wearable healthcare systems," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 6703-6706.
- [23] L. Citi, E. Brown and R. Barbieri, "A point process local likelihood algorithm for robust and automated heart beat detection and correction," *Computing in Cardiology*, pp. 293-296, September 2011.
- [24] X. He, R. A. Goubran and X. P. Liu, "Signal enhancement of wearable ECG monitoring sensors based on Ensemble Empirical Mode Decomposition," in *2011 IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA)*, Bari, 2011, pp. 433-436.
- [25] P. X. Quesnel, A. D. C. Chan and H. Yang, "Signal quality and false myocardial ischemia alarms in ambulatory electrocardiograms," in *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Lisboa, 2014, pp. 1-5.
- [26] P. X. Quesnel, A. D. C. Chan and H. Yang, "Real-time biosignal quality analysis of ambulatory ECG for detection of myocardial ischemia," in *2013 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, Gatinéau, 2013, pp.1-5.

- [27] J. Gang, A. D. C. Chan and A. Cuhadar, "Heart rate reliability for the Smart Rollator," in *Instrumentation and Measurement Technology Conference (I2MTC)*, Austin, 2019, pp. 1044-1047.
- [28] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas*, vol. 28, no. 3, pp. pp. 1-39, 2007.
- [29] G. W. J. Clarke, A. D. C. Chan and A. Adler, "Effects of motion artifact on the blood oxygen saturation estimate in pulse oximetry," in *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Lisboa, 2014, pp. 1-4.
- [30] F. P. Wieringa, F. Mastik and A. F. v. d. Steen, "Contactless multiple wavelength photoplethysmographic imaging: a first step toward "SpO2 camera" technology," *Ann. Biomed. Eng.*, vol. 33, no. 8, pp. 1034-1041, Aug 2005.
- [31] K. Humphreys, T. Ward and C. Markham, "Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry," *Rev. Sci. Instrum.*, vol. 78, no. 4, p. 044304, Apr 2007.
- [32] G. Ogedegbe, "White-coat Effect: Unraveling Its Mechanisms," *Am. J. Hypertens.*, vol. 21, no. 2, p. 135, Feb 2008.
- [33] C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Med. Eng. Phys.*, vol. 29, no. 8, pp. 853-857, Aug 2007.
- [34] W. Verkruijsse, L. O. Svaasand and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt.Express*, vol. 16, no. 26, pp. 21434-21445, Dec 2008.

- [35] M. Poh, D. McDuff and R. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, no. 10, pp. 10762-74, May 2010.
- [36] A. M. Rodríguez and J. R. Castro, "Pulse rate variability analysis by video using face detection and tracking algorithms," in *Proc. 37th Annu. Int. Conf. IEEE EMBS*, Milan, 2015, pp.5696-5699.
- [37] C. Park and H. j. Choi, "Motion artifact reduction in PPG signals from face: Face tracking & stochastic state space modeling approach," in *Proc. 36th Annu. Int. Conf. IEEE EMBS*, Chicago, IL, 2014, pp.3280-3283.
- [38] C. Huang, X. Yang and K. T. T. Cheng, "Accurate and efficient pulse measurement from facial videos on smartphones," in *Proc. IEEE Winter Conf. Applications Comp. Vision*, Lake Placid, NY, 2016, pp.1-8.
- [39] D. N. Tran, H. Lee and C. Kim, "A robust real time system for remote heart rate measurement via camera," in *IEEE International Conference on Multimedia and Expo (ICME)*, Turin, 2015, pp.1-6.
- [40] L. K. Mestha, S. Kyal, B. Xu, L. E. Lewis and V. Kumar, "Towards continuous monitoring of pulse rate in neonatal intensive care unit with a webcam," in *Proc. 36th Annu. Int. Conf. IEEE EMBS*, Chicago, IL, 2014, pp. 3817-3820.
- [41] R. Y. Huang and L. R. Dung, "A motion-robust contactless photoplethysmography using chrominance and adaptive filtering," in *Biomedical Circuits and Systems Conference (BioCAS)*, Atlanta, GA, 2015, pp. 1-4.
- [42] F. Bousefsaf, C. Maaoui and A. Pruski, "Remote assessment of the heart rate

- variability to detect mental stress," in *Int. Conf. Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, Venice, Italy, 2013, p.348-351.
- [43] L. Feng, L. M. Po, X. Xu and Y. Li, "Motion artifacts suppression for remote imaging photoplethysmography," in *19th International Conference on Digital Signal Processing*, Hong Kong, 2014, pp.18-23.
- [44] L. Shan and M. Yu, "Video-based heart rate measurement using head motion tracking and ICA," in *6th International Congress on Image and Signal Processing (CISP)*, Hangzhou, 2013, pp.160-164.
- [45] J. Gunther, N. Ruben and T. Moon, "Model-based (passive) heart rate estimation using remote video recording of moving human subjects illuminated by ambient light," in *IEEE International Conference on Image Processing (ICIP)*, Quebec City, QC, 2015, pp. 2870-2874.
- [46] Z. Guo, Z. Wang and Z. Shen, "Physiological parameter monitoring of drivers based on video data and independent vector analysis," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Florence, 2014, pp. 4374-4378.
- [47] J. Shi and C. Tomasi, "Good Features to Track," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp.593–600.
- [48] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. CVPR*, 2001, pp.511-518.
- [49] M. Köstinger and e. al., "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011, pp. 2144-2151.

- [50] P. Werner, A. Al-Hamadi, S. Walter, S. Gruss and H. C. Traue, "Automatic heart rate estimation from painful faces," in *IEEE ICIP*, Paris, 2014, pp. 1947-1951.
- [51] M. C. Li and Y. H. Lin, "A real-time non-contact pulse rate detector based on smartphone," in *International Symposium on Next-Generation Electronics (ISNE)*, Taipei, 2015, pp. 1-3.
- [52] L. Feng, L. M. Po, X. Xu, Y. Li and R. Ma, "Motion-Resistant Remote Imaging Photoplethysmography Based on the Optical Properties of Skin," *IEEE Transactions on Circuits and Systems in Video Technologies*, vol. 25, no. 5, pp. 879-891, May 2015.
- [53] Y. P. Yu, R. Paramesran and C. L. Lim, "Video based heart rate estimation under different light illumination intensities," in *Proc. Intel. Signal Proc. and Comm. Syst.*, Kuching, 2014, pp.216-221.
- [54] K. Jack, *Video demystified: a handbook for the digital engineer*, Elsevier, 2011.
- [55] A. Lam and Y. Kuno, "Robust Heart Rate Measurement from Video Using Select Random Patches," in *Proc. IEEE Int. Conf. Comp. Vision*, Santiago, 2015, pp.3640-3648.
- [56] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70-73, 1967.
- [57] M. Grundmann, V. Kwatra and I. Essa, "Auto-directed video stabilization with robust 11 optimal camera paths," in *Proc. IEEE Conf. CVPR*, Colorado Springs, 2011, pp. 225-232.

- [58] J. R. Estepp, E. B. Blackford and C. M. Meier, "Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, San Diego, 2014, pp. 1462-1469.
- [59] A. Sikdar, S. K. Behera and D. P. Dogra, "Computer Vision Guided Human Pulse Rate Estimation: A Review," *IEEE Reviews in Biomedical Engineering*, 2016.
- [60] J. C. X. Li, G. Zhao and M. Pietikäinen, "Remote Heart Rate Measurement from Face Videos under Realistic Situations," in *Proc. IEEE Conf. CVPR*, Columbus, OH, 2014, pp. 4264-4271.
- [61] "BioHarness 3 – Wireless Professional Heart Rate Monitor & Physiological Monitor," [Online]. Available: <https://www.zephyranywhere.com/products/bioharness-3>. [Accessed July 2016].
- [62] [Online]. Available: <http://www.affectiva.com/q-sensor/>. [Accessed April 2011].
- [63] G. Moody, W. Muldrow and R. Mark, "A noise stress test for arrhythmia detectors," *Computers in Cardiology*, pp. 381-384, 1984.
- [64] "Wikipedia Sony A7," [Online]. Available: https://en.wikipedia.org/wiki/Sony_A7. [Accessed 10 July 2016].
- [65] "AFE4490 EVM," Texas Instruments, [Online]. Available: <http://www.ti.com/tool/afe4490spo2evm>. [Accessed 10 July 2016].
- [66] E. Richard and A. Chan, "Design of a gel-less two-electrode ECG monitor," in *2010 IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA)*, 2010, pp. 92-96.

- [67] "The MIT-BIH Noise Stress Test Database," [Online]. Available: <http://physionet.org/physiobank/database/nstdb/>.
- [68] "Heart Rate Variability Analysis with the HRV Toolkit," [Online]. Available: <http://physionet.org/tutorials/hrv-toolkit/>.
- [69] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss and P. J. Schwartz, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354-381, march 1996.
- [70] H. Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM T. Graphic.*, vol. 31, no. 4, p. 65, Jul 2012.
- [71] [Online]. Available: <http://people.csail.mit.edu/mrub/evm/#code>. [Accessed March 2016].
- [72] S. L. Bennett, R. Goubran and F. Knoefel, "Adaptive eulerian video magnification methods to extract heart rate from thermal video," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Benevento, 2016, pp.1-5.
- [73] X. He, R. A. Goubran and X. P. Liu, "Wrist pulse measurement and analysis using Eulerian video magnification," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Las Vegas, NV, 2016, pp. 41-44.
- [74] S. L. Bennett, R. Goubran and F. Knoefel, "The detection of breathing behavior using Eulerian-enhanced thermal video," in *2015 37th Annual International*

Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 7474-7477.

- [75] X. He, R. A. Goubran and X. P. Liu, "Using Eulerian video magnification framework to measure pulse transit time," in *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*, Lisboa, 2014, pp. 1-4.