SIGNAL PROCESSING METHODS FOR THE ESTIMATION OF PHYSIOLOGICAL PARAMETERS FROM OPTICAL RECORDINGS

by

EIRINI CHRISTINAKI

BSc, Technological Educational Institute of Crete, 2013

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

DEPARTMENT OF INFORMATICS ENGINEERING

SCHOOL OF ENGINEERING

TECHNOLOGICAL EDUCATIONAL INSTITUTE OF CRETE

2016

Approved by:

Professor Manolis Tsiknakis

Statement of Originality

The work contained in this thesis has not been previously submitted for a degree or diploma at any other higher education institution or any other purpose. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except as specified in references, acknowledgments or in footnotes. I certify that the intellectual content of this thesis is the product of my own work and all the assistance received in preparing this thesis and sources have been acknowledged.

Eirini Christinaki

Acknowledgements

I would like to express the deepest appreciation and my utmost gratitude to my supervisor Prof. Manolis Tsiknakis. His exceptional scientific knowledge, experience and creative thinking have been a source of inspiration and motivation for me while his patience and encouragement were determinant for the accomplishment of this thesis. In addition, I would like to express my sincere acknowledgement in the continuous support, constructive suggestions and guidance of my thesis advisor Franco Chiarugi.

I am also heartily thankful to my fellow and lab mates in Biomedical Informatics & eHealth Laboratory (BMI Lab) of the Technological Educational Institute of Crete (TEI of Crete) and Computational BioMedicine Laboratory (CBML) of ICS-FORTH for the fruitful discussions. Finally, I would like to acknowledge the EU funded project with the acronym SEMEOTICONS (FP7-ICT-2013-10-611516) for the *SEMEOTICONS reference dataset*, which I have extensively used in most of my experiments for this thesis.

Abstract

Monitoring of physiological signals of an individual via remote and contactless means is an important scientific challenge, whose resolution will enable the development of novel, non-intrusive mHealth and wellness-management systems and services. Measurement of vital signs, such as heart rate (HR), breathing rate (BR), heart rate variability (HRV), blood oxygen saturation level (SpO2) and blood pressure (BP), is important both in clinical environments and at home due to the fundamental role of such signs in the diagnosis of health conditions and monitoring of well-being. Currently, the gold standard techniques for measuring these signs are based on sensors with skin contact, but this is not convenient in all scenarios. Therefore, nonintrusive methods for accurate vital sign measurement have become important, particularly in the context of pervasive services for health or wellness management.

During the last decade, researchers have begun to explore the use of simple and low-cost methods for non-contact vital sign extraction. One of the approaches explored for contactless detection and estimation of these signs is based on photoplethysmography (PPG), a noninvasive optical technique for measuring blood flow in tissue, the use of video imaging and ambient light. Exploiting this method has a high potential benefit since there is a need for low-cost physiological monitoring solutions that are easy to use, accurate and can be utilized outside from conventional clinical settings. However, PPG systems using digital camera and ambient light, face several issues and challenges.

This thesis describes the development and the optimization of the necessary methods in order to measure a set of physiological signs such as HR, HRV and respiratory rate from facial videos using PPG and ambient light. It attempts to identify different factors that need to be taken into account when constructing an optimal computational pipeline for the estimation of these signs. It explores different signal processing techniques to overcome issues that emerge during the whole procedure. The experimental validation and analysis using a large dataset, proves that highly accurate measurement of the order of 97-99% for the HR in real life conditions is possible. It also demonstrates positive indications on the feasibility of the HRV analysis and respiratory rate estimation since we observed that many types of artefacts had major effects on computed HRV indices which allow us to provide evidence only for the feasibility and not for the reliability of our method.

Περίληψη

Η παρακολούθηση των βιοσημάτων (physiological signals) ενός ατόμου εξ' αποστάσεως και χωρίς επαφή αποτελεί μια σημαντική επιστημονική πρόκληση της οποίας η επίλυση θα επιτρέψει την ανάπτυξη καινοτόμων, μη παρεμβατικών υπηρεσιών ηλεκτρονικής υγείας (mHealth) και υπηρεσιών διαχείρισης της ευεξίας (wellness management).

Η μέτρηση ζωτικών σημείων όπως ο *καρδιακός παλμός*, ο *ρυθμός αναπνοής*, η *μεταβλητότητα του καρδιακού ρυθμού*, τα *επίπεδα κορεσμού του οξυγόνου στο αίμα (SpO2)* και η *αρτηριακή πίεση*, είναι σημαντική τόσο σε κλινικά περιβάλλοντα όσο και σε άλλα περιβάλλοντα διαβίωσης (π.χ. στο σπίτι) εξαιτίας του θεμελιώδη ρόλου που τα εν λόγω σήματα έχουν στην ιατρική διαγνωστική διαδικασία. Σήμερα, ο κανόνας αναφορικά με τον τρόπο μέτρησης απαιτεί τη χρήση αισθητήρων που έρχονται σε επαφή με το δέρμα (contact sensors). Αυτός όμως ο τρόπος δεν είναι βολικός σε όλα τα σενάρια χρήσης ούτε συχνά επιθυμητός. Ως εκ τούτου, οι μη παρεμβατικές μέθοδοι για την ακριβή μέτρηση βιοσημάτων και ζωτικών σημείων ενός ατόμου είναι σημαντικές, ιδίως στο πλαίσιο ανάπτυξης διάχυτων υπηρεσιών (pervasive eHealth services) ηλεκτρονικής υγείας για τη διαχείριση της υγείας ή της ευεξίας.

Κατά την τελευταία δεκαετία, έχει ξεκινήσει η διερεύνηση της χρήσης απλών και χαμηλού κόστους μεθόδων για την ανέπαφη μέτρηση των ζωτικών σημείων. Μία από τις υπό διερεύνηση προσεγγίσεις για την ανέπαφη ανίχνευση και εκτίμηση αυτών των σημείων βασίζεται στη φωτοπληθυσμογραφία, μια μη επεμβατική οπτική τεχνική για τη μέτρηση της ροής του αίματος στους ιστούς με τη χρήση βίντεο-απεικόνισης και το φως του περιβάλλοντος. Η αξιοποίηση αυτής της μεθόδου είναι πολύ σημαντική, δεδομένου ότι υπάρχει η ανάγκη για χαμηλού κόστους λύσεις για τη παρακολούθηση ζωτικών σημείων ενός ατόμου που να είναι εύκολες στη χρήση, ακριβείς και να μπορούν να χρησιμοποιηθούν και εκτός από τα συμβατικά κλινικά περιβάλλοντα. Ωστόσο, τα φωτοπληθυσμογραφικά συστήματα που χρησιμοποιούν καταγραφές από βίντεο και το φως του περιβάλλοντος, αντιμετωπίζουν πολλά προβλήματα και προκλήσεις.

Αυτή η διατριβή περιγράφει την ανάπτυξη και τη βελτιστοποίηση των απαραίτητων μεθόδων προκειμένου να μετρήσουμε μια σειρά από φυσιολογικές ενδείξεις όπως ο *καρδιακός παλμός*, η *μεταβλητότητα του καρδιακού ρυθμού* και ο *ρυθμός αναπνοής* από βίντεο του προσώπου χρησιμοποιώντας τη φωτοπληθυσμογραφία και το φως του περιβάλλοντος. Προσδιορίζει μια σειρά από παράγοντες που πρέπει να ληφθούν υπόψη κατά την κατασκευή μίας βέλτιστης

υπολογιστικής διαδικασίας για την εκτίμηση των εν λόγω σημείων και διερευνά διαφορετικές τεχνικές επεξεργασίας εικόνας και σήματος για την αντιμετώπιση θεμάτων που προκύπτουν κατά τη διάρκεια της όλης διαδικασίας.

Η εργασία επίσης εστιάζει με έμφαση στην πειραματική επικύρωση των μεθόδων που αναπτύχθηκαν αλλά και της συνολικά προτεινόμενης διαδικασίας. Προς την κατεύθυνση αυτή χρησιμοποιήθηκε ένα μεγάλο σε όγκο σύνολο δεδομένων. Η πειραματική αυτή επιβεβαίωση αποδεικνύει ότι εξαιρετικά ακριβής μετρήσεις της τάξης του 97-99% για τον *καρδιακό παλμό* σε πραγματικές συνθήκες ζωής είναι δυνατόν να επιτευχθούν. Η εργασία επίσης παρέχει θετικές ενδείξεις σχετικά με τη δυνατότητα της υπολογιστικής εκτίμησης της *μεταβλητότητας του καρδιακού ρυθμού* και της εκτίμησης του *ρυθμού αναπνοής*.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ANS	Autonomic Nervous System
AVI	Audio Video Interleave
BP	Blood Pressure
bpm	Beats Per Minute
BR	Breathing Rate
BSS	Blind Source Separation
BVP	Blood Volume Pulse
cICA	constrained Independent Component Analysis
CMOS	Complementary metal–oxide–semiconductor
ECG	Electrocardiogram
EEG	Electroencephalogram
FDA	Food and Drug Administration
FFT	Fast Fourier Transformation
fMRI	functional Magnetic Resonance Imaging
fps	Frames Per Second
HCI	Human Computer Interaction
HF	High Frequency
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Inter-Beat Interval
ICA	Independent Component Analysis
JADE	Joint Approximate Diagonalization of Eigen-matrices
KLT	Kanade-Lucas-Tomasi
kNN	k-Nearest Neighbour
LDA	Linear Discriminant Analysis
LE	Laplacian Eigenmap
LF	Low Frequency
LLE	Locally Linear Embedding
LPP	Linearity Preserving Projection
MJPG	Motion JPEG (Joint Photographic Experts Group) video compression format
ML	Machine Learning
MP4	MPEG-4 Part 14 (Moving Picture Expert Group) digital multimedia format
MVU	Maximum Variance Unfolding
NLMS	Normalized Least Mean Square

NMF	Non-negative Matrix Factorization
NN intervals	Normal-to-Normal intervals
PCA	Principal Component Analysis
PNS	Parasympathetic Nervous System
PPG	Photoplethysmography
PRV	Pulse Rate Variability
PSD	Power Spectral Density
QRS	Combination of the main three graphical deflections seen on a typical electrocardiogram
RGB	Red-Green-Blue
RMSE	Root Mean Squared Error
ROI	Region Of Interest
RR intervals	Beat-to-Beat intervals
SCA	Sparse Component Analysis
SNR	Signal-to-Noise Ratio
SNS	Sympathetic Nervous System
SP approach	Smoothness Priors approach
SpO2	Blood oxygen saturation level
SVD	Singular Value Decomposition
TFR	Time Frequency Representation
TLV	Time Lapse Video
ULF	Ultra-Low Frequency
VLF	Very Low Frequency

1 Introduction

The human face is a precious revealer of key information about the health status of individuals, in the form of a combination of physical signs and expressive features. It has always been considered to be a mirror of emotions and mood. During the Pythagorean and Socratic period, physiognomy used to consider the face as a descriptor of human character. One of the first scholars particularly concerned with physiognomy was Leonardo da Vinci, who published some fundamental observations in 1651[1]. Later, René Descartes carefully investigated ocular saccades, face gesture, color variations and breath rhythms, in order to establish their relationship with the emotional and pathological status of an individual [2]. In the modern age, Charles Darwin contributed to the study of face and expressions with a relevant book [3].

The quantification of facial signs into meaningful measures and computational descriptors can be exploited for the automatic assessment of cardiovascular disease risk and the evaluation of a wellness index [4]. In this context, a patient's physiological signal monitoring via nonintrusive, remote and noncontact means is an important objective. The nonintrusive measurement of vital signs, such as heart rate (HR), breathing rate (BR), heart rate variability (HRV), blood oxygen saturation level (SpO2) and blood pressure (BP), is important both in clinical environments and at home due to their fundamental role in the diagnosis of health conditions and wellness monitoring.

Currently, the gold standard techniques to measure vital signs are based on contact sensors such as chest straps, electrocardiographic (ECG) electrodes, finger pulse oximeters and BP cuffs (Figure 1). The most widely used and gold standard tools and techniques involve these obtrusive devices attached to the body. In some cases, the contact between skin and electrodes or adhesive gel patches or chest stripes may cause skin irritation, stress and generally discomfort. Furthermore, contact-based sensors are not convenient in all scenarios. Ethical issues, such as stigmatization, can also be introduced by visible or not sufficiently hidden contact-based sensors.

Currently, the proposed solutions for contactless measurement of vital signs, such as HR and respiratory rate, include laser Doppler [5], microwave Doppler radar [6], thermal imaging [7] or even the use of the camera of a mobile phone [8]. The noncontact assessment of HRV represents a greater challenge due to a number of limitations and specific requirements, and few attempts have been made to respond to the challenge so far [9]. One of the approaches explored for non-contact detection and measurement of these vital signs is based on the analysis of video images

using techniques that are able to provide estimates of the vital signs either directly or indirectly (i.e. by calculating relative measures). A variety of techniques are employed, such as photoplethysmography (PPG), Eulerian video magnification [10] and time lapse video (TLV) [11].

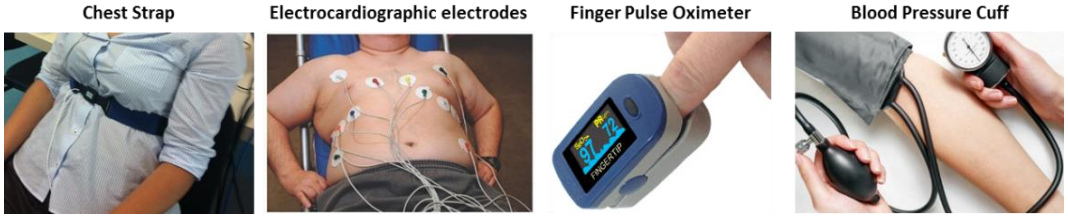


Figure 1: *Example of chest strap, ECG electrodes*[1]*, finger pulse oximeter*[2] *and BP cuff*[3]

Non-contact methods for vital sign measurement using a camera have already shown initial promising results [12], [13]. In particular, PPG is an optical method used to measure blood volume change in body extremities such as the face, finger and earlobe. The foundation of the approach is based on the fact that as the heart pumps blood, the volume of blood in the arteries and capillaries varies by a small amount in synchronization with the cardiac cycle. Such variations in blood volume in the arteries and capillaries underneath the skin induce subtle skin color changes. The goal of a color camera-based vital sign monitoring system is to estimate the PPG waveform, which is proportional to these changes in skin color. Vital signs such as HR, HRV, SpO2 and BR can be derived from a well-acquired PPG waveform.

However, PPG systems using digital cameras and ambient light face several issues and challenges. The most important challenge is that these systems are susceptible to motion artifacts due to an individual's movement in front of the camera and artefacts induced by changes in light conditions [13].

In the context of the present thesis, we will focus on the PPG principle, a non-invasive technique, and the development and evaluation of appropriate signal processing methods for the estimation of physiological parameters from optical recordings of the face. The fundamental principle of the approach is the fact that a PPG signal is extracted from video recordings of the human face, using ambient or diffuse white light; subsequently this signal is appropriately processed for the detection and measurement of a set of semeiotics signs of the face.

---

[1] http://what-when-how.com/paramedic-care/diagnostic-ecgthe-12-lead-clinical-essentials-paramedic-care-part-7/

[2] http://fingerpulseox.com/oximetry/pulse-oximetry-is-a-technology-used-to-measure-the-oxygen-level-of-body/

[3] http://www.theweeklyhaul.com/tag/medicine/

1.1 Scope and Objective

The present thesis aims to explore advanced signal processing methods for the estimation of physiological parameters from optical recordings. The scope of this thesis is the development, implementation and optimization of the necessary methods and algorithms to easily and accurately extract HR and respiration rate parameters as well as performing HRV analysis from video-based PPG signals. The key contribution of the present work relates to the identification and validation of an optimal image processing pipeline for the accurate contactless measurement of these parameters from facial videos. The main objective relates to the establishment of a robust computational pipeline and its validation in terms of measurement accuracy, when compared to classical ECG measurements, using a large dataset.

In specific, the present thesis attempts to quantify the different factors influencing the estimation of physiological parameters from optical recordings and explore signal processing techniques to overcome these issues. It attempts to identify the elements of an optimal measurement computational pipeline and its appropriate configuration for estimating the most suitable parameters at each step of this pipeline. It considers the errors introduced at each stage of the process, it evaluates the whole process and estimates the overall accuracy of the computational pipeline and its methods. Finally, it validates the measurement accuracy of the PPG based contactless monitoring of physiological parameters from optical recording as compared to classical ECG measurements.

1.2 Thesis Overview

The thesis is organized in eight chapters, as follows:

- **Chapter 1:** the current chapter, includes a brief introduction of the topic, highlights the scope and the main objectives of this dissertation, as well as the publications that were produced in the context of the present Master's thesis.
- **Chapter 2:** provides the theoretical background for understanding the cardiovascular system, HR, respiration rate and HRV. It also includes the analysis and measurement techniques based on ECG and PPG.
- **Chapter 3:** presents an elaborate review of related work in an effort to identify, review and analyze the findings of all related studies published during the last six years.
- **Chapter 4:** describes the datasets used for our exploratory and main validation studies.

- **Chapter 5:** explains in detail the technical implementation. In this chapter, the exploratory study and the proposed signal processing pipeline are detailed. It also makes reference to the ECG signal processing for the extraction of the relevant information.
- **Chapter 6:** presents the validation results for the estimation of HR, HRV and respiratory rate.
- **Chapter 7:** includes the discussion over the results.
- **Chapter 8:** is the last chapter of this thesis and includes the conclusion and the possible directions to follow in future work.

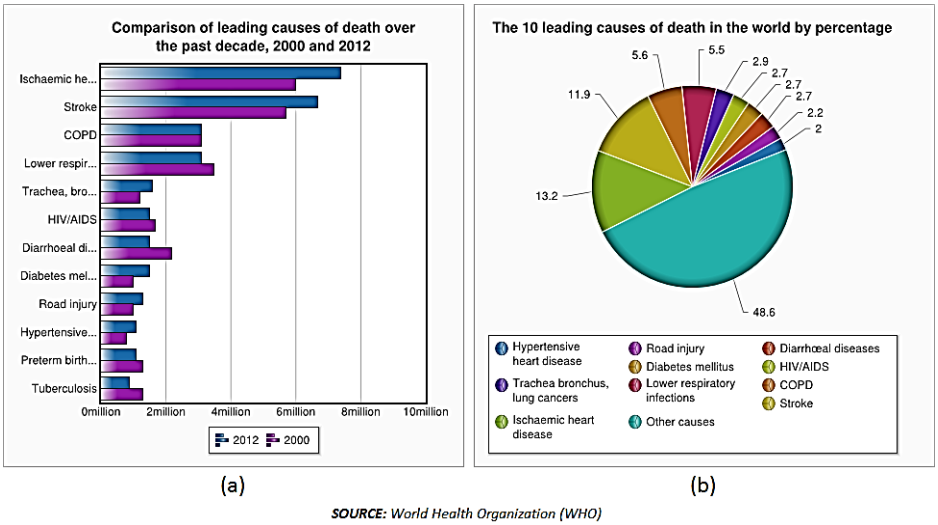
1.3 Publications

Two different publications have been produced during the Master's degree, reporting on findings obtained from this dissertation. The first paper entitled "*Comparison of blind source separation algorithms for optical heart rate monitoring*", was accepted as a full paper (peer-reviewed) and was presented at the 4th International Conference on Wireless Mobile Communication and Healthcare (MobiHealth 15) [14], while the second one entitled "*Contactless Heart Rate Measurement from Facial Videos*" has been submitted as a journal paper in the IOPScience Physiological Measurement Journal[4]. The work reported in this thesis has also contributed important material in the scientific publication "*Pedro Henriquez, et al., Mirror on the wall... an unobtrusive intelligent multisensory mirror for well-being status self-assessment and visualization"*, which is currently under peer review.

[4] http://iopscience.iop.org/journal/0967-3334

2 Background

Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels which are the leading cause of deaths worldwide [15]. The World Health Organization (WHO) estimates that CVDs are responsible for the death of 17.5 million people in 2012, which accounts for 3 in every 10 deaths. Of these, about 6.7 million people died of stroke and 7.4 million from ischemic heart disease in 2012 (Figure 2a) which was about 13.2 percent of all deaths (Figure 2b).



SOURCE: *World Health Organization (WHO)*

Figure 2: *(a) Leading causes of death worldwide (b) Global distribution*

According to the WHO, CVD is caused by disorders of the heart and blood vessels, and includes coronary heart disease (heart attacks), cerebrovascular disease (stroke), raised BP (hypertension), peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. Four out of five CVD deaths are due to heart attacks and strokes. Individuals at risk of CVD may demonstrate raised BP, glucose, and lipids as well as overweight and obesity. The major causes of cardiovascular disease are tobacco use, physical inactivity, an unhealthy diet and harmful use of alcohol.

2.1 Cardiovascular System

The cardiovascular system is designed to transport oxygen and nutrients to the cells of the body and remove carbon dioxide and metabolic waste products from the body. The control of the cardiovascular system is accomplished in part by the Autonomic Nervous System (ANS) which plays an important role in the regulation of the physiological processes of the human organism during normal and pathological conditions. The ANS is generally conceived to have two major

branches, the sympathetic system associated with energy mobilization and the parasympathetic system associated with vegetative and restorative functions. These divisions have opposite effects and usually work reciprocally. This extrinsic control mechanism can dominate intrinsic regulatory mechanisms that modulate HR and cardiac output. The sympathetic branch of the ANS, when activated, increases HR, whereas activation of the parasympathetic branch reduces HR. For example, under conditions of physical or emotional activity like stress sympathetic tends to increase HR and under conditions of rest the parasympathetic reduces HR. Normally, the activity of these branches is in dynamic balance.

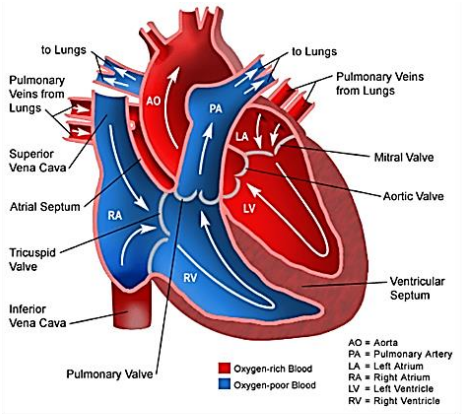


Image taken from:
https://www.urmc.rochester.edu/Encyclopedia/Content.aspx?ContentTypeID=90&ContentID=P02362

Figure 3: *The normal heart*

The heart is the muscular organ of the circulatory system that constantly pumps blood throughout the body (Figure 3). In order to pump blood through the body, the heart is connected to the vascular system of the body. The normal, healthy, adult heart is about the size of a closed fist and sits in the thorax in the center of the chest, between right and left lungs, rotated about 30 degrees. It is designed as a pump with four chambers called right atrium (RA), right ventricle (RV), left atrium (LA), and left ventricle (LV). The two atria are the smaller, upper chambers of the heart and contract simultaneously. The two ventricles are the larger, lower chambers of the heart and also contract simultaneously. The LV is generally about twice as thick as the RV because it needs to generate enough force to push blood through the entire body while the RV only needs to generate enough force to push blood through the lungs.

The right side of the heart pumps blood to the lungs through the pulmonary artery (PA), pulmonary capillaries, and then returns blood to the LA through the pulmonary veins (PV). The

left side of the heart pumps blood to the rest of the body through the aorta, arteries, arterioles, systemic capillaries, and then returns blood to the RA through the venules and great veins.

Each pump or beat of the heart consists of two parts or phases, the diastole and the systole. During diastole the ventricles are filling and the atria contract. Then during systole, the ventricles contract while the atria are relaxed and filling. The heart also has four valves. The tricuspid valve is between the RA and RV. The pulmonary valve is between the RV and the PA. The mitral valve is between the LA and the LV and the aortic valve is between the LV and the aorta. The valves, under normal conditions, insure that blood only flows in one direction in the heart.

2.2 Heart Rate

The heart's pumping action is regulated by an electrical conduction system that coordinates the contraction of the various chambers of the heart. The electrical conduction system of the heart is responsible for initiating each heartbeat and determining the heart's rate and rhythm. The HR, also known as pulse rate, is a measure of the cardiac activity. It is defined as the number of times the heart beats in a given time interval and it is measured by the number of contractions of the heart per unit of time, typically beats per minute (bpm). The pulse can vary according to the body's physical needs, including the need to absorb oxygen and excrete carbon dioxide. Changes can also be provoked by physical exercise, sleep, anxiety, stress, illness, food intake, and drugs. The normal HR varies from person to person with the normal resting HR ranging from 60-100 bpm, but in infants the rate may be as high as 120 bpm. The normal results for resting HR, according to the National Institutes of Health (NIH)[5], are presented in Table 1.

Table 1: *Normal results for resting HR*

Newborns 0 to 1 month old	70 to 190 bpm
Infants 1 to 11 months old	80 to 160 bpm
Children 1 to 2 years old	80 to 130 bpm
Children 3 to 4 years old	80 to 120 bpm
Children 5 to 6 years old	75 to 115 bpm
Children 7 to 9 years old	70 to 110 bpm
Children 10 years and older, and adults (including seniors)	60 to 100 bpm
Well-trained athletes	40 to 60 bpm

Measuring the pulse gives important information about health. Any change from normal HR can indicate a health problem. Resting HR that is continually high (tachycardia) or below the

[5] https://www.nlm.nih.gov/medlineplus/ency/article/003399.htm

normal values (bradycardia) may mean a problem. A pulse that is very firm (bounding pulse) and lasts for more than a few minutes or an irregular pulse can also indicate a health condition. Moreover, a pulse that is hard to locate may mean blockages in the artery. These blockages are common in people with diabetes or hardening of the artery from high cholesterol. The pulse rate can also help determine if the patient's heart is pumping in emergency situations. Pulse measurement has other uses as well, e.g. during or immediately after exercise, the pulse rate gives information about your fitness level and health.

A heartbeat is a two-part pumping action (Figure 4) that takes about a second. As blood collects in the upper chambers (the right and left atria), the heart's natural pacemaker, the sinoatrial (SA) node, sends out an electrical signal that causes the atria to contract. This contraction pushes blood through the tricuspid and mitral valves into the resting lower chambers (the right and left ventricles). This pumping phase (the longer of the two) is called diastole. The second part of the pumping phase begins when the ventricles are full of blood. The electrical signals from the SA node travel along a pathway of cells to the ventricles, causing them to contract. This is called systole. As the tricuspid and mitral valves shut tight to prevent a back flow of blood, the pulmonary and aortic valves are pushed open. While blood is pushed from the right ventricle into the lungs to pick up oxygen, oxygen-rich blood flows from the left ventricle to the heart and other parts of the body. After blood moves into the pulmonary artery and the aorta, the ventricles relax, and the pulmonary and aortic valves close. The lower pressure in the ventricles causes the tricuspid and mitral valves to open, and the cycle begins again. This series of contractions is repeated over and over again, increasing during times of exertion and decreasing while you are at rest.

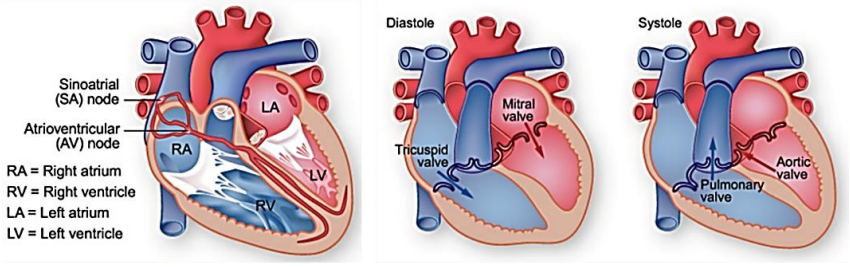


Image taken from:
http://www.texasheart.org/HIC/Anatomy/systole.cfm

Figure 4: *The two parts of a heartbeat called diastole and systole*

The SA node generates impulses about 100-120 times per minute at rest. However in healthy individual resting HR would never be that high. This is due to continuous control of the

ANS over the output of SA node activity, which net regulatory effect gives real HR. Therefore, the heart does not work alone. The brain tracks the conditions around such as climate, stress, and level of physical activity and adjusts the cardiovascular system to meet those needs.

2.3 Heart Rate Variability

The time interval between consecutive heartbeats, even under rest, is constantly and slightly changing. The estimations of this naturally occurring beat-to-beat variations provide specific, well-defined, quantitative indicators of cardiac autonomic function. According to the European Society of Cardiology and the North American Society of Pacing and Electrophysiology there is a significant relationship between the ANS and the cardiovascular mortality [16]. A commonly used measure of the ANS is the variation over time of the period between consecutive heartbeats called HRV, as it represents one of the most promising quantitative markers of the autonomic activity [17]. This physiological phenomenon indicates the heart's ability to adapt to changing circumstances by detecting and quickly responding to unpredictable stimuli. It is a measure that reflects the balance between sympathetic and parasympathetic (vagal) input to the heart. The HRV analysis enables the assessment of the overall cardiac health and the state of the ANS responsible for regulating the cardiac activity. Therefore, the HRV constitutes an important factor for the cardio-metabolic risk assessment.

The assessment of HRV provides quantitative information about the modulation of HR by the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). Interactions of SNS and PNS using HRV have been well studied and their importance established for a number of cardiac diseases including myocardial infarction [18] and patients at risk of sudden cardiac death [19]. HRV analysis is currently also used to assess the dynamics describing the sympathetic and parasympathetic modulation of HR in various other clinical conditions like diabetes [20] and the sleep apnea syndrome [21]. Furthermore, its analysis has valuable relevance in psychophysiology and, in particular, in studies of physical, emotional, and mental function, as well as in psychological and psychiatric disorders [22], [23]. Thus, changes in the HRV patterns provide a sensible and advanced indicator of health status. Higher HRV is a signal of good adaptation and characterizes a healthy person with efficient autonomic mechanisms, while lower HRV is, frequently, an indicator of abnormal and insufficient adaptation of the ANS, provoking poor patient's physiological function [24]. Such diminished variability suggests decreased vagal

input and increased sympathetic input, which have been associated with an increase in malignant ventricular arrhythmias and mortality [25]. Patients with major depressive disorder also exhibit decreased HRV and the severity of the impairment correlates with the clinical severity of depression [26].

The various measurable parameters used to estimate HRV are usually calculated by analyzing the time series of normal sinus beat-to-beat (RR) intervals from the ECG signal. It mainly derives from ECG data analysis, since the ECG is the most widely used gold standard technique for monitoring the cardiac activity, in order to detect various abnormalities and conditions associated with the heart. The HRV signal is usually obtained from short-term (usually 5 minutes) or long-term (usually 24 hours, at least 18 hours) ECG recordings. It can be analyzed by different methods, which are usually categorized as time domain methods, frequency domain methods, and methods based on the nonlinear dynamics of HR [27]. The time and frequency domain measures of HRV have been most commonly used. The time domain measures are easier to calculate but tend to provide less detailed information than the frequency domain approaches. In cases where these conventional methods fail to identify nonlinear patterns, the nonlinear methods are preferred for the analysis, as they can represent the non-stationary characteristics of HRV. Overall, in the analysis of stationary short-term recordings, more experience and theoretical knowledge exist on physiological interpretation of the frequency domain measures compared with the time domain measures derived from the same recordings.

Table 2: *Selected time-domain measures of HRV*

Statistical Measures		
Variable	**Units**	**Description**
SDNN	ms	Standard deviation of all NN intervals
SDANN	ms	Standard deviation of the averages of NN intervals in all 5-minute segments of the entire recording
RMSSD	ms	The square root of the mean of the sum of the squares of differences between adjacent NN intervals
SDNN index	ms	Mean of the standard deviations of all NN intervals for all 5-minute segments of the entire recording
SDSD	ms	Standard deviation of differences between adjacent NN intervals
NN50 count	-	Number of pairs of adjacent NN intervals differing by > 50 ms in the entire recording
pNN50	%	NN50 count divided by the total number of all NN intervals

All HRV parameters are calculated on normal-to-normal (NN) inter-beat intervals (IBIs) caused by normal heart contractions paced by sinus node depolarization. In time domain analysis, the NN intervals are measured over the period of recording and a variety of statistical variables (Table 2) can either be calculated from these intervals directly or can be derived from the differences between these intervals. For short-term recordings (5 minutes) the SDNN and RMSSD

measures are recommended for time domain HRV assessment. SDNN is a standard deviation of the NN intervals, which is the square root of their variance and it is measured in milliseconds. As referred in [27], the variance is mathematically equivalent to the total power of spectral analysis, so it reflects all cyclic components' of the variability in recorded series of NN intervals. On the other hand, the RMSSD is the square root of the mean squared differences of successive NN intervals and it is also measured in milliseconds. This measure estimates high-frequency variations in HR in short-term NN recordings that reflect an estimate of parasympathetic regulation of the heart.

Table 3: *Selected frequency domain measures of HRV*

Analysis of short-term recordings (5 min)			
Variable	**Units**	**Description**	**Frequency range**
5-min total power	ms^2	The variance of NN intervals over the temporal segment	≈≤0.4 Hz
VLF	ms^2	Power in VLF range	≤0.04 Hz
LF	ms^2	Power in LF range	0.04-0.15 Hz
LF norm	nu	LF power in normalized units LF/(total power−VLF)×100	-
HF	ms^2	Power in HF range	0.15-0.4 Hz
HF norm	nu	HF power in normalized units HF/(total power−VLF)×100	-
LF/HF	-	Ratio LF [ms^2]/HF[ms^2]	-

The frequency domain methods employ power spectral density (PSD) analysis on HRV signal in order to provide the basic information of how power is distributed as a function of frequency. The PSD analysis uses either parametric or non-parametric approaches that, in most instances, provide comparable results. The non-parametric methods employ the Lomb-Scargle periodogram [28] or the Fast Fourier Transformation (FFT) while the parametric method use autoregressive modelling [29]. The power spectral analysis decomposes the HRV signal into different frequency components ranging from 0 to 0.4 Hz and can be classified into four distinct bands:

- **ULF -** Ultra Low Frequency (0-0.003 Hz)
- **VLF -** Very Low Frequency (0.003-0.04Hz)
- **LF -** Low Frequency (0.04-0.15Hz)
- **HF -** High Frequency (0.15-0.4Hz)

The main spectral components calculated from short-term recordings, differ from the spectral components calculated from long-term recordings. In short-term recordings (5 minutes) main components are HF, LF and VLF. The HF and LF are the prominent bands since, the VLF assessed from short-term recordings is a dubious measure. The most common HRV frequency domain measures for the analysis of short-term recordings are presented in Table 3. In the case of

long-term recordings where the frequency analysis is done over 24-hour signals the main components then include the ULF in addition to HF, LF and VLF components, since the frequency resolution increases with the total observation time.

2.4 Respiration Rate

Additional information on ANS functioning, on top of what is provided by cardiovascular measures, can be obtained by measuring of the respiratory activity, which is basically an indicator of how fast a person is breathing. The respiratory system is remarkably complicated and sensitive to a variety of psychological variables, including emotions. Physical activity and emotional arousal are reported to cause faster and deeper respiration, while peaceful rest and relaxation are reported to lead to slower and shallower respiration [30]. The rate and volume of respiration has been used to measure levels of stress, generally in conjunction with other physiological measures [31]. Hence, an additional generic sign reflecting the body's functions could be the respiratory rate. It is defined as the number of breaths a person takes per minute. The rate is usually measured when a person is at rest and simply involves counting the number of breaths per minute. Normal respiration rates for an adult at rest range from 12 to 20 breaths per minute. Most respiration monitoring systems require individuals to wear a belt around their chest. This type of system is intrusive and may affect individuals from performing their regular activities.

2.5 Electrocardiography

The origin of heartbeat is located in a small strip of heart tissue known as the sinoatrial (SA) node, or pacemaker. Each heartbeat is stimulated by an electrical impulse that originates in the SA, where a group of specialized cells continuously generate an electrical impulse spreading all over the heart muscle through specialized pathways and creating the process of heart muscle contraction well synchronized between both atriums and ventricles. The electrocardiography is the process of recording this electrical changes and constitutes a fundamental part of cardiovascular assessment. Changes in an ECG from the normal tracing may indicate one or more of several heart-related conditions. Disorders that are not associated with heart conditions may also cause changes in the ECG.

The heart contracts about 60 to 100 times a minute (normally at rest) as the electrical impulse moves through the heart. The heart activity generates small electric waves and as the electrical signal propagates throughout the body, it is captured by the electrodes. An ECG sensor

detects the electrical signal generated by the heart muscle each time it contracts and records the strength and timing of electrical signals as they pass through the heart (Figure 5). The electrical activity associated with the cardiac cycle can be detected at the surface of the body, amplified, and recorded as a time record of the electrical events occurring during each cardiac cycle. By placing electrodes on certain parts of the body, a graphic representation, or tracing, of the electrical activity can be obtained. The ECG signal, thus obtained, displays information about this electrical activity. These signal are recorded for a short or a long period of time since short-term (usually 5 minutes) or long-term (usually 24 hours) analysis is mainly performed.

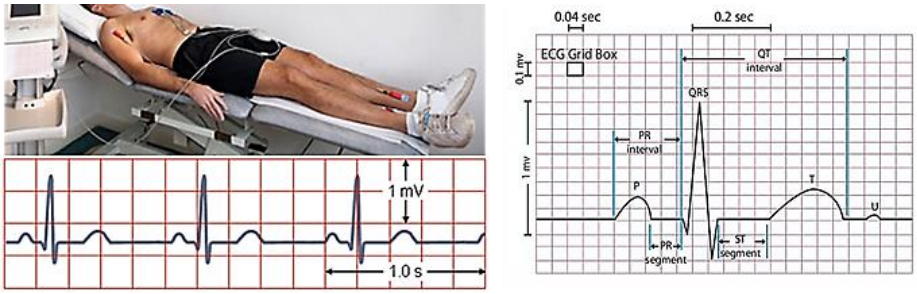


Figure 5: *ECG electrodes placed on body and the ECG components of a healthy heart beat*[6]

The ECG signals consist of waveform components which indicate electrical events during one heartbeat. Each ECG cycle consists of a series of positive and negative waves. The length of the cardiac cycle can be determined by measuring each occurrence of a specific component of the ECG waveform. On a normal ECG, there are typically five visible waveforms corresponding to different phases of the heart activities. These are the P wave, the Q wave, the R wave, the S wave, T wave and the U wave (Figure 5) and they represent potentials between rested and depolarized or depolarized and repolarized parts of whole heart. Amplitude and duration of these waves correspond to electrical power fluctuation in entire heart.

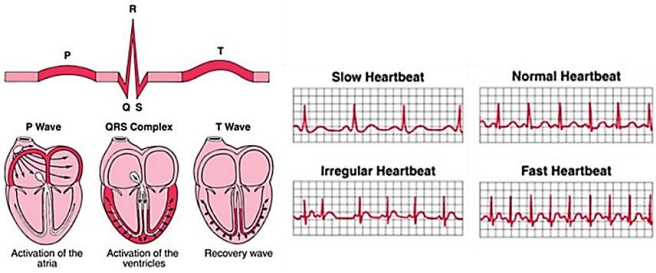


Figure 6: *The ECG waves produced in the cardiac cycle*[7]

[6] http://www.theheartattackblog.com/wp-content/uploads/2016/04/ekg01.jpg

[7] http://images.slideplayer.com/25/7612898/slides/slide_10.jpg

As presented in Figure 6, the ECG trace consists of waves and a standard wave comprises the PQRST elements [32]. The analysis of the ECG tracing not only gives information about the amount of electric activity passing through the heart muscle but also the duration of the electrical wave crossing the heart, which in turn reveals whether the electrical activity is normal, fast, slow or irregular. By convention, the ECG tracing is divided into different segments and intervals which are related directly to phases of cardiac conduction. The main propagation characteristics are the PR interval, the PR segment, the QRS complex, the QT interval, the ST segment and the RR interval. The normal limits [33] for some waves and intervals are:

- **RR interval:** 0.6-1.0 seconds (HR: 60-100 bpm)
- **P wave:** 100 milliseconds
- **PR segment:** 50-120 milliseconds
- **PR interval:** 120-200 milliseconds
- **QRS complex:** 60-100 milliseconds
- **QT interval:** 400-430 milliseconds
- **Q wave:** 40 milliseconds
- **T wave:** 160 milliseconds
- **ST segment:** 80-120 milliseconds

The QRS complex is a prominent waveform in an ECG, which gives the basis for analyzing HR as well as its variability. The instantaneous HR can be calculated from the time between any two QRS complexes. The drawback of this method is that the calculated HR can be significantly different from the measured pulse, even in a normal person, due to variations in the HR associated with respiration (the sinus arrhythmia).

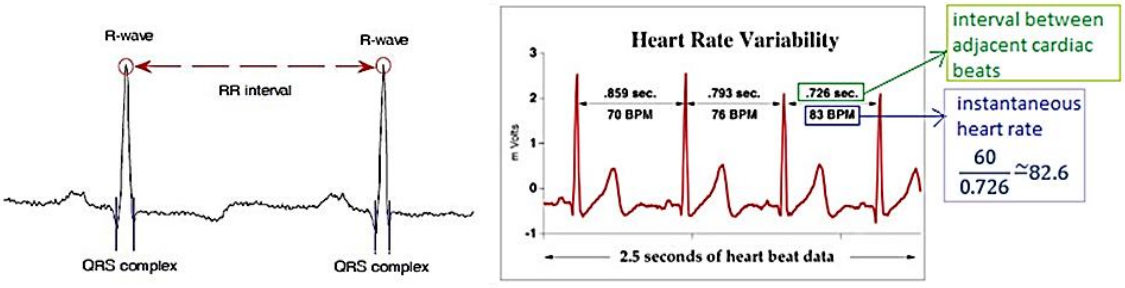


Figure 7: *Time differences between R-wave and variations of instantaneous HR*

The QRS complex is necessary for efficient extraction of RR intervals from long ECG recordings. For the HRV estimation, the peak of the R-wave is the most commonly used measure. It is manifested as a sharp positive peak followed by a negative deflection in the ECG waveform. The peak of the R-wave is normally greater in amplitude than all other peaks in the ECG, making it easily distinguishable. Since, HRV describes both variations of both instantaneous HR and

intervals between successive fiducial points of R peaks of the QRS complex (RR intervals), its assessment is performed from the analysis of intervals between adjacent cardiac beats or from instantaneous HR values for each cardiac beat from continuous ECG recordings (Figure 7). Accuracy of the RR series is crucial for reliable HRV analysis and even a single artifact in the RR signal could impart substantial spurious variance into all commonly analyzed frequency bands.

2.6 Photoplethysmography

Photoplethysmography comes from the Greek word of "plethysmos" meaning increase and "graph" which is the word for write. It is a non-invasive optical technique for measuring blood flow in tissue, widely used in pulse oximeters. The blood volume changes are recorded from finger, toe, ear lobe, forehead and other locations based on the application requirements. This method has widespread clinical application [34], since it can be used to monitor vital body signs such as HR and to provide valuable information about the cardiovascular system. Traditionally, PPG uses dedicated light sources (e.g. red or infrared light) and a skin contact photodiode to measure the amount of light either transmitted or reflected (Figure 8). The transmission mode measures how light is obstructed and absorbed by the tissue while in the reflectance mode estimates of how the light source is reflected from the skin surface are obtained. The PPG signal represents an average of all blood volume in the arteries, capillaries, and any other tissue through which the light has passed. The amount of light that returns to a PPG sensor's photodetector is proportional to the volume of blood in the tissue. Thus, this method can be used to detect subtle color changes caused by variations in reflected light due to change of volume in the underlying blood vessels during the cardiac cycle.

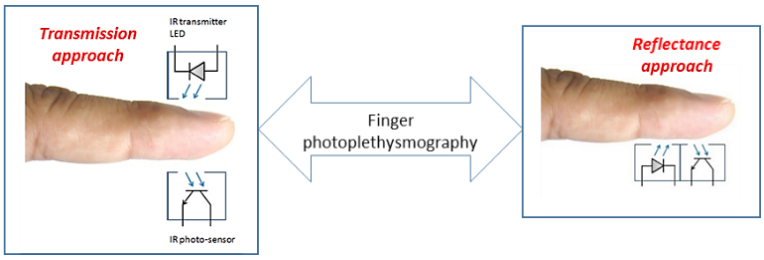


Figure 8: *Transmission and reflectance types of PPG*

The appearance of the PPG pulse is commonly divided into two phases, the anacrotic and the catacrotic phase. The anacrotic phase is the rising edge of the pulse, whereas the catacrotic phase is the falling edge of the pulse as shown in Figure 9a. The first phase is primarily concerned with systole, and the second phase with diastole and wave reflections from the periphery. The

distance between two consecutive systolic peaks is referred as the Peak-Peak interval and the distance between the beginning and the end of the PPG waveform is referred as the Pulse interval (Figure 9b). The Pulse Rate Variability (PRV) extracted from PPG has been studied as a potential surrogate of the HRV and results demonstrate that PRV in PPG signal and HRV in ECG signal are highly correlated [35]. These findings suggest that PPG signals could be used as an alternative measurement of HRV.

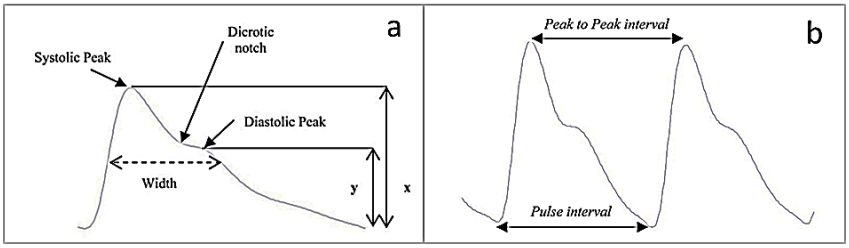


Figure 9: *(a) A typical PPG waveform where x is the amplitude of the systolic peaks while y is the amplitude of the diastolic peak and (b) two consecutive PPG waves* [35]

The use of PRV, as a surrogate of HRV, could be useful in applications where, for example, ECG is not available. The main difference between HRV and PRV is the time taken by the pulse wave to travel from the heart to the finger. This time is called the pulse transit time (PTT) and is typically measured as the difference between the peak of the R-wave on the ECG and the peak value of the corresponding pulse in the finger pad measured by PPG (Figure 10) [36].

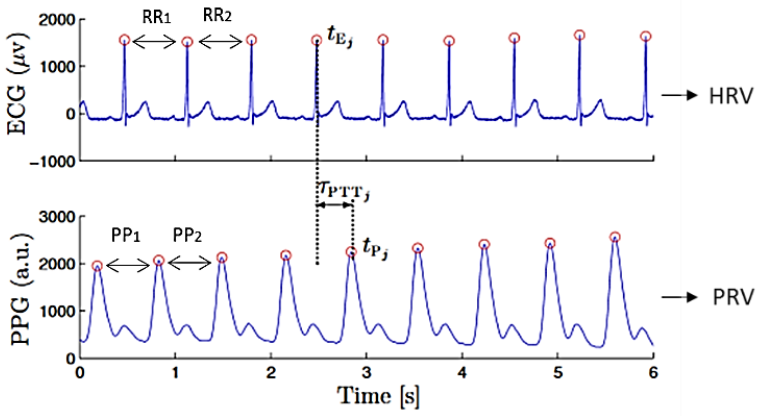


Figure 10: *The ECG and PPG pulse wave comparison*

During the last decade, researchers have begun to explore the use of simple and low-cost methods for non-contact vital sign extraction and PPG signal measurement. Initially, the skin contact photodiode was replaced with a camera and several studies report that this approach is capable of monitoring blood perfusion based on both transmission and reflection modes (Figure 11) without making contact with the tissue [37], [38].

Subsequently, research focused on measuring signs and obtaining PPG using ambient light rather than a dedicated light source. Takano and Ohta [11] proposed such a low cost solution to measure pulse rate by capturing the time-lapse image from a handy video complementary metal–oxide–semiconductor (CMOS) camera under illuminance conditions of 860 lx. The system could measure the 30-s average HR based on the changes in the brightness of a region of interest (ROI) set around the cheek. Their analysis showed that even in cases where the scene illumination is between 270 and 1500 lx, which is the typical illuminance range for human daily activities, the measurement could be satisfactorily conducted.

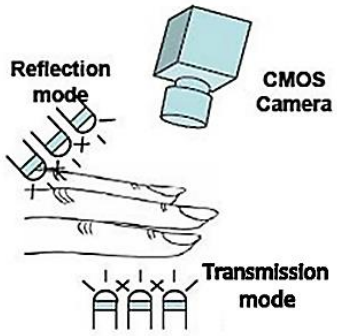


Figure 11: *Imaging PPG system setup based on transmission and reflection modes* [38]

In 2008, Verkruysse et al. [39] reported for the first time that remote, i.e. contactless, measurement of the cardiac pulse and respiratory could be obtained from PPG with the use of a simple digital, consumer level photo camera in movie mode and ambient light as the only illumination source. In this study, the analysis was performed on color movies recorded at either 15 or 30 frames per second (fps) with pixel resolution of 640 x 480 or 320 x 240 and saved in Audio Video Interleave (AVI) format. Their results suggested that not only respiratory and pulsatility but also phase information regarding the cardio-vascular waveform might be deduced. The main purpose of this paper was to demonstrate that video sequences of the human face, recorded with a simple digital camera and ambient light as the only source of illumination, contain robust reflectance PPG signals.

This initial efforts have generated significant research interest on the potential use of PPG-based methods for the accurate, contactless estimation of HR and HRV. The most important developments reported during recent years are thoroughly reviewed and analyzed in the subsequent Chapter.

3 Literature Review

As previously said, several studies have been reported focusing on contactless detection of the PPG signal, mainly for HR estimation from the human face, with a camera as the detector and ambient light as the illumination source. A variety of methods have been proposed for addressing issues and important questions that emerged from this research. In 2010, Poh et al. [12] proposed a methodology for non-contact measurements of the cardiac pulse. The authors state in their study that attempts to that point of time were non-automated, susceptible to motion artefacts and typically expensive and that the method proposed in their publication was able to overcome these problems. The technique employed by the authors was based on Blind Source Separation (BSS) of the color channels into independent components. The color channels were decomposed into three independent source signals using Independent Component Analysis (ICA), based on the joint approximate diagonalization of eigen-matrices (JADE) algorithm. The desired source signal was defined as the second component, since this component typically contains a strong PPG signal. The cardiac pulse rate extracted from videos recorded by a basic webcam was compared to a Food and Drug Administration (FDA)-approved finger blood volume pulse (BVP) sensor and their analysis reported high accuracy and correlation, even in the presence of movement induced artefacts. Furthermore, they applied their technique to perform HR measurements from three participants simultaneously and demonstrated the capability of performing concomitant measurements on more than one person at a time.

The previous methodology was extended to quantify multiple physiological parameters and more specifically HR, respiratory rate as well as HRV [40]. ICA, based on the JADE algorithm was used again, but this time the authors reported that the order in which ICA returned the independent components was random. Hence, the component whose power spectrum contained the highest peak was selected as the desired source signal. In this later study, the analysis of HRV was performed by PSD estimation using the Lomb periodogram. The LF and HF components were measured and by locating the center frequency of the HF peak, the authors were able to determine the respiratory rate. The results indicated a strong correlation between the parameters derived from the webcam recordings and the standard reference sensors.

The pulse rate estimation based on webcam recordings was also explored by Lewandowska et al. [41] who managed to obtain the pulse component using the FastICA algorithm and Principal

Component Analysis (PCA). The reported results show that those two methods extracted the pulse component with similar accuracy. However, although the accuracy of extraction of cardiac pulse signal by PCA was comparable to that obtained by FastICA, PCA was the less computationally demanding of the two. Concerning the ROI, the analysis was performed for two different regions, i.e. the whole face and the forehead. The forehead was proposed since it does not contain any moving elements, like blinking eyes or moving lips. The results indicated that the forehead ROI was representative for the whole face region but, as the size of the ROI decreases, the level of noise increases.

The potential of reliable HR measurement from facial videos recorded using a smartphone camera was explored in [42]. ICA, based on the JADE algorithm, was used and the average HR was extracted from both the green channel and the second independent component, using the maximum power frequency of the FFT in each signal. In addition, an FDA-approved ECG recording system was used to obtain the reference ECG signal. The reported analysis indicated that the green channel performed better than the second independent component and the very low error rates indicated that a reliable average HR could be measured using the smartphone camera.

Regarding the camera, Sun et al. [43] compared the performance between a low-cost webcam and a high-sensitivity CMOS camera to assess HR at rest and after exercise. The analysis showed that the webcam and CMOS camera exhibited comparable capabilities in measuring HR. The influence of the intensity of ambient light on the physiological information was also considered. The authors' results reveal that variations in ambient light intensity did not affect the ability of both camera systems to track accurately HR at rest and changes in heart associated with exercise. The HR values were obtained by performing the time frequency representation (TFR) on the green channel's PPG and the estimations were compared to the pulse signal measured with a commercial pulse oximetry contact sensor.

Wei et al. [44] reported that image sequences of human face contain the information of BVP for the estimation of HR, but the camera captures the source signal as a mixture of BVP along with other source of fluctuations in light. Therefore, there is a need for methods to recover the source signals from the set of the observed mixtures. The authors of the above paper performed extensive experiments on alternative solutions using nine dimensionality reduction methods, namely: Isomap, Laplacian Eigenmap (LE), Locally Linear Embedding (LLE), Local Tangent Space Alignment (LTSA), Maximum Variance Unfolding (MVU), Linearity Preserving Projection

(LPP), Linear Discriminant Analysis (LDA), PCA and ICA. Through the comparison of these methods, the authors concluded that LE obtained the highest degree of agreement with measurements taken with a finger pulse oximetry. In addition, Tsouri et al. [45] report that the algorithms based on ICA suffer from a sorting problem which hinders their performance. Hence, they proposed an algorithm based on constrained ICA (cICA) in order to extract the PPG signal, resolve the sorting problem and improve performance. According to the authors, previous studies have reported that the PPG signal can be extracted from ICA when selecting either the second component or the component with the maximum spectral peak. Their analysis showed that the cICA produced much better results and provided measurement accuracy similar to the one achieved with a finger probe oximeter. On the other hand, the cICA approach required more processing time than the ICA approach.

The performance of three BSS algorithms, based on ICA, in order to perform a comparative evaluation of their accuracy and convergence capability for optical HR estimation was studied at [14]. The selected algorithms were the JADE, the FastICA and the RobustICA and these methods were applied to decompose the signal into independent components. Then, the desired source signal was defined with two different approaches for each method, either as the green channel or the component with the maximum spectral peak. Furthermore, the average HR was extracted from the cheek of the participants and was compared with the values determined from a finger pulse oximeter. The analysis indicated that both JADE and FastICA performed well with the FastICA algorithm performing slightly better. Concerning the extraction of the desired source signal, the results showed that a reliable HR could be measured by exploiting only the green channel.

A comparative study on the influence of selecting different face regions as the ROI for contactless extraction of the HR in the visible spectrum was presented in [46]. ICA, based on the JADE algorithm, was applied to decompose the signals into independent components and the average HR was extracted from ten facial ROIs. The Mobi 8 device produced by TMS International B.V. was used for the reference pulse measurement. The analysis indicated that, although the order of the resulting independent components is not known, the second channel contained the HR signal. Concerning the ROI, the best results were achieved from the cheek followed by the forehead, while the eye regions provided the lowest performance. A reported result of interest is the fact that when the whole face is used as ROI the results were not the optimal.

At approximately the same time, Lempe et al. [47] also presented an evaluation of the suitability of various ROIs for PPG measurement. Their study compares the average signal quality of different ROIs in order to examine whether it is possible to improve the PPG signal's quality by determining the best suited ROI. For this purpose, the Verkuysse's method [39] of pulse amplitude mapping was extended by using precise reference signal. The ground truth reference was represented by the ECG together with the PPG, recorded from an ear clip on the right earlobe. For the analysis only the green channel was considered, since this channel contained the strongest PPG signal. The results showed that the cheeks provided the highest mean signal quality.

Monkaresi et al. [48] evaluated the technique proposed by Poh et al. [12] in a controlled situation, i.e. in a naturalistic human computer interaction (HCI) environment and during exercising scenarios which contained more user motion and different lighting conditions. For comparison, HR was measured simultaneously using an ECG device. The results demonstrate the feasibility of this methodology to measure HR at rest. However, this method was less accurate in naturalistic HCI and indoor exercise situations. The evaluation of the performance of the three output components of ICA in estimating actual HR revealed that each component could estimate the HR better than others on different occasions. Therefore, the method could not rely only on one component to obtain the HR signal, particularly on noisy occasions when the observed signals are affected by head movements and changes in illumination. For this reason, machine learning (ML) techniques were proposed by the authors to estimate the HR from the three ICA components automatically. They reported that by obtaining some information from each output component and training a model to find the spectrum that contains the HR signal, a better estimation could be obtained. For each participant, specific models were built and trained using ML techniques. More specifically, the linear regression and k-nearest neighbor (kNN) classifier were used to estimate HR from nine features extracted from PSD analysis. The experimental results suggested that the kNN-based technique outperformed in the HCI scenario and improved the accuracy in indoor exercising conditions. However, the accuracy achieved is not acceptable in some healthcare settings (e.g., emergency severity index triage), where a more accurate estimation must be provided.

Li et al. [49] proposed a framework which utilized face tracking and Normalized Least Mean Square (NLMS) adaptive filtering methods to count the influence of motion and illumination variations. They tested their framework on the MAHNOB-HCI dataset [50] and demonstrated that

their method substantially outperformed previous methods. They also used their method for long term HR monitoring in a game evaluation scenario and achieved promising results. Regarding the analysis, the face was considered as the ROI and the green channel was selected, since a preliminary test showed that this channel worked best. The proposed framework was evaluated using three experiments. In the first experiment, authors re-implemented previously proposed methods and tested them on a simple database collected by themselves. The purpose of this experiment was only to demonstrate that the previously proposed methods have been correctly re-implemented. In this test, the number of participants was 10, the HR was measured at rest and a Polar S810 HR monitor system was used to record the ground truth. Also ambient illumination variations and body movement were negligible during the video recordings, hence another experiment was carried out in order to test the robustness of these methods in relation to these challenges. In this second experiment, the previously proposed methods were tested again on the MAHNOB-HCI database, the number of participants was 27 and the ground truth reference was the simultaneously recorded ECG. Finally, a third experiment was carried out in order to test the proposed framework in a game playing scenario. For this reason, the face video of 1 subject for 10 minutes was recorded while playing a videogame in the same setting as in the first experiment. The HR monitoring during game playing was presented as an example of applications of this framework under realistic situations.

McDuff et al. [51] reported that traditional red-green-blue (RGB) cameras have limited frequency resolution and alternative color bands could be useful in recovering the BVP. Therefore, they presented an automated method for recovering the BVP waveform from video recordings using a camera with five band channels; red, green, blue, cyan and orange. For the analysis, the ICA, based on JADE algorithm was applied to decompose the signals into independent components and the appropriate source signal was defined as the component whose power spectrum contained the highest peak. The experimental results show that alternate frequency bands, in particular an orange band, allow physiological measurements much more highly correlated with an FDA-approved contact PPG sensor. Moreover, the combination of cyan, orange and green color channels was the best for recovering the HR. In this study, the selected ROI was the face excluding the section around the eyes, the number of participants was 10 and the HR was measured at rest and under cognitive stress. The results showed that correlations of over 0.92

($p<0.01$) were obtained for HR, BR and HRV measurements and the remotely measured HRV spectrograms closely matched those from the contact approach.

Kumar et al. [52] reported that the major challenges in estimating PPG using a camera is the extremely low signal strength of the color-change signal, particularly for darker skin tones and/or under low lighting conditions, and the motion artefacts due to an individual's movement in front of the camera. Therefore, they proposed a method of combining skin-color change signals from different tracked regions of the face using a weighted average, where the weights depended on the blood perfusion and incident light intensity in order to improve the signal-to-noise ratio (SNR) of a camera-based estimate. Their method was tested on people having diverse skin tones, under various lighting conditions and natural motion scenarios. It showed significant improvement in the accuracy of HR estimation on single channel (green) video data recorded with a monochrome camera having a green filter in front of it. Furthermore, the PRV derived from the PPG signal estimated using distancePPG showed significant improvement in the root mean squared error (RMSE). Using distancePPG method, the RMSE in PRV for light and medium skin tone was less than 16 ms and since a 30 fps camera was used for recording the video, one solution proposed by the authors to further reduce the RMSE of PRV is to use higher frame rate camera (e.g., 100 fps). The proposed method was also compared with an ICA-based method using two sets of color videos, a static and a talking dataset. In the static scenario a small improvement in SNR was observed whilst for the talking scenario both methods performed poorly, possibly because of large motion artefacts which rendered small PPG signal estimation difficult. The authors in this study used color videos, the selected ROI was the face, the number of participants was four, the ground truth was based on pulse oximeter recordings from the earlobe and the HR was measured on a static and a talking dataset. The authors reported that the talking dataset was deliberately chosen to be extremely harsh (lower light, non-Caucasian skin tones, and large motion during talking) to highlight the scenario where current known algorithms (including the proposed method) showed unsatisfactory performances.

Mannapperuma et al. [53] investigated the limits of detection of the HR while reducing the video quality. They compared the performance of three ICA methods (JADE, FastICA, RADICAL) and discussed the different sources of increasing error and other limiting conditions in three situations of reduced signal-to-noise ratio. In the first one, the area of the analyzed face was decreased from 100 to 5%, in the second the face area was progressively re-sampled down to

a single RGB pixel and in the third, the HR signal was severely reduced with respect to the boundary noise. The results indicated that for JADE, RADICAL and FastICA, the reduction of the ROI area had little effect on the HR error and only when the ROI was reduced to 5% the error increased significantly. Concerning the video sequence length, a 15 seconds segment showed similar trends to a 30 seconds for all ICA algorithms used.

Finally, Wedekind et al. [54] proposed Markov models to describe and subsequently identify the cardiac component from the output of BSS technique applied to camera-based PPG. This method was compared with five alternative approaches and the analysis showed that good results could be obtained by combining different simple Markov models. The main purpose of their study was the investigation of automated component selection algorithms after applying PCA and FastICA to the camera-based PPG.

Table 4: *Overview of measured signs and proposed ROIs, separation and selection methods*

		References
Measured sign	*HR*	[12], [14], [40]–[49], [51]–[54]
	Respiration	[40], [51]
	HRV	[40], [51], [52]
Region of interest	*Face*	[12], [40]–[49], [51]–[53]
	Cheek	[14], [46], [47]
	Forehead	[41], [46], [47], [54]
	other	[46], [53]
Separation method	*JADE*	[12], [14], [40], [42], [44], [46], [48], [51], [53]
	FastICA	[14], [41], [53], [54]
	PCA	[41], [44], [54]
	other	[14], [44], [45], [53]
Selection method	*Highest peak*	[14], [40], [41], [51], [53]
	2nd Component	[12], [42], [46]
	Green channel	[14], [42], [43], [47], [49], [52]
	other	[41], [45], [48], [54]

Thorough reviews of non-contact monitoring methods which include PPG measurement techniques can be found in [55]–[57]. In summarizing the results of our analysis of reported research findings to date, it is evident that there is an increasing research activity focusing on the issues related to physiological measurement from facial videos using PPG. However, most of the reviewed studies have focused on HR estimation and limited attempts have been made to measure HRV and respiratory rate. These few studies addressing the HRV analysis were also mainly based on the use of BVP sensors or pulse oximeters as a reference and they lack of direct quantitative

comparison between the video-based PPG and ECG signals. Finally, it is worth mentioning that most of these studies use their self-recorded datasets and only in [49], in one of their experiments, the publically available dataset MAHNOB-HCI was used. Moreover, in most cases the experiments were conducted with low-cost RGB cameras except from [43], [47], [54], in which a high performance camera was used and [51] in which a five-band digital camera was used. Yet the results of the studies reviewed indicate that camera-based, non-contact vital sign measurement is definitely feasible, with convincing results already obtained.

The relevant studies reported above are summarized in Table 4 and categorized with respect to the measured sign, the selected ROIs and where applicable the various separation and component selection methods.

Table 5: *Overview of studies on monitoring HR with video-based PPG*

Study	*Sample size*	*Experiment (measurement)*	*Records x Duration*	*Sampling rate*	*Reference*
Poh et al. [12]	12	a) at rest b) during motion	24 x 1 min	15 fps	finger BVP sensor
Poh et al. [40]	12	at rest	12 x 1 min	15 fps	finger BVP sensor
Lewandowska et al. [41]	10	at rest	10 x 0.5 min	20 fps	ECG
Know et al. [42]	10	at rest	10 x 1 min	29.9 fps	ECG
Sun et al. [43]	10	a) at rest b) after exercise 1 c) after exercise 2 d) recovery	80 x 0.5 min	50 fps (CMOS) 30 fps (webcam)	pulse oximeter
Wei et al. [44]	20	at rest (sunlight, spotlight)	40 x 0.5 min	30 fps	finger pulse oximeter
Tsouri et al. [45]	45	at rest	45 x 1 min	15 fps	finger probe oximeter
Datcu et al. [46]	a) 6 b) 1	a) at rest b) after physical effort	7 x 0.5 min	15 fps	Mobi 8 device
Lempe et al. [47]	20	at rest	20 x 3 min	100 fps	ECG, PPG ear clip
Monkaresi et al. [48]	a) 10 b) 10 c) 1	a) at rest b) HCI scenario c) indoor cycling	a) 10 x 0.5 min b) 10 x 30 ± 10 min c) 10 x 17 min	a) 30 fps b) 30 fps c) 15 fps	ECG
Li et al. [49]	a) 10 b) 27 c) 1	a) at rest b) MAHNOB-HCI c) playing video game	a) 10 x 0.5 min b) 527 x 0.5 min c) 1 x 10 min	a) 30 fps b) 60 fps c) 30 fps	a) Polar S810 monitor b) ECG c) Polar S810 monitor
McDuff et al. [51]	10	a) at rest b) under cognitive stress	a) 9 x 2 min b) 10 x 2 min	30 fps	contact PPG sensor
Christinaki et al. [14]	7	at rest	16 x 1 min	30 fps	finger pulse oximeter
Kumar et al. [52]	4	a) static dataset b) talking dataset	a) 4 x 1 min b) 4 x N/A	30 fps	earlobe pulse oximeter
Mannapperuma et al. [53]	18	at rest	9 x 1.5 min	15 fps	finger pulse oximeter
Wedekind et al. [54]	18	at rest	18 x 20 ± 7 min	100 fps	ECG, finger PPG

Based upon the presented comprehensive review and on a critical analysis of the available research results, it becomes obvious that, although significant progress has already been made, a

number of issues still remain to be successfully addressed. These relate to a) which is the optimal ROI to be used for HR measurement, b) which method is most appropriate for forming the traces and c) which is the most appropriate method for selecting the best component for the measurement of the cardiac pulse.

However, based upon the presented comprehensive review and on a critical analysis of the available research results, it becomes obvious that, although significant progress has already been made, a number of issues still remain to be successfully addressed. These relate to a) which is the optimal ROI to be used for HR measurement, b) which method is most appropriate for forming the traces and c) which is the most appropriate method for selecting the best component for the measurement of the cardiac pulse.

Furthermore, the concise overview of the studies presented in Table 5 indicates some factors that can influence the evaluation and are likely to impact on the accuracy of measurements. These are related to the relatively small number of participants used in most studies, the limited duration of the available data for analysis and the low frame rate used in several datasets. Taking into consideration that in the present work our focus is on evaluating the accuracy obtained in measuring HR with our proposed computational pipeline and methods, these findings have to be taken into account. An additional reason for taking these findings into consideration is the fact that a subsequent objective is to extend the methods presented for contactless HRV analysis.

4 Description of Available Datasets

Reference datasets are important in order to be able to fulfil the research requirements and, although there are datasets publically available that are relevant to our research questions, they do not completely cover the needs of the present study. As an example, there are datasets that even though they include both video and ECG signals, the recording data were acquired only in a noisy scenario or the data includes near-infra-red (NIR) facial videos. For similar reasons, the studies reviewed in the previous Chapter (Chapter 3) have used their own datasets.

The estimation of physiological parameters from the face requires the acquisition of frontal facial videos with an adequately high frame rate and a duration of at least 5 minutes (required for the HRV estimation), in order to be clinically valuable. Furthermore, the simultaneously recorded reference ECG is required, as ground truth, in order to assess the validity of the HR and HRV algorithms derived from the video processing. Other basic requirements are an adequate number of participants, good illumination conditions and limited participant's motion during the whole recording.

4.1 Self-recorded Videos

Taking into consideration the above requirements from the dataset, our initial attempt focused on generating a number of self-recorded videos. Such self-recorded videos were acquired in our lab. The experimental setup consisted of a built-in laptop camera (iSight on an Apple Macbook Pro) and a basic webcam (Logitech 720p HD). All videos were recorded at 30 fps, in 24-bit RGB color with a resolution of 640 × 480 pixels, and were stored in uncompressed AVI format. The experimental setup included a commercially available fingertip pulse oximeter (Heal Force Prince-100A), which was used to obtain the participant's reference cardiac pulse during the video recordings. The experiments were conducted indoors, at different times of the day, with a varying amount of indirect sunlight and fluorescent lights as the only sources of illumination. The participants were seated at a distance of approximately 0.5 m from the cameras. Seven subjects (3 males and 4 females) participated at the recordings, out of which two wore glasses and had light beard. They were asked to sit naturally and look straight at the cameras during video capturing. They were free to make facial expressions and to move their head or body slightly while remaining seated. Two facial videos, each lasting one minute, were simultaneously recorded for all

participants from the cameras. One participant was also recorded after a running session in order to support the motion induced noise/artefact removal experiment. In total, 16 separate videos were stored for analysis (6 subjects x 2 videos (both cameras) + 1 subject x 4 videos (before/after running session, both cameras)).

4.2 Mahnob-HCI

To the best of our knowledge, none of the existing freely available datasets covers completely the needs of our research and only the publically available multimodal database, the MAHNOB-HCI, fully described in [50], seems to be a good approximation to our needs, especially with respect to the required high frame rate. Furthermore, the physiological signals in this dataset e.g. ECG, can be used as ground truth for validating results of image processing algorithms employed for detecting these signals through facial imaging.

Figure 12: *The MAHNOB dataset experimental setup (image taken from* [50]*)*

The MAHNOB-HCI was recorded in response to affective stimuli with the goal of emotion recognition and implicit tagging research. It includes data from two experiments: the emotion elicitation experiment and the implicit tagging experiment. Thirty subjects, from both genders and different cultural backgrounds, participated in these experiments. In the first one, they watched 20 emotional videos and self-reported their felt emotions using arousal, valence, dominance, and predictability as well as emotional keywords. In the second, short videos and images were shown once without any tag and then with correct or incorrect tags. The experimental setup, as shown in Figure 12, was arranged for synchronized acquisitions of face videos, audio signals, eye gaze data, and peripheral/central nervous system physiological signals. The participant's behavior was recorded using six cameras, a microphone, and a gaze tracker, whereas the physiological responses were recorded using a Biosemi Active II system. The facial videos were recorded at 60 fps with a

space resolution of 780 x 580 pixels. The corresponding ECG signals were recorded using three sensors attached on the participants' body. The recorded videos and bodily responses were segmented and stored in a database.

The MAHNOB-HCI database was selected due to the fact that, apart from the fact that it is publicly available, it contains a wide range of vital signs acquired in parallel with the two-dimensional (2D) videos. In our study only the frontal videos and the ECG were used, since our computational processing is performed within the facial region and the measured ECG provides the ground truth. In this dataset, the data recorded from participants 9 and 15 are not complete due to technical problems and the physiological responses of participant 12 are missing due to recording difficulties. Therefore, the data recorded from these three participants (P9, P12, P15) were not analyzed due to technical problems and unfinished data collection. Moreover, we have excluded from our analysis the recordings in which the face detection algorithm failed to detect the face within an adequate number of frames. Hence, the analysis results of this thesis are only based on the frontal RGB videos recorded from 21 participants resulting to 50 recordings available for analysis and the corresponding ECG.

4.3 SEMEOTICONS Reference Dataset

The SEMEOTICONS reference dataset includes data produced in the context of the EU funded project with the acronym SEMEOTICONS (http://www.semeoticons.eu/) which was acquired during a dedicated acquisition campaign in Pisa, Italy (May 2015). This reference dataset comprises of data collected from 24 volunteers (17 males and 7 females), aged 47.2±9.3 years old, including both two dark skinned male subjects as well as participants having beard and wearing glasses. Some of the participants are used as controls and some others are individuals with diagnosed cardiometabolic risks, e.g. hypercholesterolemia, hypertension not requiring drug treatment. Each study participant gave a written informed consent for her/his participation, an informed consent for the scientific use of the acquired images, as well as an informed consent for the publication of audio-visual material for dissemination purpose. The study was approved by the North-West Tuscany ESTAV (Regional Health Service, Agency for the Technical-Administrative Services of Wide Area) Ethical Committee with approval number 213/2014 (25/09/2014).

The experimental setup (Figure 13) consisted of a Point Grey Grasshopper 3 (GS3-U3-41C6C-C) USB camera with lens of fixed focal length of 16 mm. All videos were recorded at 90

fps, in 24-bit RGB color with a resolution of 1600 × 1216 pixels. Contrary to other published efforts which employ low-cost web cameras, SEMEOTICONS has used a high end camera, since the overall vision and requirements of the project necessitate demanding emotional analysis for the detection and monitoring of adverse psychological states such as anxiety, stress or fatigue. These needs dictate specific requirements on the selected camera for the acquisition of raw data of such quality that could enable the computation of the selected descriptors with satisfactory accuracy. In addition, the experimental setup included a Micromed Brain Spy Plus Holter electroencephalographic (EEG)/ECG device with a sampling rate of 256 Hz, which was used to record the participants' ECG in parallel with most of the video recordings.

During the recording, the subjects were seated at a distance of approximately 0.6 m in front of the camera and were asked to limit their movements. Conditions of good diffuse illumination were provided with 2 x 500 W halogen lamps pointed away from the participant to produce diffused light (reflected on white wall). Moreover, calibration was performed by using a 50% grey card in front of the camera and a black backdrop was placed behind subjects to minimize reflection and provide a constant background scene.

Figure 13: *The SEMEOTICONS dataset experimental setup (published with subject's permission)*

Several videos of the 24 subjects were collected while participants were: a) in a neutral state, b) describing themselves, c) mimicking situations of stress or anxiety or fatigue, d) during a Stroop test [58], and e) watching a set of relaxing or stressful images or videos. A large high-resolution dataset was, thus, collected using the high-end camera together with suitable experimental protocol for a reliable assessment of the ground truth. In total 12 trials, as presented in Table 6, were recorded and in each trial the color facial video sequence was saved as MP4 format, compressed with a MJPG codec. The duration of the video varied from 60 to 300 seconds. The longest duration videos are to be used for the estimation of HRV. As a result, the entire dataset

consists of more than 290 videos and the total duration of collected data was more than 430 minutes. Hence, to the best of our knowledge, it constitutes one of the largest datasets, with a large number of facial videos combined with the corresponding ground truth represented by a simultaneously recorded ECG, to date to be used for the estimation of physiological parameters from facial videos, when compared to the videos reported in the studies summarized in Table 5.

In the present thesis, the dataset used for the validation of our methods was slightly smaller, since the video recordings of some participants were not available for analysis and for other participants some recordings had specific issues, e.g. non-availability of the ECG signal due to technical problems during the recording, which forced us to exclude these data from the analysis. Moreover, we have excluded from our analysis the recordings in which the face detection algorithm failed to detect the face within an adequate number of frames due to specific factors, such as the large degree of head tilting or occlusion of the face. Hence, the analytic results of this thesis are based on 247 videos recorded from 24 participants (17 males) when the whole dataset is used and on 229 videos recorded from 22 participants (15 males) when the two dark skinned participants are excluded. The duration of these videos varies from 60 seconds to 120 seconds according to the recording scenario per trial as reported in Table 6.

Table 6: *The recording scenarios of the SEMEOTICONS dataset*

TRIAL	DURATION	TYPE
1	1 minute	*Reference period*
2	1 minute	*Self-describing*
3	1 minute	*Mimicking neutral/rest state*
4	1 minute	*Mimicking anxiety state*
5	1 minute	*Mimicking stress state*
6	1 minute	*Mimicking fatigue state*
7	2 minutes	*Viewing calming and non-calming images*
8	2 minutes	*Mental Task (Stroop test)*
9	1 minute	*Recording with blank monitor*
10	2 minutes	*Viewing calming videos*
11	2 minutes	*Viewing adventure film*
12	2 minutes	*Viewing action videos*

It must be noted that the complete dataset contains the recordings of the dark skinned participants in each trial, apart from trial 5 where the face detection algorithm failed to detect the face for both of these participants. Moreover, the estimation of HRV is based on five minutes of continuous video recordings which were collected from 5 participants (1 video per participant). Finally, we were provided with the corresponding 247 reference HR measurements extracted from the recorded ECG signals to be used as ground-truth for algorithm verification.

5 Technical Implementation

The technical implementation refers to the implemented signal processing pipeline for non-contact estimation of physiological parameters from optical recording by means of PPG. The implemented and validated methods perform estimations from facial video streaming via both spatial and temporal analysis.

In the context of this thesis, a measurement computational pipeline will be configured with alternative choices for estimating the most suitable parameters at each step of the pipeline. In most of these steps, there are open issues that have emerged from the review presented in Section 3, which need to be addressed in order to optimally configure our signal processing pipeline. These open issues relate to the following questions: a) which is the most appropriate ROI to be used for HR estimation, b) which is the optimal method for forming the signal traces, c) which is the optimal BSS algorithm to overcome the influence of motion artefacts and d) which is the most appropriate method for the optimal selecting of the signal component for further analysis, i.e. the one with the highest SNR, containing the cardiac pulse.

An overview of the general proposed pipeline for the HR, HRV and respiratory rate estimation is presented in Figure 14.

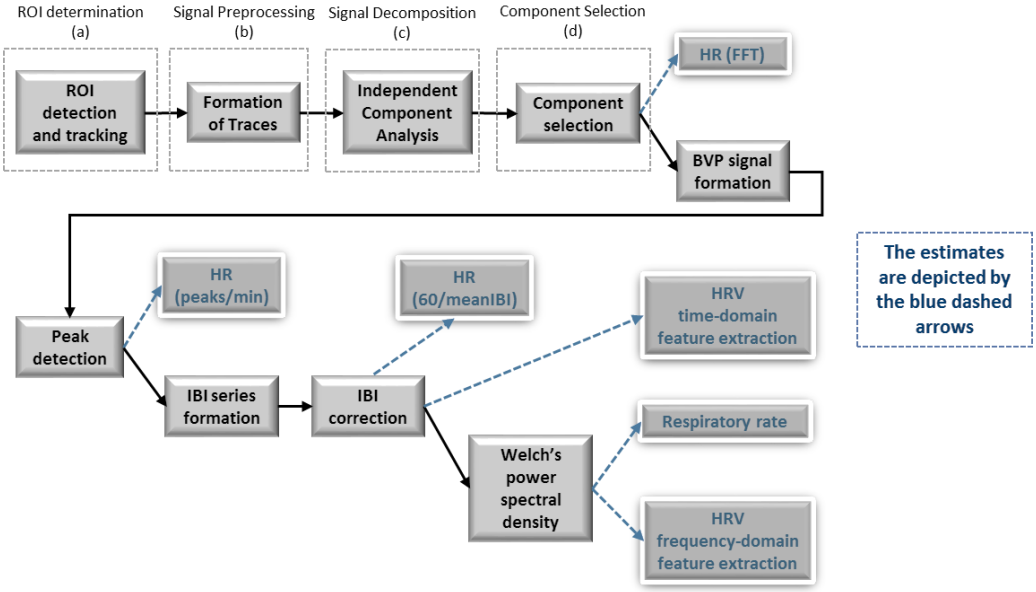


Figure 14: *Overview of the signal processing pipeline for the HR, HRV and respiratory rate estimation and the exploratory sub-questions (a)-(d)*

The error induced at each stage will be considered during the investigation of the appropriate signal processing techniques for deriving the cardiovascular information in the context of the above-mentioned issues.

5.1 Exploratory Study

Currently, the proposed methods for non-contact vital signs, such as HR, measurement from video mainly implement a general analysis framework comprises of the following five main steps: (1) *ROI determination*, (2) *signal pre-processing*, (3) *signal decomposition*, (4) the *component selection* and (5) *HR estimation*. *ROI determination* refers to the detection and tracking of the optimal region in which the average brightness will be measured. The *signal pre-processing* refers to the optimal method to form the traces, a series of operations involving detrending and normalization to enhance and prepare the signal for analysis. *Signal decomposition* refers to the optimal method to decompose mixtures of source signals into a set of source components. *Component selection* refers to the identification of the component which contains the best replica of the BVP signal and finally *estimation* refers to the pulse rate measurement.

Given the aforementioned state-of-the-art and the questions at hand, we have designed an exploratory study with the objective to estimate HR using the FFT analysis and to recover the BVP signal. The present section reports our findings, utilizing a self-recorded dataset and a publically available dataset, for the evaluation of the adequacy, in terms of measurement accuracy, of camera-based HR monitoring outside of conventional clinical settings. The exploratory study refers only to HR measurement since the resulting method will lead in finding the underlying BVP signal and will be exploited as a pre-processing step for the HRV and respiratory measurement algorithm. Specifically, we focus on a set of important sub-questions (Figure 14) that relate to (a) which is the most appropriate ROI to be used for the estimations, (b) which is the optimal method for forming the signal traces, (c) which is the optimal BSS algorithm to overcome the motion artefacts limitation, and (d) which is the corresponding best method for selecting the most appropriate of those signals. The main contribution relates to the identification and validation of an optimal image processing pipeline for the accurate contactless measurement of HR from facial videos. For our analysis, we decided to adopt the RMSE and the percent error. The percent error is a useful tool to determine the accuracy of our estimation and is defined as:

$$\boldsymbol{percent\ error} = \left|\frac{\boldsymbol{observed\ value - true\ value}}{\boldsymbol{true\ value}}\right| * \mathbf{100} \qquad \textbf{(Equation 1)}$$

Accordingly, the RMSE of the estimates is a measure of variability derived from the squared residuals and is calculated using the Equation 2, where y_i represents the *ith* predicted value, $\hat{y}_i$ is the actual value for the *ith* observation, $y_i - \hat{y}_i$ is the residual and n is the sample size:

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}} \qquad \textbf{(Equation 2)}$$

5.1.1 Region of Interest

The first question studied relates to optimal ROI selection. In addressing this question we have specifically focused on comparing the face and the cheek as potential ROIs to be used for HR estimation. According to previous published research the most commonly selected ROIs are the face, followed by the forehead and the cheek (Table 4). The forehead is a good candidate, due to the fact that the optical reflectance from this region is typically strong because of the relatively thin skin covering the skull combined with a higher density of blood vessels. However, based on the experimental data on our hands, the forehead may not always be a visible part due to occlusion by hair. Concerning the cheek and the face region (Figure 15), both are reported to show good PPG signal quality. The cheek, being a smaller region, leads to shorter computation time but, on the other hand, the face contains more skin pixels for the extraction of the pulse signal.

In order to test these alternatives and obtain evidence in relation to the question concerning the selection of the most appropriate ROI, we implemented the method described by Poh et al. [12] in MATLAB environment. The evaluation was performed on two different datasets, the publically available MAHNOB-HCI and the self-recorded dataset specially designed and acquired to support the experiment, both described in Section 4. From the self-recorded videos we have excluded the 4 videos recorded before and after the running session (1 participant) since they were acquired in order to support the noise/artefact removal experiment (section 5.1.3).

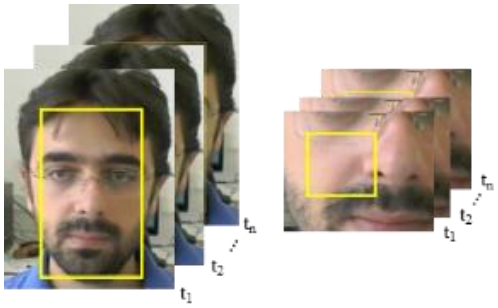


Figure 15: *The face (left) and the cheek (right) as potential ROIs (published with subject's permission)*

The comparative results presented in Table 7 indicate that face ROI provides smaller error when compared to the cheek although the differences are relatively small. Both these ROIs produced similar results in each of the datasets leading us to the conclusion that even though the face region is slightly better, the cheek region can be also used especially in cases where the computation time needs to be shorter. Regarding the remaining sub-questions in this section (5.1), both these ROIs will be adopted depending on the task.

Table 7: *Estimated accuracy of ROIs tested on self-recorded videos and Mahnob-HCI dataset*

Dataset	ROI	Number of subjects	Number of recordings	Reference	percent error	RMSE
Self-recorded videos	**Face**	6	12	Finger pulse oximeter	**1.63 %**	1.58
Self-recorded videos	Cheek	6	12	Finger pulse oximeter	1.91 %	1.68
Mahnob-HCI	**Face**	21	50	ECG	**23.87 %**	22.02
Mahnob-HCI	Cheek	21	50	ECG	24.61 %	23.35

Moreover, the significant increase in error when the MAHNOB-HCI dataset was used, implies that this dataset would be rather inappropriate for the needs of our main analysis, in order to evaluate the overall accuracy of our computational pipeline and its methods.

5.1.2 Signal Preprocessing

The second question we addressed during our exploratory study relates to identifying the most appropriate method for the formation of the raw signal traces. The most popular published methods measure the average brightness inside the ROI by spatially averaging the pixel values for each color channel (RGB) in each frame. A slightly different approach could refer to spatially averaging the pixel values excluding outliers. In this way we make the algorithms more robust to inaccuracies due to errors in face or landmarks detection and tracking, which might produce slight errors in the proper boundary of the selected ROIs in some frames. The calculation of the pixel values, excluding outliers, used in this work was performed by using the trimmed mean MATLAB function, hereafter referred as trimmean, in order to compute the 10% trimmed mean of the samples where the lowest 10% and the highest 10% values were discarded.

The evaluation was performed on the MAHNOB-HCI and the self-recorded dataset excluding the videos recorded before and after the running session, as in the previous section (5.1.1). The comparative results presented in Table 8 indicate that trimmean method provides smaller or the same in one case error when compared to the mean method.

Table 8: *Estimated accuracy of the trace formation methods tested on self-recorded videos and the Mahnob-HCI dataset*

ROI	Dataset	Trace formation method	Number of subjects	Number of recordings	Reference	percent error	RMSE
Face	Self-recorded videos	Mean	6	12	Finger pulse oximeter	1.63 %	1.58
	Self-recorded videos	**Trimmean**	6	12	Finger pulse oximeter	**1.51 %**	1.44
	Mahnob-HCI	Mean	21	50	ECG	23.87 %	22.02
	Mahnob-HCI	**Trimmean**	21	50	ECG	**23.79 %**	22.13
Cheek	Self-recorded videos	**Mean**	6	12	Finger pulse oximeter	**1.91 %**	1.68
	Self-recorded videos	**Trimmean**	6	12	Finger pulse oximeter	**1.91 %**	1.68
	Mahnob-HCI	Mean	21	50	ECG	24.61 %	23.35
	Mahnob-HCI	**Trimmean**	21	50	ECG	**23.09 %**	21.77

Another possible approach for improving performance could be the adoption of a suitable method to detrend the RGB signals before decomposition. The goal of detrending is to remove oscillations in the signal due to external factors, such as movement, in order to obtain the periodic component and a signal with a constant mean. The removal of the trend in the signal was performed with two different approaches. In the first one, a conventional linear detrending technique was applied to the traces, while in the second, a detrending filter based on a smoothness priors approach [59] was used for reducing the slow non-stationary trend of the signal. In order to remove the linear trends, the method computes the least-squares fit of a straight line (or composite line for piecewise linear trends) to the data and subtracts the resulting function from the data. The linear detrending is based on first order or higher order polynomial but detrending with higher order polynomials usually suffers from overfitting issues [60].

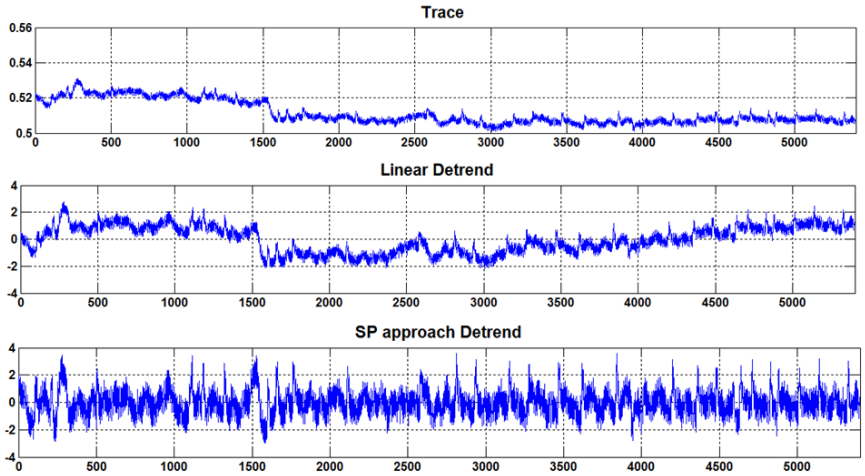


Figure 16: *An example of linear detrending and detrending based on SP approach*

Moreover, detrending by polynomial fitting can have negative effects on spectral analysis techniques and thus an alternative method for removing the slow non-stationarity trends from the

signal before the analysis was tested. The alternative method is a procedure based on smoothness priors approach, hereinafter referred to as "*SP approach*", in which the frequency response can be adjusted adequately to different situations by a single parameter. The smoothness parameter λ was in our case set to 2000 in order to allow only very low-frequency components of the signal to be removed while avoiding to damage high-frequency information. An example of linear detrending and detrending based on *SP approach* applied on the initial trace is presented in Figure 16.

The evaluation of these methods was performed on the MAHNOB-HCI and the self-recorded dataset excluding the videos recorded before and after the running session, as previously. The comparative results presented in Table 9 indicate that detrending based on a *SP approach* improves the results slightly and provides smaller or almost the same error when compared to the linear detrending method.

Table 9: *Summary of detrending methods tested on MAHNOB-HCI and self-recorded videos*

ROI - Trace formation method	Dataset	Detrend method	Number of subjects	Number of recordings	Reference	percent error	RMSE
Face Trimmean	Self-recorded videos	**Linear**	6	12	Finger pulse oximeter	**1.51 %**	1.44
	Self-recorded videos	*SP approach*	6	12	Finger pulse oximeter	1.52 %	1.55
	Mahnob-HCI	Linear	21	50	ECG	23.79 %	22.13
	Mahnob-HCI	***SP approach***	21	50	ECG	**19.06 %**	18.50
Cheek Trimmean	Self-recorded videos	Linear	6	12	Finger pulse oximeter	1.91 %	1.68
	Self-recorded videos	***SP approach***	6	12	Finger pulse oximeter	**1.77 %**	1.61
	Mahnob-HCI	Linear	21	50	ECG	23.09 %	21.77
	Mahnob-HCI	***SP approach***	21	50	ECG	**18.08 %**	18.15

In addition, the significant increase in error when the MAHNOB-HCI dataset was used, confirms that this dataset would be rather inappropriate for the needs of our analysis due to a number of constraints, such as variations in illumination and movements of the subjects. As it turned out, this dataset was acquired in conditions of inappropriate illumination and in some cases, with very intense movements of the participating subjects. The motion induced artefacts and the bad performance of the face segmentation and tracking, due to both movements and scarce illumination, may not provide a substantial help in the computational pipeline's methods comparison. However, these results reaffirm the influence of a subject's motion on the accuracy of HR measurement. On the positive side, though, we were able to obtain a set of limitations for the implemented computational pipeline. Specifically, the very high percent error resulted from

the computed HR measurement compared to ECG was mainly due to the fact that the videos were recorded in a relatively dark environment and included illumination variations. The low ambient light conditions and the blink of reflected light from the computer display caused illumination variations leading to signal corruption. Furthermore, the participant's rigid and non-rigid movements introduce motion-induced noise since there were sudden non-uniform movements and in some cases the head rotation was almost 90 degrees.

5.1.3 Signal Decomposition

In recent years, as already described, researchers have begun to explore the use of video imaging, ambient light and PPG for vital signs estimation. While the initial results obtained appear encouraging, there are drawbacks including susceptibility to motion artefacts [13]. The reason for this is the fact that reflected PPG signal acquired from the camera is a mixture of the physiological pulse along with other sources of fluctuation in light due to motion. Motion is reported to affect the accuracy of measurements and BSS techniques have been considered as a solution to overcome this limitation [12], [13]. Nevertheless, the reliability of these methods has not been systematically explored yet, with respect to optical estimation of HR, HRV and respiratory rate.

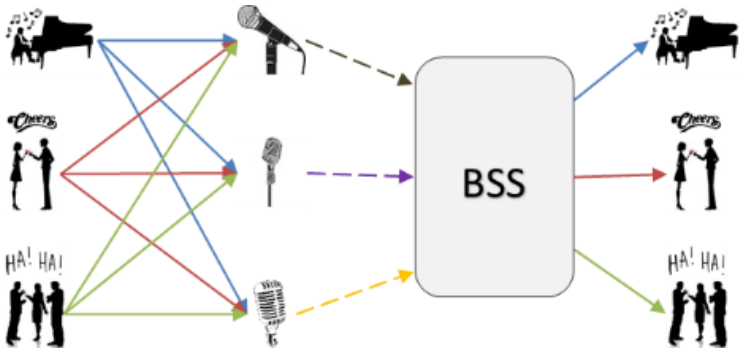


Figure 17: *The Cocktail party problem*

Typically, physiological signals are recorded from a set of sensors, where each sensor collects information of both the signal we wish to analyze and an unwanted noise component. The BSS is one of several methods used for noise or artefact removal from physiological signals [61]. It is a family of signal processing techniques in which a set of observation signals, which are produced by linear mixture of the original signals, are decomposed into a set of source components. In our case, BSS's objective is to separate the desired (e.g. cardiac pulse) signal from noise and artefacts by means of decorrelation and utilization of statistical independence. This process is carried out without any information regarding the mixing model or the source signals and for this

reason the separation is called blind. Based on the assumptions on the source signals and the mixture matrices, different approaches of BSS can be utilized, such as ICA, PCA, Sparse Component Analysis (SCA) and Non-negative Matrix Factorization (NMF). These techniques are used to generate an un-mixing matrix for the estimation of the original sources and consequently the removal of the sources representing the artefacts.

A typical example of the BSS method application is the "cocktail party problem" presented in Figure 17. In this example, the problem is the decomposition of a complex mixture of audio signals coming from different sources (at a real party it can be music, talking, singing, applauding, etc.) and mixed by several sensors (party microphones distributed around the party hall) with weights decreasing with the distance from the sensor to the source of the signal. This problem is solvable given statistical independence of the signal sources.

The mixing and unmixing process involved in BSS is diagrammatically illustrated in Figure 18 [62]. Initially, the independent sources are mixed by the matrix A (which is unknown in this case) and we seek to obtain a vector y that approximates s by estimating the unmixing matrix W. If the estimate of the unmixing matrix is accurate, we obtain a good approximation of the sources. With this spatial separation assumption in mind, we can model the mixing process with matrix multiplication as follows:

$$\boldsymbol{x(t) = As(t)} \qquad \textbf{(Equation 3)}$$

where A is an unknown matrix called the mixing matrix and $x(t)$, $s(t)$ are the two vectors representing the observed signals and source signals respectively.

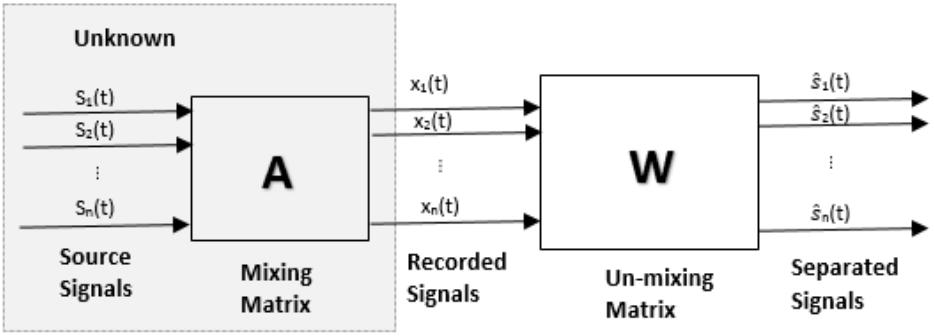


Figure 18: *The BSS block diagram with mixing and unmixing process*

The objective is to recover the original signals, $s_i(t)$, from only the observed vector $x_i(t)$. We obtain estimates for the sources by first obtaining the unmixing matrix W, where, $W = A$-1. This enables an estimate, $\hat{s}(t)$, of the independent sources to be obtained:

$$\hat{\boldsymbol{s}}(\boldsymbol{t}) = \boldsymbol{W}\boldsymbol{x}(\boldsymbol{t}) \qquad \textbf{(Equation 4)}$$

From the most commonly employed BSS method, the ICA decomposes mixtures of source signals into statistically independent, or as independent as possible, components. On the other hand, PCA uses an orthogonal matrix composed by the eigenvectors of the covariance matrix of the original variables in order to identify the principal components, a small subset of which, when some of the original variables are correlated, is able to describe a very large variation of the original data. Both of these methods have been successfully applied for noise or artefact removal to EEG [63]–[65], ECG [66], [67] and functional magnetic resonance imaging (fMRI) data [68]. Good results for the removal of artefacts in EEG were reported in [65] by applying PCA computed via Singular Value Decomposition (SVD) and in [63] by applying ICA computed via the JADE algorithm. He et al. [66] tested an ICA algorithm on three-channel ECG signals showing that this method could effectively detect and remove a considerable amount of noise and artefacts. Correa et al. [68] compared different ICA methods (i.e. Infomax (maximum likelihood) [69], FastICA [70], JADE [71] and eigenvalue decomposition [72]) to study the performance of BSS algorithms for fMRI analysis and reported that each of them had its strengths in specific areas.

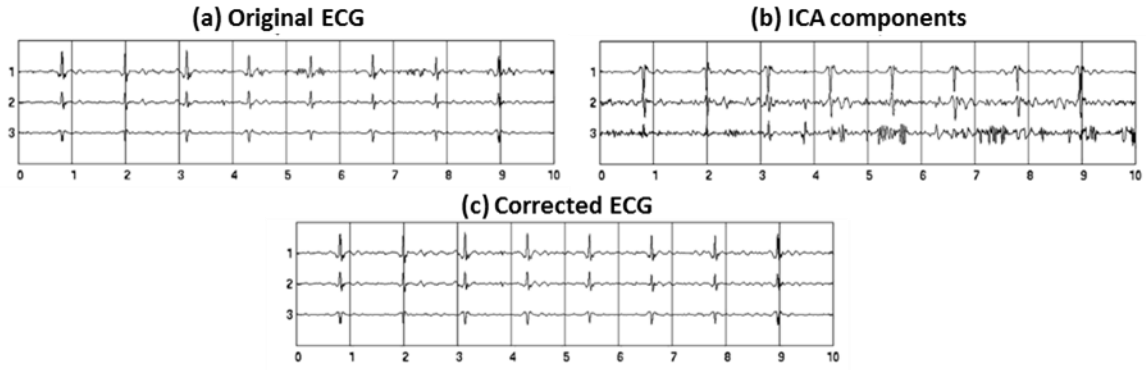


Figure 19: *An example of ECG artefact removal by ICA (taken from* [66]*)*

An example of BSS for removing noise and artefacts from ECG using the ICA is presented in Figure 19 taken from [66]. The ICA algorithm was tested on 3-channel ECG recordings; Figure 19a presents 10 seconds of the original ECG data with channel 1 contaminated with noise, seen as abnormal oscillations either side of the 5th, 7th and 8th QRS complexes. Figure 19b shows the corresponding components derived by ICA with the noise in the original ECG to be separated as ICA component 3, while Figure 19c shows the corrected ECG when the noise component is removed.

For the needs of the present thesis, the question studied and the results reported in this section relate to the optimal BSS algorithm selection for the accurate estimation of HR. The algorithms compared here are JADE, FastICA and RobustICA. These algorithms are variations of ICA (the most commonly used method) with respect to the criteria employed to identify the unique character of the individual unmixed components. The JADE algorithm uses fourth order moments of cumulant matrices in order to diagonalize eigenmatrices, FastICA uses a fixed-point iterative algorithm utilizing negentropy in order to maximize the non-Gaussianity of the sources and RobustICA is a modification of FastICA, which optimizes the kurtosis contrast by using an optimal step size, leading to a higher convergence speed.

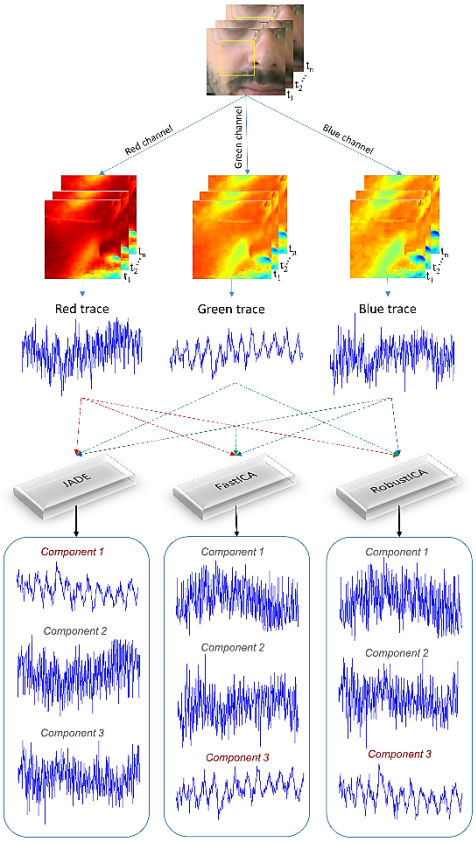


Figure 20: *The processing pipeline leading to the estimation of the Independent Components by applying the JADE, FastICA and RobustICA algorithms*
(published with subject's permission)

Our objective is to perform a comparative evaluation of their accuracy and convergence capability, for the optical estimation of HR. The evaluation of these BSS algorithms for the optical estimation of HR was performed by implementing the method described by Poh et al. [12] in MATLAB environment and the processing pipeline is graphically depicted in Figure 20.

In more details, initially, a ROI was detected for each frame on the right cheek of the participant. The cheek was initially selected as it is a smaller region that could lead to shorter computation time. The video in this ROI was then divided into the three color channels (RGB). The pixel values for each color channel in each frame were spatially averaged and the resulting frame-to-frame time series of the RGB signals formed the three raw traces, which were subsequently de-trended and normalized (zero mean, unit variance). The resulting raw signals contained pulse and motion noise. Thus, subsequent processing was performed for motion artefact removal by applying the three aforementioned ICA-based (JADE, FastICA, RobustICA) BSS methods to decompose the normalized raw traces into independent components. Then, FFT was computed for each component to obtain their power spectrums. The component whose power spectrum contained the highest peak was selected for the HR measurement. Specifically, the frequency with the highest power in the band 0.75-4 Hz (corresponding to 45-240 bpm), was used as the estimated HR. An example of the power spectrum obtained through such analysis of the recorded signal for one subject is shown in Figure 21. The frequency related to the maximum power was 1.333 Hz, which corresponded to HR = 79.98 bpm. The result was in very good agreement with the measured reference HR.

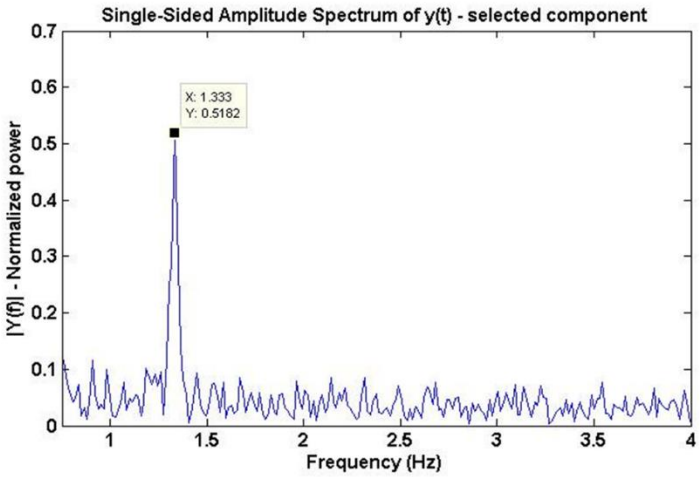


Figure 21: *Result of applying FFT to one of the components produced by the BSS stage*

The preliminary evaluation concerning the question studied in this section was performed on the self-recorded videos obtained from the built-in laptop camera (iSight on an Apple Macbook Pro) and the basic webcam (Logitech 720p HD). The whole dataset and the experimental setup are described in more details in Section 4.1. The HR was computed by following the procedure

previously described and the results were compared with the finger pulse oximeter's measurements.

For our analysis, we decided to adopt the percent error (Equation 1). The results from the two cameras presented in Table 10 indicate that both JADE and FastICA performed well with the FastICA algorithm performing slightly better. However, FastICA, in one of our cases, failed to initialize properly and was not able to produce results, since it is an iterative algorithm that requires proper initialization and setting of parameters. It is also important to mention that the worst results obtained with RobustICA are mainly due to the bad result yielded in just one case. This specific video presented the subject under investigation with heavy head motion, which constitutes a common problem for these algorithms. When we excluded this case from our analysis, the outcomes are comparable with those produced by the other two algorithms. Finally, the discernible increment in the errors, observed in iSight compared to the Logitech, reveals that results are camera dependent. These results were published and presented in the 4th International Conference on Wireless Mobile Communication and Healthcare [14].

Table 10: *Results obtained from videos recorded with iSight and Logitech cameras*

Measured HR	***CAMERA A –ISIGHT (% error)***			***CAMERA B –LOGITECH (% error)***		
	JADE	FastICA	RobustICA	JADE	FastICA	RobustICA
69.00	0.62	0.62	32.49	3.19	3.19	3.19
102.00	53.47	53.47	53.47	0.82	0.82	50.88
63.00	1.84	1.84	1.84	1.84	1.84	1.84
77.00	1.83	1.83	1.83	1.84	0.45	1.84
78.00	2.54	2.54	2.54	2.54	2.54	2.54
69.00	4.45	4.45	4.45	4.45	-	4.45
61.00	3.74	0.58	0.58	0.59	0.59	0.59
76.00	0.54	0.54	0.54	0.54	0.54	0.54
Average % error	**8.63**	**8.24**	**12.22**	**1.98**	**1.42**	**8.23**
Average % error *excluding 102 bpm*	**2.22**	**1.77**	**6.32**	**2.14**	**1.53**	**2.14**

From the results reported, the conclusion is that the optimal BSS algorithm is JADE, since the results from the comparative evaluation on the different ICA algorithms indicated that both JADE and FastICA performed well whereas the FastICA algorithm in one case failed to initialize properly and was not able to produce results. In general, the fact that the choice of BSS algorithm does not significantly influence the estimates is interesting. Percentage errors are lower than 2.5% (6.5% in one case) when the outlier value of 102 bpm is excluded from analysis. It is also noticeable that all algorithms present inadequate response in videos of high HR and notable head motion (related to the recording after physical exercise). This finding needs to be studied more

extensively and its physical origins requires explanation. An assumption may be that the channel selection algorithm requires improvement, especially in cases with low SNR, which in our experiment appears to be the case with the high HR (high head motion), and techniques for enhancing SNR may be applied in the process pipeline before the application of BSS.

5.1.4 Component Selection

Our fourth research question relates to the selection of the optimal component, based on measures obtained from the green channel and the component with the maximum power frequency, hereinafter referred to as "max selection". Based on the results of previous published studies [39], [43] the green channel features the strongest plethysmographic signal. Thus, a further analysis was performed on the raw traces recorded with both cameras bypassing the BSS stage, where the HR was computed by following the procedure previously described in section 5.1.3.

Table 11: *Results obtained using the raw traces bypassing the BSS stage on the videos recorded with both cameras and using the JADE separation technique and the max selection*

Measured HR	bypassing BSS						JADE- *max selection*			
	iSight camera			Logitech camera			iSight camera		Logitech camera	
	Estimated HR	Percent error	Selected channel	Estimated HR	Percent error	Selected channel	Estimated HR	Percent error	Estimated HR	Percent error
69.00	68.55	0.65	Green	66.80	1.90	Green	69.43	0.62	66.8	3.19
102.00	100.19	1.77	Green	71.20	20.64	Red	47.46	53.47	102.83	0.82
63.00	64.16	1.84	Green	64.16	1.84	Green	64.16	1.84	64.16	1.84
77.00	76.46	0.70	Blue	77.34	0.44	Green	75.59	1.83	75.59	1.84
78.00	79.98	2.54	Green	79.98	2.54	Green	79.98	2.54	79.98	2.54
69.00	72.07	4.45	Green	72.07	4.45	Green	72.07	4.45	72.07	4.45
61.00	60.64	0.59	Green	60.64	0.59	Green	63.28	3.74	60.64	0.59
76.00	75.59	1.70	Green	75.59	0.54	Green	75.59	0.54	75.59	0.54
Average percent error		**1.78**			**4.12**			**8.63**		**1.98**

In order to test the alternatives and obtain evidence in relation to the question concerning the most appropriate approaches for component selection, the self-recorded videos were used; for comparing the results obtained, we have used the percent error (Equation 1). The results, reported in Table 11, are in agreement with these studies, as the outcomes show that the green channel was largely more selected. Only in very few cases, better results were obtained through the selection of another channel, implying that a reliable HR can be measured by exploiting only the green channel. Finally, it is wort noting that in few cases there is a very light improvement when using separation techniques, while in some others the percent error remains unchanged.

Based on the results obtained from the present study, we have demonstrated the feasibility of using a simple webcam to estimate HR. It is evident from the experimental data that the type of camera used does have an effect on the observed accuracy of estimation, although the changes are not relevant in the large majority of cases.

Furthermore, this preliminary evaluation showed that the face ROI and the trimmean method yielded the lowest errors and a high agreement was observed across all measures. Nevertheless, we obtained non-conclusive evidence with respect to the most appropriate signal containing the BVP. These results were also published and presented in the 4th International Conference on Wireless Mobile Communication and Healthcare [35].

These results must be considered in the light of several limitations. Such limitations include: a) the inaccuracy of the fingertip pulse oximeter, b) the limited amount of subjects and measures on the available subjects, c) the uncontrolled head motion which may contaminate with high noise the methods (investigations using controlled movements at different velocities should be performed in order to evaluate the capabilities of the various algorithms of filtering the motion artefacts and to understand the amount of movements that are permitted in order not to skew results) and finally d) the distance from camera and camera resolutions may be further studied since they are confounding variables that could potentially influence the results. Our findings therefore, point to the fact that extensive validation and continued systematic exploration of these variables is still required.

5.2 Signal Processing Pipeline

The findings obtained from the exploratory study previously presented (section 5.1) were used in forming the procedure for the HR estimation from video recordings. The pipeline of the proposed camera-based HR monitoring method is graphically depicted in Figure 22.

The results, taking the accuracy of estimation as the parameter to compare with, indicate that the optimal ROI is the face and the most appropriate approach to measure the average brightness inside this ROI is the trimmean method that excludes outliers. Concerning the removal of the trend in the signal, the detrending method based on a *SP approach* is selected and for the BSS algorithm the optimal method is the ICA based on JADE.

Finally, regarding the component selection approach, we obtained non-conclusive evidence with respect to the most appropriate signal containing the BVP and the findings will be

further verified. In order to configure the HR measurement computational pipeline, the FFT of each resulting component is computed to obtain the corresponding amplitude spectrums and the component whose power spectrum contains the highest peak is identified as the component containing the cardiac signal.

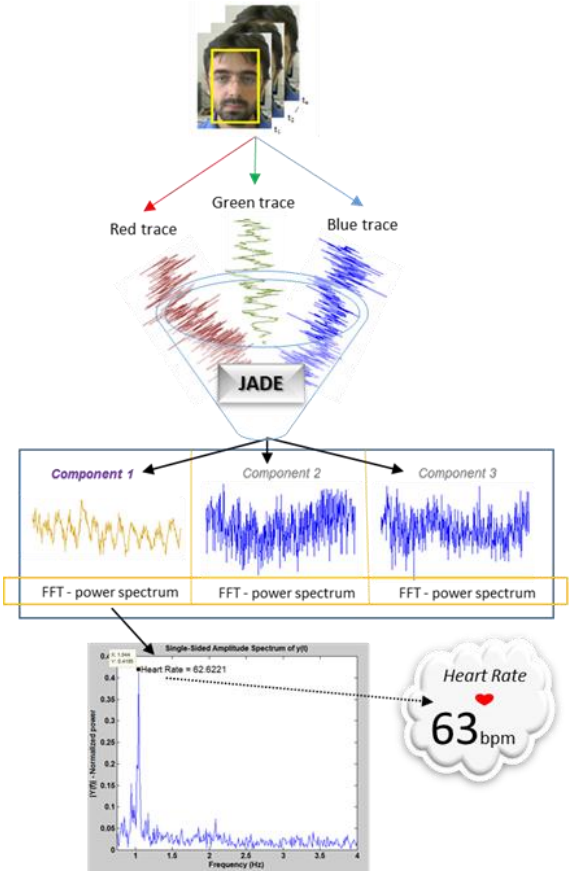


Figure 22: *The HR measurement processing pipeline (published with subject's permission)*

In more details, the Viola-Jones algorithm [73] is initially used to detect the face and localize the measurement ROI within the first video frame. The algorithm returns the x and y-coordinates along with the height and width that define a box around the face. From this output, the measurement ROI is defined as the rectangle enclosing the detected face reduced by 40%, in order to exclude pixels with background noise. This decision, i.e. to scale down the estimated ROI, was based on published evidence that reduction of the ROI area has little effect on the HR error. Specifically, it has been reported that reduction in the ROI from 100% through to 40%, the percentage of data representing HRs with respect to the reference value is close to 100% [53]. Therefore, the ROI used for our estimation is the rectangle defined as the central 60% of the width and full height of the rectangle containing the face region (Figure 15-left image). Inside this ROI,

feature points are detected and tracked by employing the Kanade-Lucas-Tomasi (KLT) algorithm [74].

The video signal corresponding to this ROI is then processed for the three color channels (RGB) and the pixel values for each color channel in every frame are spatially averaged, excluding outliers (10% trimmed mean). The resulting frame-to-frame time series of the RGB signals form the three raw traces, which are subsequently de-trended and normalized so that they exhibit a zero mean and unit variance. The final raw signals, thus obtained, contain pulse, motion noise and potential variations due to environmental lighting. Therefore, subsequent processing is performed for artefact removal by applying the ICA based on JADE algorithm to decompose the color channels into three independent source signals. This method has a major drawback since the output components are returned in a random order and therefore, a key challenge is to determine a method for selecting the optimal component for computing the HR. In order to identify the component containing the cardiac signal, the FFT of each resulting component is computed to obtain the corresponding amplitude spectrums.

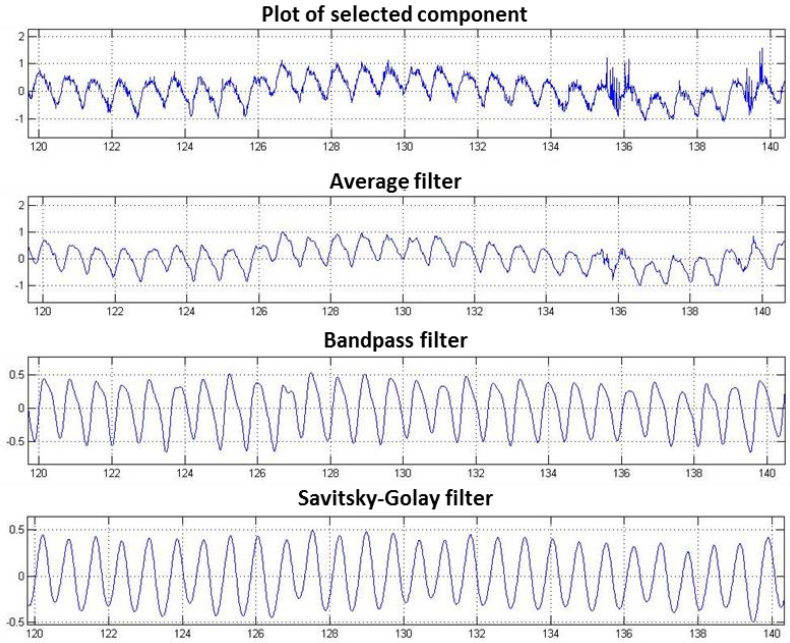


Figure 23: *Example of filtering and smoothing process on the selected component*

Subsequently, the component whose power spectrum contains the highest peak is selected and the pulse frequency is estimated as been the frequency that corresponds to the highest power of the selected component's spectrum within an operational frequency band range of 0.75-4 Hz (corresponding to 45-240 bpm). This frequency is finally used as the estimated HR.

The method was extended in order to measure HRV and respiratory rate by further processing the component identified as the signal containing the best replica of the BVP signal (Figure 23). More specifically, the selected source signal was smoothed using a five-point moving average filter and bandpass filtered using 128-point Hamming window (0.7–4 Hz). Then, the Savitzky-Golay filter was applied to further smooth the signal and refine the BVP peaks. The reason for selecting Savitzky-Golay amongst other filters with comparative characteristics was the fact that, when it is appropriately designed to match the waveform of an oversampled signal corrupted by noise, it tends to preserve the width and height of peaks in the signal waveform [75]. For this filter the 4th order polynomial fit and a window size equal to the fps was the most efficient in our case.

Then, peak detection was performed (Figure 24) using a custom algorithm for maxima detection with a minimum peak-to-peak distance that removes peaks differing less than 30% from the median of their intervals over the whole signal in order to avoid detecting double peaks due to the nature of the signal. From the time interval between consecutive peaks we extracted the IBI series to be used for the computation of the HRV.

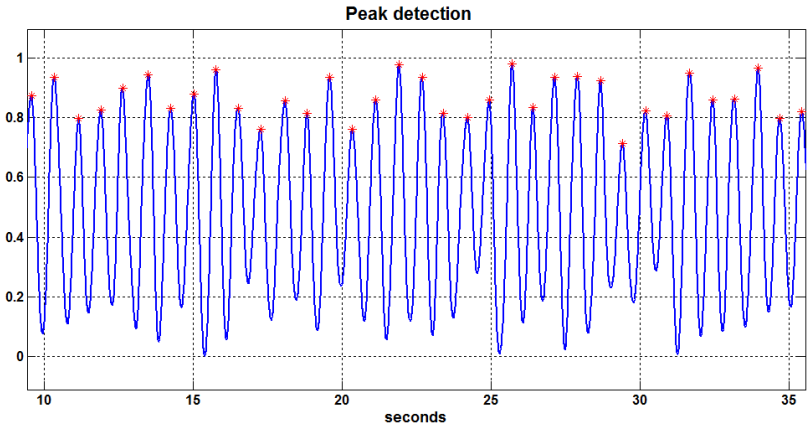


Figure 24: *The result from peak detection*

All HRV parameters are calculated on NN IBIs caused by normal heart contractions paced by sinus node depolarization. Thus, before such calculation all abnormal heartbeats and artefacts must be removed from consideration. In our case, before the feature extraction, the IBI needs to be corrected to prevent artefacts due to ectopic beats or motion. Especially in case of video-based PPG signals the interval correction is important, since these signals are very sensitive to motion artefacts that may cause unreliable HRV feature extraction. A common practice for the IBI correction is to adjust the inter-beat tachogram in order to remove the effect that abnormal beats

would have on HRV estimation. Hence, the IBI tachogram was filtered using an impulse rejection filter [76] for reducing the impact of artefact on the estimated PSD of IBI series.

Afterward, the HRV signal was constructed and its analysis was performed in both time and frequency domain. In this task, since it was unrealistic to have the subject in front of the camera for more than a few minutes, the analysis of HRV was performed, under the hypothesis of short-term recording. The time domain extracted features were the SDNN and the RMSSD and were calculated directly from the IBI series. In the frequency domain, the analysis was performed by PSD estimation using the Welch periodogram. This method requires the analyzed time signal to be stationary and evenly sampled. In order to remove any non-stationarities within IBI time series, a linear detrending was performed before HRV analysis. Moreover, the cubic spline interpolation was applied with a 10 Hz interpolation frequency in order to equidistantly sample the data. Then, the LF and HF powers were measured as the area under the PSD curve corresponding to 0.04-0.15 Hz and 0.15-0.4 Hz, respectively. Finally, the extracted features were the LF power, the HF power, the LF power in normalized units, the HF power in normalized units and the LF/HF ratio. In addition, the respiratory rate was calculated from the center frequency of the highest peak between 0.15-0.4 Hz of the HRV power spectrum and the HR was also extracted by counting the number of BVP peaks per minute or calculated as $60/\overline{IBI}$, where the $\overline{IBI}$ is the mean of the interbeat intervals.

5.3 ECG Signal Processing

Assessment of HRV requires the extraction of RR intervals time series, which is usually performed by detecting R peaks from digitized ECG data. A main problem of signal digitization is its interference with other noisy signals like power supply interference, patient motion artefacts, electrode contact noise and baseline wandering. Among these noises, the power line interference and the baseline drift are the most significant and can strongly affect ECG signal analysis. The power line interference is a narrow-band noise cantered at 50 Hz (or 60 Hz) with a bandwidth of less than 1 Hz. On the other hand, the baseline drift usually originates from respiration at frequencies ranging between 0.15 and 0.3 Hz. These noisy elements have to be removed before the signal is used for the HR and HRV estimation.

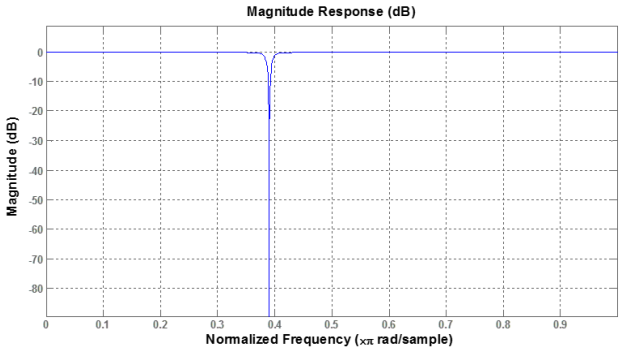


Figure 25: *An IIR notch filter with the notch located at 50 Hz or 0.3906π radians per sample*

Since power line interference occurs at a frequency of 50 Hz, we have implemented an IIR notch filter able to remove this high frequency noise (Figure 25). The notch filter is characterized by the sharp attenuation dip in frequency domain. This filter type is generally used to remove a disturbance of a known narrow band frequency.

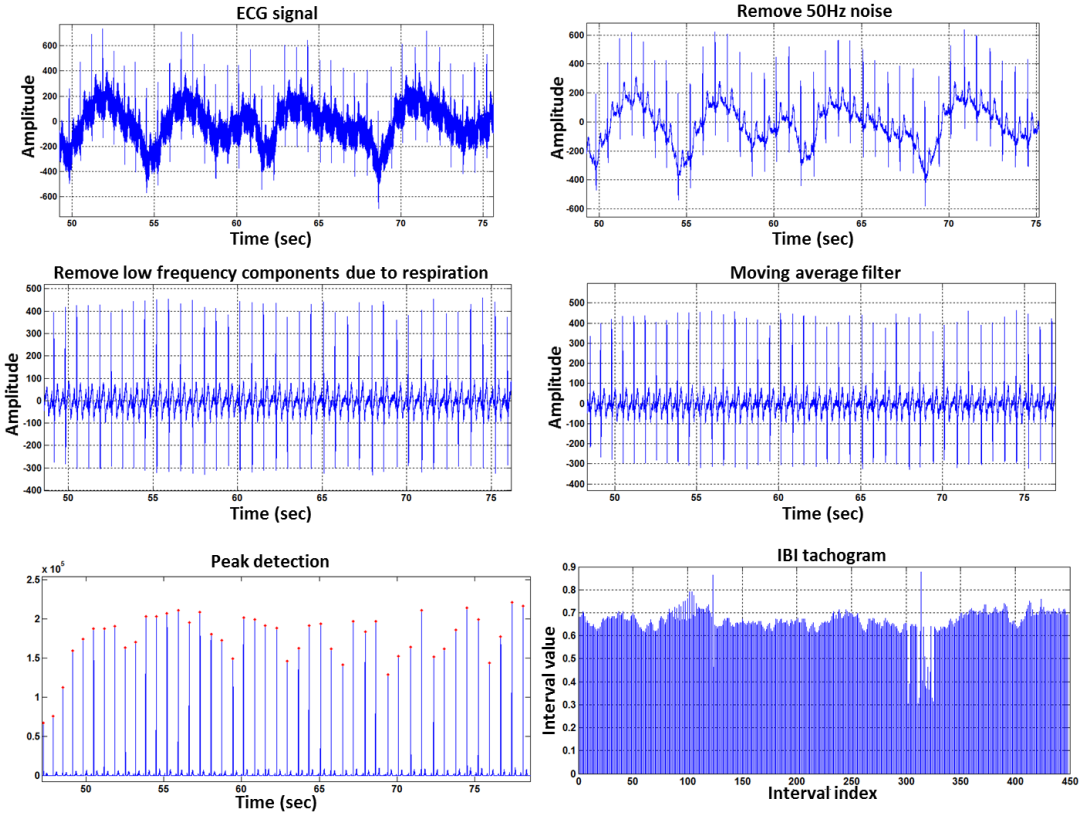


Figure 26: *Example of ECG signal pre-processing and RR intervals extraction*

We also implemented an IIR 2nd order Butterworth high pass filter (cut-off frequency = 0.5 Hz) to remove low frequency components due to respiration and at the same time to preserve the interesting ECG frequency components. After removing the baseline wandering by eliminating

the trend, the resulting signal is more stationary than the original. Then, the signal was adjusted to have zero mean by using a running mean filter and subtracting the smoothed data from the signal. Further process included the enhancement of the R peaks to make them easier to detect by using the square magnitudes of the data. Afterwards, the extraction of the RR intervals from the filtered signal was performed. Figure 26 shows an example of the original ECG signal, the resulting signal processed by the notch and high-pass filter, the detection of R peaks and the extracted RR intervals. Finally, the HRV analysis was performed, following the same procedure as for the IBI signal previously described in section 5.2.

6 Experimental Validation

Following the exploratory analysis presented in the previous chapter and in order to verify the findings obtained as well as in evaluating the overall accuracy of our computational pipeline and its methods, we engaged in a subsequent study, employing the SEMEOTICONS reference dataset that was acquired with the ECG recorded in parallel with the videos and was made available in due time by the SEMEOTICONS project.

The data processing algorithm were configured in order to measure the HR, the HRV and the respiratory rate and we utilized this dataset to address the main research question, i.e. accuracy of their contactless estimation. This chapter reports on our findings, utilizing one of the largest, if not the largest, dataset available, in evaluating the adequacy, in terms of measurement accuracy, of camera-based cardiac information monitoring. The ECG signals from this dataset provided the reference HRs, used to determine the accuracy of our method. The performance of our approach was calculated by computing the percent error (Equation 1) and the RMSE (Equation 2) between the estimated HR from the video-based PPG and the reference HR from the ECG.

6.1 Heart Rate Estimation Analysis

At first, for a more extensive validation on the channel selection methods (see Section 5.1.4), the SEMEOTICONS reference dataset was utilized, as it includes more cases than the dataset used in our exploratory study and includes the reference HR extracted from the recorded ECG signals. In order to validate the effectiveness of the max selection algorithm, we further explored the three ICA outputs to manually identify and select the component containing the BVP signal. For this reason, the HR was estimated from each of the three ICA outputs using the pulse frequency that corresponds to the highest power of each component's spectrum. Then, we compared the results produced from each ICA component with the ECG results. The most appropriate component was defined as the component leading to the most precise estimation of the HR. Thus, the component with the minimum percent error was selected as the best component for comparison and this method hereinafter referred to as "manual selection".

The results presented in Figure 27 indicate that even though the max selection method's results showed reasonable accuracy in most cases, the overall error produced by this method in each trial was higher than the error produced by the manual selection method showing that this

method in some cases probably failed to select the appropriate component. The exact cause of failure need to be identified.

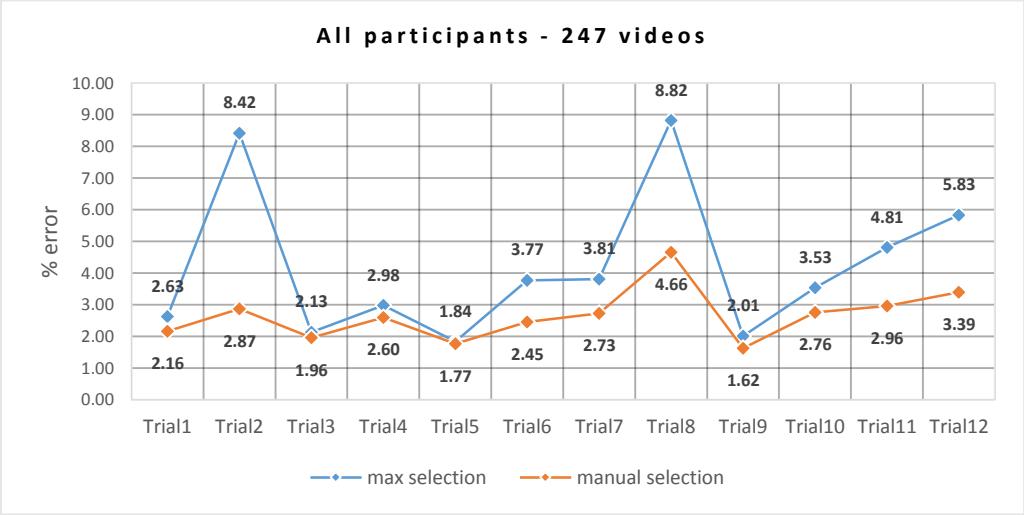


Figure 27: *Max and manual component selection method comparison (all participants)*

As it can be seen from the above graph, the trend for the max selection method is similar to the manual selection method and the worst results for both methods are observed in the Stroop test task (trial 8). The performance of both methods deteriorates for this talking scenario, which is probably due to the non-rigid deformation observed on the face during expressions. A very high percent error (8.42%) is also observed in the self-describing task (trial 2) for the max selection method. These results state that this method leads to poor performance during talking scenarios, where a large portion of the face goes through non-rigid deformation. However, in more normal conditions, in which the recordings contain rigid and non-rigid deformation in a small portion of frames, such as in the mimicking fatigue state (trial 6), the produced error is smaller (3.77%).

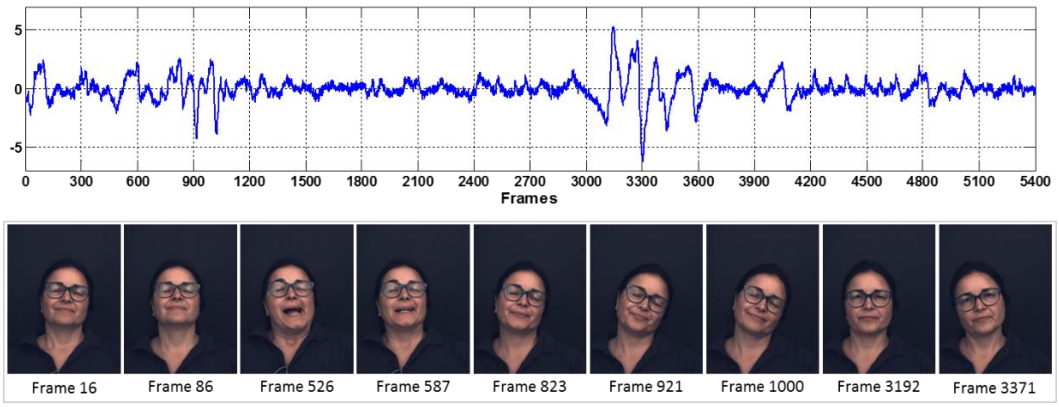


Figure 28: *Video frames and the raw trace from trial 6 (mimicking fatigue) showing extreme moves and intense facial expressions*
(published with permission of the subject)

An example, from trial 6, of motion artefacts due to extreme subject motion and intense facial expressions as displayed in video frames and in the recorded signal can be seen in Figure 28. The corresponding raw trace shows that head motion and facial expressions add noise variations in the reflected light from the face which are captured as sudden changes in the recorded signal. However, in this particular case, the ICA managed to separate properly the sources and the actual HR was successfully extracted from the signal.

Eye blinking constitutes another type of artefacts that can also influence the results. In some of our signals, eye blinking artefacts were observed in the form of peaks that constituted components which were not related to the signal of interest. Exported frames corresponding to the signal peaks reveal that the recorded signal was related to ocular artefacts and the observed peaks were produced by the eye blinks. In Figure 29, the presented signal and the frames were taken from a video recorded during trial 1 (reference period) where the subject was asked to keep still in a neutral state. In this video, the actual blinks, after visual inspection, were 29 and occurred at the time where the peaks occur in the corresponding signal. In this particular case, the ICA, as a noise cancelation algorithm for eye blink artefacts, was successfully applied and the actual HR was extracted from the signal. However, in recording scenarios that include motion, the eye blinking and/or the mouth openness provided low performance and negatively affected the signal. Thus, the exclusion of the eye and mouth region, when the whole face is used as the ROI, could possibly lead to improvement of the HR estimation under motion scenarios.

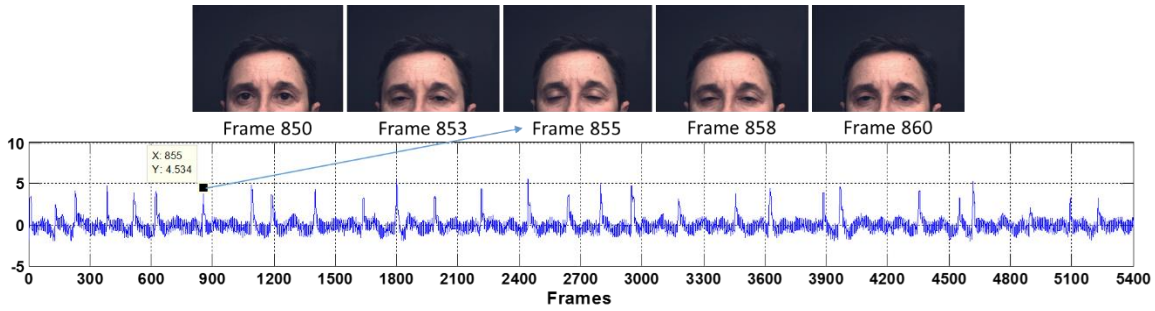


Figure 29: *Example of eye blink artefacts in the form of signal peaks*

Overall, based on the results presented in Figure 27, there is an error increase observed between trial 2 (8.42%) or trial 8 (8.82%), in which a higher level of motion was observed and trial 3 (2.13%) or trial 9 (2.01%), where the subjects demonstrated the least motion. In these cases, the error appears to increase due to extreme movements and intense facial expressions. However, although this is true, when extreme move and the intense facial expressions appear only in some

parts of the video as in the case of trial 6, the error falls to 3.77% for the max selection method. Moreover, moderate motion and non-extreme changes in facial expressions of the subject are present in all dataset and appear to be well tolerated by our analysis pipeline for the HR estimation.

As discussed in section 5.1.3, motion can affect the accuracy of measurements and BSS techniques have been considered as a solution to overcome this limitation. However, based on the results presented in Figure 27, an error from 1.62% to 4.66% exists even when the best ICA output component is selected by the manual selection method. For example, in trial 8 for the manual selection method the error was of the order of 4.66% showing that even though the selecting component was the one (out of three) leading to the most precise estimation of the HR, the percent error was still high. This imply that the ICA based on JADE may sometimes was not able to separate properly the sources. In order to test this assumption and also further verify our findings from the signal decomposition (5.1.3) and component selection (5.1.4) sections, an analysis was performed on the videos from trial 9, since they were evaluated as being mostly suitable for the task. These videos, apart from good diffuse illumination, exhibit limited movements of the participating subjects as specified by the corresponding recording scenarios. However, in some cases, the collected data still contained spontaneous movements and facial expressions. Moreover, the recording conditions are similar to the videos from the self-recorded dataset used in the previous sections (5.1.3 and 5.1.4) and their duration is the same. In this trial, two recordings had issues, e.g. non-availability of the ECG signal due to technical problems during the recording, which forced us to exclude them from the analysis. Hence, our analysis was based on 22 recordings from 22 participants, including the two dark skinned male subjects, leading to a total of approximate 22 minutes of video data.

The comparative results presented in Table 12 indicate that the selection of a single appropriate component may not be necessary for estimating the correct HR, since, in some cases, a correct HR may also be accessible form other components (e.g. case 3). This result confirms recent research findings (see [54]).

Moreover, the selected component may not always be the same since it appears that in some cases (e.g. case 20) the ICA-based BSS method was not able to separate properly the sources but produced mixed-source signals in the output (Figure 30).

The evaluation of the performance of the three output components of ICA in estimating actual HR also shows that each component could estimate the HR better than others on different

occasions. Probably this mean that we may cannot rely on one component to estimate the HR even for the same participant.

Table 12: *Max and manual component selection using Trial 9 from the SEMEOTICONS reference dataset*

Trial 9	Max selection		Manual selection		Comments*	Correct HR component
Videos	HR	Comp.	HR	Comp.		
Case 1	1.41	3	1.41	2-3	Component (S) - HR (S)	More than one
Case 2	0.00	2	0.00	2	Component (S) - HR (S)	One
Case 3	0.00	2	0.00	1-2-3	Component (S) - HR (S)	More than one
Case 4	0.00	2	0.00	1-2-3	Component (S) - HR (S)	More than one
Case 5	3.75	3	3.75	2-3	Component (S) - HR (S)	More than one
Case 6	2.78	2	2.78	2-3	Component (S) - HR (S)	More than one
Case 7	1.72	2	1.72	2-3	Component (S) - HR (S)	More than one
Case 8	3.70	2	3.70	2-3	Component (S) - HR (S)	More than one
Case 9	1.43	3	1.43	3	Component (S) - HR (S)	One
Case 10	1.43	3	1.43	3	Component (S) - HR (S)	One
Case 11	0.00	2	0.00	2-3	Component (S) - HR (S)	More than one
Case 12	1.27	3	1.27	1-2-3	Component (S) - HR (S)	More than one
Case 13	1.72	3	1.72	2-3	Component (S) - HR (S)	More than one
Case 14	0.00	2	0.00	2	Component (S) - HR (S)	One
Case 15	1.67	3	1.67	2-3	Component (S) - HR (S)	More than one
Case 16	1.75	2	1.75	2	Component (S) - HR (S)	One
Case 17	6.35	2	1.59	3	Component (D) - HR (D)	One
Case 18	0.00	2	0.00	2-3	Component (S) - HR (S)	More than one
Case 19	1.25	2	1.25	2-3	Component (S) - HR (S)	More than one
Case 20	5.33	1	2.67	2-3	Component (D) - HR (D)	More than one
Case 21	0.00	1	0.00	1-2	Component (S) - HR (S)	More than one
Case 22	8.75	2	7.50	1	Component (D) - HR (D)	One
Average % error	**2.01**		**1.62**		**Comments show whether the component selected and the measured HR was the same (S) or different (D)*	
RMSE	**2.21**		**1.75**			

One of the known limitations of ICA is that it demands the same number of observers as are the number of signal sources, i.e. if there are three people talking at the cocktail party and only two recordings, one cannot separate all the sources from each other. Furthermore, when different noise factors exist, the number of signal sources will be unknown and may be greater than the number of recordings. This means that the signal sources we wish to record (HR) cannot be separated perfectly in this cases. A solution could be to remove as many noise and unwanted signal sources as possible and/or increase the number of video recordings e.g. more cameras or

using a five band camera. However, in most cases this is not a good solution due to the cost and increased complexity.

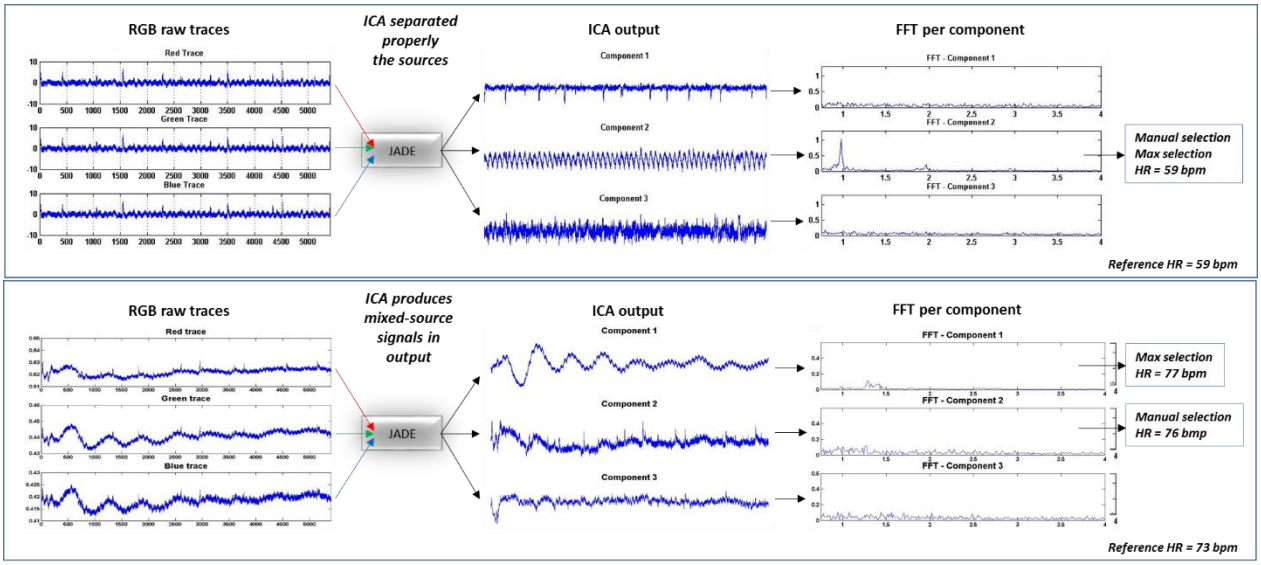


Figure 30: *ICA-based BSS separated properly the sources or produced mixed-source signals in output*

Another major challenge in estimating PPG using a camera is the extremely low signal strength of the color-change signal, particularly for darker skin tones as it is claimed in [52]. In order to test this assumption, the recordings of the dark skinned participants were excluded from our analysis. The results presented in Figure 31 indicate that the accuracy of the estimation is significantly influenced when dark colored subjects are involved. In such case, the measures represent an increase for both methods which confirms the previous research findings. An exception is observed in case of trial 5 which does not include dark skinned participants' recordings in the whole dataset and thus the error is the same in both cases. Moreover, the manual selection method seems to produce a peculiar result in case of trial 2. The results are slightly better when the dark skinned participants are included than when excluded because of the precise estimation of HR from these participants.

In more details, in cases where the subjects demonstrated the least motion and non-extreme changes in facial expressions (trial 1, 3, 9) both methods show the lowest error differences. However, in cases where other sources of artefacts are included e.g. in trial 8 where a higher level of motion is present, the manual selection method showed higher improvement when dark skinned participants were excluded. This observation, reveals that the BSS method has a poor performance for people having darker skin tones, in the presence of facial expression and motion between the

camera and the subject whilst in relatively motionless scenarios our method can address the aforementioned challenge.

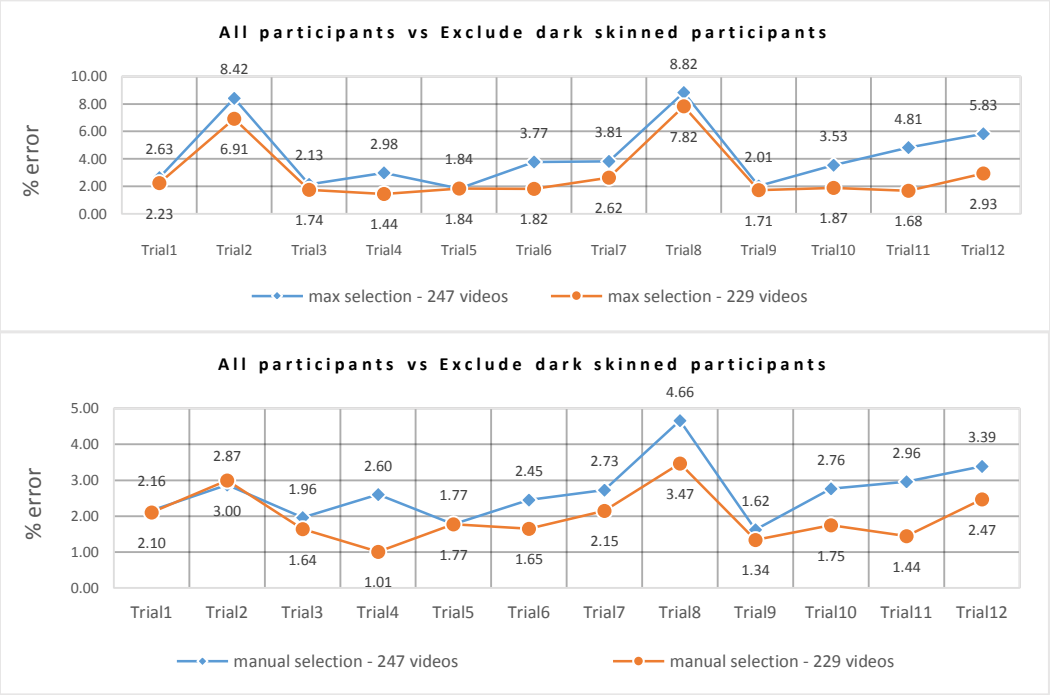


Figure 31: *Max and manual component selection method comparison for all participants and without dark skinned participants*

Although it is known that camera-based vital sign monitoring does not perform well for subjects having darker skin tones, the recording conditions also influence the results. In the case of the SEMEOTICONS dataset, the choice of using of a black backdrop (Figure 32) placed behind participants to minimize reflection and provide a constant background scene may have influenced the results.

Figure 32: *Dark skinned subject in black background (published with subject's permission)*

As we have previously mentioned, for our test, the Viola-Jones algorithm [73] was used to detect the face and localize the measurement ROI within the first video frame. This algorithm

returns the location of a rectangular bounding box containing the face in a relatively clean background. For the dark skinned subjects in our dataset the black background may produce a false detection by the means of providing a bounding box containing many background pixels that causes an influence on the results, even in our case where the ROI is slightly reduced. A proposed solution could be the use of white background when the skin is dark and a black background in all other cases during the recordings.

The next step performed was a thorough validation of the HR measured from the processing of facial RGB videos when compared to classical ECG measurements. The analysis was performed on the whole SEMEOTICONS dataset which, as previously reported, contains 247 recordings and a total of approximately 355 minutes of video data. The agreement between 247 pairs of measurements from 24 participants was tested by Bland-Altman analysis (Figure 33 - left) and the mean bias $\bar{d}$ was 2.5 bpm with 95% limits of agreement −11.3 to 16.3 bpm.

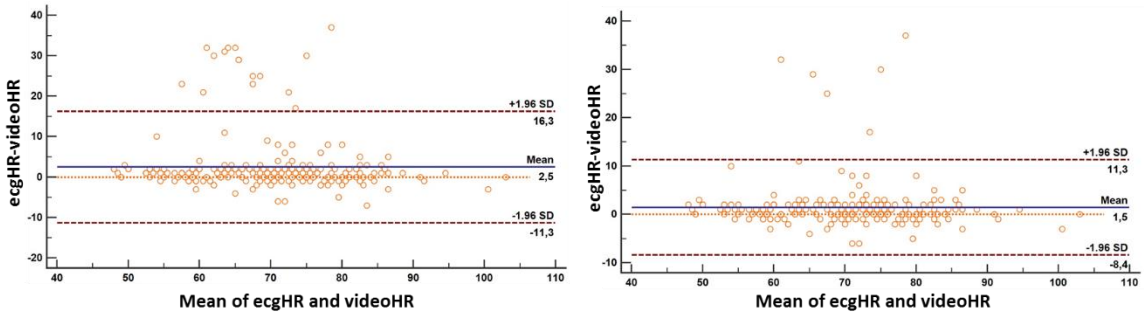


Figure 33: *The agreement between HR measurements obtained from ECG and video (Left: 247 measurement pairs from 24 participants – Right: 229 measurement pairs from 22 participants)*

Subsequently, the recordings of the dark skinned participants were excluded from our analysis; thus, our dataset, as previously reported, contains 229 recordings and a total of approximately 329 minutes of video data. Again, the agreement between 229 pairs of measurements from 22 participants was tested by Bland-Altman analysis (Figure 33 - right) and this time the mean bias $\bar{d}$ was 1.5 bpm with 95% limits of agreement −8.4 to 11.3 bpm. Although there are some outlier points located outside the 95% limit of agreement interval, an accurate estimation of the HR, derived from the 229 pairs of measurements, is shown to be feasible in most cases. Visual inspection of the few videos producing deficient estimates point to the fact that these videos include extreme responses from the participants to the presented stimuli, e.g. sudden facial expressions accompanied by nodding.

The HR was also extracted by counting the number of BVP peaks per minute or calculated as $60/\overline{IBI}$. These approaches were tested using the whole SEMEOTICONS reference dataset (with and without dark skinned participants). In Figure 34, the results from all participants for both approaches are presented graphically. The scatter plot comparing the ECG and video-based HR measurements from all trials shows a strong relationship between those two measures. In the case of HR calculated from peaks, the correlation coefficient was 0.97 (P<0.0001 and 95% CI 0. 9578 to 0.9742) while in case of HR calculated from IBI the correlation coefficient was 0.96 (P<0.0001 and 95% CI 0.9505 to 0.9697).

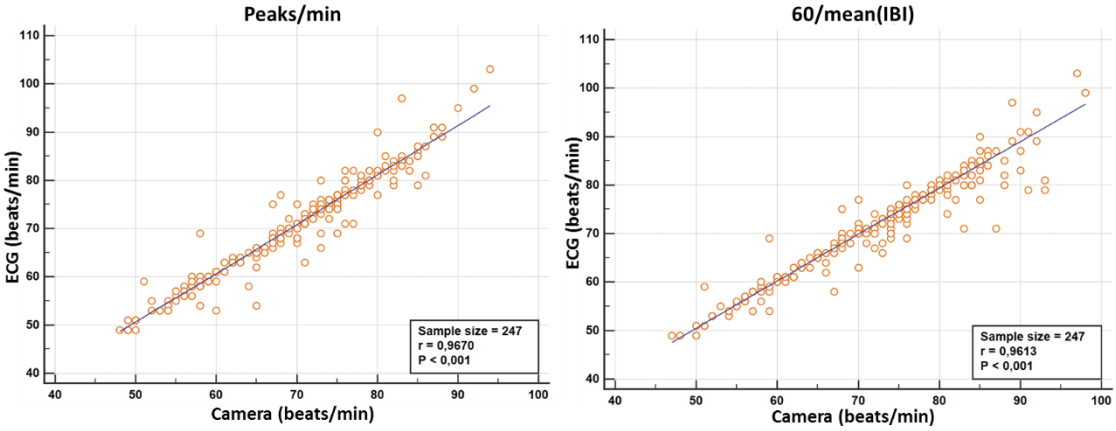


Figure 34: *Scatter plot comparing the ECG and video-based HR measurements which shows a strong relationship between those two measures*

Finally, the comparison of the results obtained from the previous methods (the frequency of the highest power of the spectrum – max selection) and those two (number of BVP peaks and $60/\overline{IBI}$) are presented in Table 13. When the dark skinned participants were excluded, in most of the cases, the estimates from the two later methods lead to a reduction of the percent error by almost 50%. The results indicate that the average percent error across all 229 measures decreased from 4.22% to 2.88% for the max selection method, from 2.35% to 2.21% for method based on the number of BVP peaks and from 2.28% to 1.76% for the method based on the mean IBI.

Comparing the errors produced from these approaches, the max selection approach appears to be mostly affected from the participant's color while the number of peaks approach tends to yield similar results. Furthermore, the average percent error was slightly increased in all three approaches not only when the dataset contains participants with dark skin but also when the dataset contains movements. According to these results, the method based on the mean IBI is the most

appropriate approach since it produced adequate results in most of the cases and generated overall the smallest percent error.

Table 13: *The % error and RMSE produced from different approach for HR measurement using the SEMEOTICONS reference dataset*

		247 measures						229 measures					
		Max selection		# peaks		IBI		Max selection		# peaks		IBI	
Trial	Duration	% error	RMSE	% error	RMSE	% error	RMSE	% error	RMSE	% error	RMSE	% error	RMSE
1	1 minute	2.63	2.85	2.97	2.86	2.04	2.46	2.23	2.33	2.92	2.89	1.48	1.86
2	1 minute	8.42	11.62	3.86	3.90	3.12	3.03	6.91	9.62	3.81	3.89	3.04	2.98
3	1 minute	2.13	2.03	1.71	1.97	1.70	1.55	1.74	1.46	1.28	1.12	1.50	1.27
4	1 minute	2.98	5.58	2.02	1.94	2.23	3.57	1.44	1.37	1.62	1.32	1.31	1.41
5	1 minute	1.84	1.61	1.63	1.47	0.75	0.79	1.84	1.61	1.63	1.47	0.75	0.79
6	1 minute	3.77	7.05	2.30	2.03	2.32	2.93	1.82	1.56	2.29	2.03	1.87	2.35
7	2 minutes	3.81	5.67	1.46	2.61	1.64	2.87	2.62	3.06	1.54	2.68	1.26	2.44
8	2 minutes	8.82	13.37	3.38	4.35	3.11	3.42	7.82	12.59	3.37	4.41	3.03	3.39
9	1 minute	2.01	2.21	2.66	2.21	3.15	3.50	1.71	1.69	2.45	1.97	2.12	1.76
10	2 minutes	3.53	6.70	2.05	2.78	2.23	3.18	1.87	1.80	2.07	2.84	1.65	2.05
11	2 minutes	4.81	8.52	2.04	2.42	2.72	4.46	1.68	1.79	1.49	1.81	1.04	0.95
12	2 minutes	5.83	8.45	2.16	2.68	2.37	2.73	2.93	3.17	2.03	2.66	2.06	2.55
Average % error		**4.22**		**2.35**		**2.28**		**2.88**		**2.21**		**1.76**	

6.2 Heart Rate Variability Estimation Analysis

The next step in our work focused on the HRV evaluation on the subset of participants (5 subjects) for which five minutes of continuous video recordings was collected in the SEMEOTICONS reference dataset. This subset contains facial videos coupled with the corresponding ground truth represented by the simultaneously recorded ECG. The analysis of the IBI series for the HRV parameters extraction was performed in a standard way according to the literature as described in section 5.2 and was initially evaluated by comparing the time-domain and frequency-domain features obtained from our method to the results obtained from the corresponding ECG.

Currently, there is a lack of agreed normative values for HRV and we therefore, determine the reference values (Table 14) from a recent quantitative systematic review of the literature [77]. According to this review, there is a wide variation in HRV measures even between participants of the same study. Moreover, even in homogenous healthy groups, measures of HRV can display wide inter-individual variations as high as 260,000%. Therefore, as stated by the authors, the data presented in the review should be used to quantify reference ranges for short-term measures of

HRV in healthy adult populations but should be undertaken with reference to methodological factors underlying disparate values.

Table 14: *The range in values for approved Task Force measures of short-term HRV*

HRV measures	*Range*
meanRR (ms)	785-1160
SDNN (ms)	32-93
RMSSD (ms)	19-75
LF (ms^2)	193-1009
LF (n.u.)	30-65
HF (ms^2)	82-3630
HF (n.u)	16-60
LF:HF	1.1-11.6

Most of our results (Table 15, Table 16) are within the reference ranges used to determine the disparate values for common measures of short-term HRV in healthy adult populations. The values falling outside the reference range are still close to the expected reference range. These results are probably affected by different factors as it is known that measures of HRV are influenced for example by age, gender, physical activity level and physical/mental stress. In our case, the protocol used to acquire HRV data contains factors such as stress and the impact on values presented, even though cannot be determined, should be taken into account.

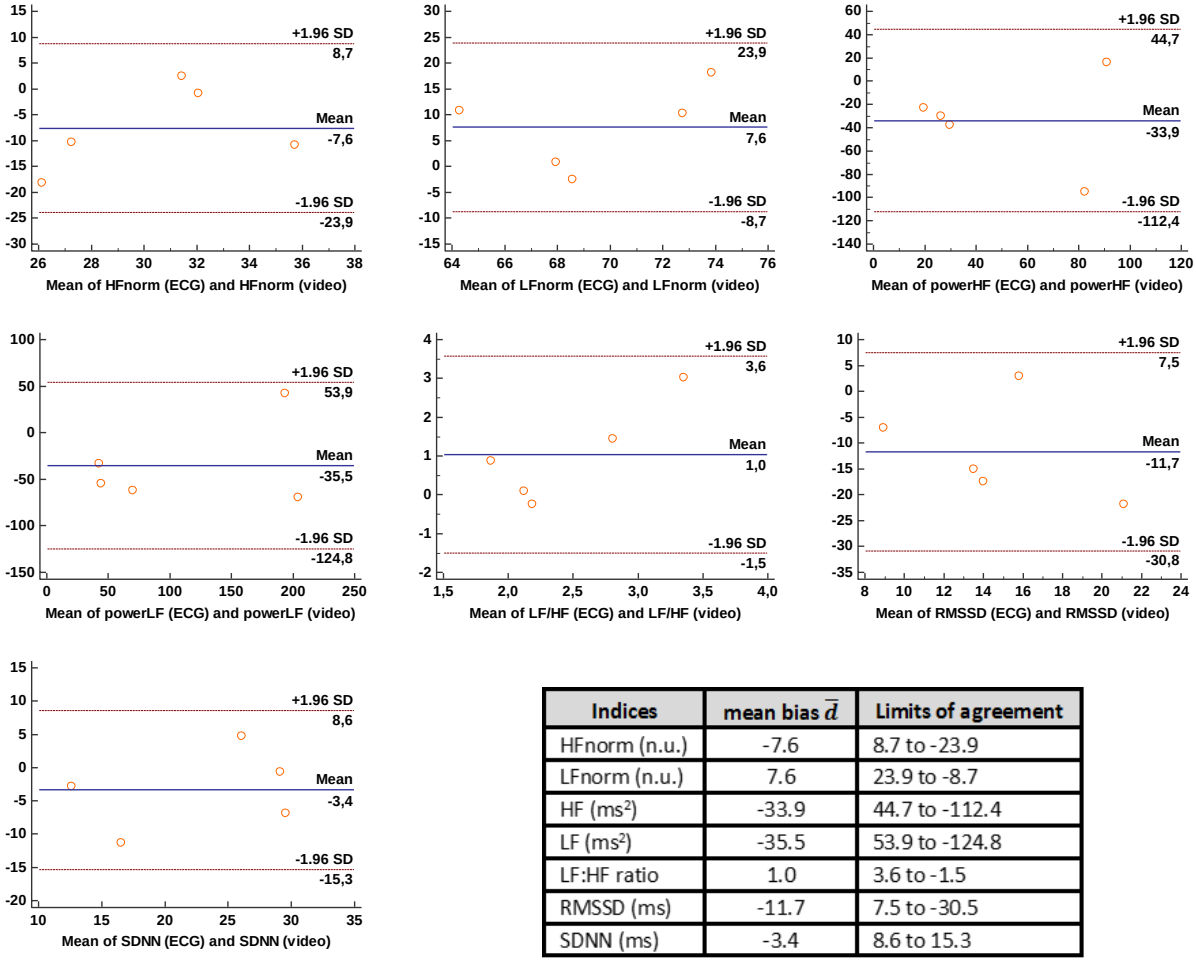
Table 15: *Mean intervals and HRV frequency domain short-term analysis results*

Signal	Mean Intervals (ms)		Frequency domain analysis									
			LF norm (n.u.)		HF norm (n.u)		LF power (ms^2)		HF power (ms^2)		LF/HF ratio	
	video	*ECG*	*video*	*ECG*	*video*	*ECG*	*video*	*ECG*	*video*	*ECG*	*video*	*ECG*
01	676	668	58.845	69.712	41.155	30.288	59.1	26.12	41.33	11.35	1.43	2.302
02	760	762	67.58	77.924	32.42	22.076	100.9	38.71	48.4	10.97	2.085	3.53
03	740	751	67.496	68.388	32.504	31.612	172.35	214.71	83	99.25	2.077	2.163
04	730	732	64.761	82.959	35.239	17.041	238.95	169.37	130.02	34.79	1.838	4.868
05	800	806	69.796	67.343	30.204	32.657	71.54	16.66	30.96	8.08	2.311	2.062

Table 16: *HRV time domain short-term analysis results*

Signal	Time domain analysis			
	SDNN (ms)		RMSSD (ms)	
	video	*ECG*	*video*	*ECG*
01	14	11.2	12.5	5.4
02	29.4	28.8	21	6
03	23.7	28.4	14.3	17.3
04	33	26.1	32	10.2
05	22.2	10.9	22.7	5.3

For a more extensive validation, the concordance correlation coefficient (pc) that evaluates the degree to which pairs of observations fall on the 45° line through the origin and the method of Bland and Altman were used to compare and evaluate the interchangeability between the ECG-based and video-based HRV indices. Figure 35 shows the results obtained from the Bland-Altman analysis. The discrepancies between ECG-based and video-based indices were reasonable although the correlation was weak for all indices. In more details, with 95% confidence interval, the mean bias $\bar{d}$ for the HFnorm (n.u.) was -7.6, for the LFnorm (n.u.) was 7.6, for the HF (ms2) was -33.9, for the LF (ms2) was -35.5, for the LF:HF ratio was 1.0, for the RMSSD (ms) was -11.7 and for the SDNN (ms) was -3.4. The wide intervals reflect the small sample size and the great variation of the differences.



Indices	mean bias $\bar{d}$	Limits of agreement
HFnorm (n.u.)	-7.6	8.7 to -23.9
LFnorm (n.u.)	7.6	23.9 to -8.7
HF (ms²)	-33.9	44.7 to -112.4
LF (ms²)	-35.5	53.9 to -124.8
LF:HF ratio	1.0	3.6 to -1.5
RMSSD (ms)	-11.7	7.5 to -30.5
SDNN (ms)	-3.4	8.6 to 15.3

Figure 35: *Bland–Altman analysis between ECG-based and video-based HRV indices*

The agreement between ECG-based and video-based measurement was also evaluated with the concordance correlation coefficient pc which is an index of how well a new test or

measurement reproduces a gold standard test or measurement. It can be calculated on as few as ten observations and it quantifies the agreement between two measures of the same variable. It contains a measurement of precision ρ and accuracy Cb:

$$p_c = pC_b$$

where:

- ρ is the Pearson correlation coefficient, which measures how far each observation deviates from the best-fit line, and is a measure of precision, and
- C_b is a bias correction factor that measures how far the best-fit line deviates from the 45° line through the origin, and is a measure of accuracy.

Like the correlation, pc ranges from -1 to 1, with perfect agreement at 1 and poor agreement for pc < 0.90. The results presented in Table 17 show that the video-based method presents a poor reproducibility for all indices. The highest accuracy (Cb=0.8836) and precision (p=0.87) was achieved by the LF (ms2) followed by the SDNN (ms) that achieved accuracy Cb=0.6625 and precision p=0.7513.

Table 17: *Concordance correlation coefficients of ECG-based and video-based HRV indices*

Indices	Concordance correlation coefficient p_c	Bias correction factor C_b (accuracy)	Pearson ρ (precision)	Confidence interval (95%)
LFnorm (n.u.)	-0,03304	0,4242	-0,07788	-0,4732 to 0,4204
HFnorm (n.u.)	-0,03304	0,4242	-0,07788	-0,4732 to 0,4204
LF (ms^2)	0,7687	0,8836	0,8700	0,05734 to 0,9623
HF (ms^2)	0,3331	0,6849	0,4864	-0,4203 to 0,8147
LF:HF (ratio)	-0,03770	0,2773	-0,1360	-0,3391 to 0,2708
SDNN (ms)	0,6625	0,8818	0,7513	-0,1843 to 0,9448
RMSSD (ms)	-0,03608	0,3107	-0,1161	-0,3714 to 0,3075

These results may imply poor quality of data, both video and ECG, or too many artefacts or even an error in pre-processing steps. In order to test the pre-processing phase for the HRV analysis, the HR extraction was evaluated on the five minutes of continuous video recordings. The results presented in Table 18 confirm the previously introduced results in case of HR measurement. Again, adequate results were produced by the number of peaks and the IBI approaches. In the case of signals 01 and 04, the artefact correction algorithm used before the IBI method led to more precise estimation of the beat-to-beat intervals and a noticeable improvement is shown in the reduction of extra-detected beats. On the other hand, in the case of signals 03 and 05, the peak detection probably was not affected or less affected by noise and artifacts meaning that the peaks

were maybe correctly detected. However, concerning the HRV analysis (Table 15, Table 16), the outcomes were unsatisfactory with considerable divergence from the values obtained from the ECG analysis. These outcomes revealed that, even though the HR results achieved from the IBI method were satisfactory, implying that the pre-processing phase for the HRV analysis was properly implemented, the constructed signal was not sufficiently representative of the BVP signal or otherwise the subsequent processing was not as accurate as required.

Table 18: *The HR results from 5-minutes video recordings*

Signal	Heart Rate				
	ECG	***#peaks - video***	***% error (#peaks)***	***IBI video***	***% error (IBI)***
01	90	87	3.33	89	1.11
02	79	79	0.00	79	0.00
03	79	79	0.00	81	2.53
04	82	81	1.22	82	0.00
05	74	74	0.00	75	1.35
Average % error			0.91		1.00

In the 5-minute recordings from SEMEOTICONS reference dataset used for the HRV analysis, the ECG and video acquisitions appears not to be ideal. Movement artefacts and technical failure were noted leading to a poor data quality that increased the need for more tests and modifications on the implemented algorithm, changes and adjustments in thresholds and parameters of the methods that may be affected from the errors imposed by even minor contamination of data with artefacts leading to the analysis of inappropriate signals. Visual inspection of the video recorded signals revealed that the signals were not of the appropriate quality for the required signal duration (i.e. five minutes) for accurate HRV analysis. We observed the presence of signal segments seriously contaminated by noise/artefacts. A more detailed inspection of the recorded videos revealed that there were motion artefacts caused from the participant's movement. As presented in Figure 36 there were cases where participants placed their hand in front of their face (Figure 36a) or made extreme body movements and facial expressions (Figure 36b) or even turned their head (Figure 36c). This can explain the unsatisfactory results since it is known that artefacts can have detrimental effects on computing HRV values, while their impact on HR is strongly mediated by the outlier removal and the averaging process, which are parts of the pipeline. Moreover, apart from the video signal (Figure 36e), the corresponding ECG signal (Figure 36d) was also affected which means that the ground truth in this case might is also be under question.

It is also known that the non-linear and nonstationary nature of HRV leads to abrupt changes in the signal so that HRV metrics are highly susceptible to outliers and artefacts [78]. The complex variation of the RR intervals are due to many interacting stimuli on many different time scales with different magnitudes. The RR interval time series constructed from the ECG recordings often contains different amounts of artefacts.

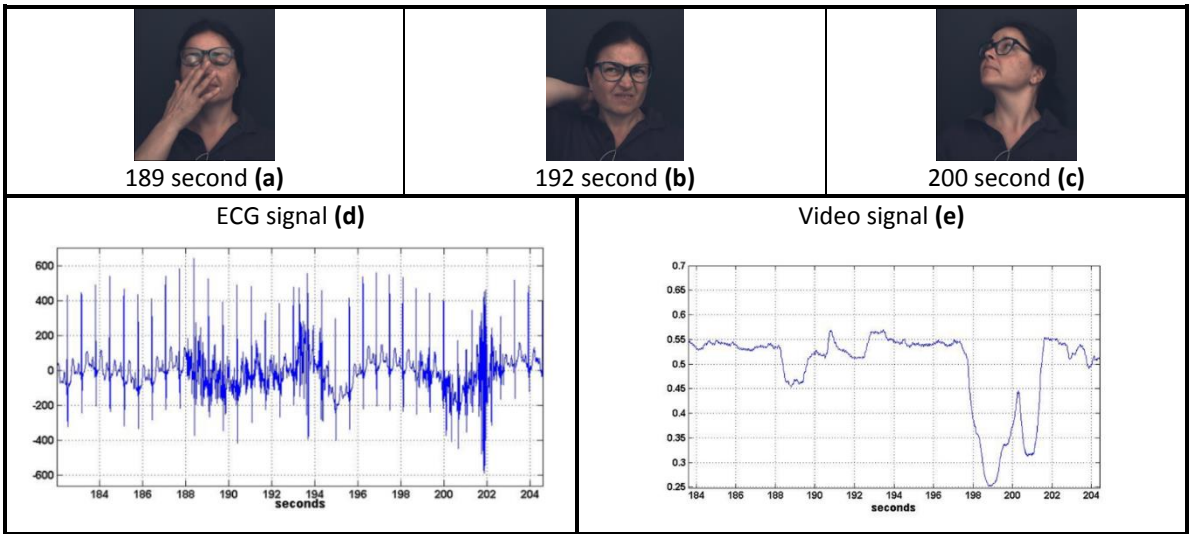


Figure 36: *Motion artefacts on ECG and video signals due to participant's movements (published with permission of the subject)*

These improperly positioned beats can be either of physiological or technical origin. For instance, technical artefact may result from poorly fastened electrodes or be due to motion of the subject. Ectopic beats and atrial fibrillation are examples of physiological artefact. Since improperly positioned or false beats are common in the RR interval time series, they complicate the reliable analysis of HRV and sometimes may make it impossible. These main problems that remain in HRV analysis from the ECG recordings are also affecting more heavily the HRV analysis from video signals. In general, it is known that in order to obtain a reliable HRV estimation, the recording conditions and pre-processing should fulfil numerous requirements. According to the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology [27] the analyzed signal should satisfy several requirements in order to obtain a reliable e.g. spectral estimation. A few examples of the main measurement standards, technical requirements and recommendations are:

- *A low sampling rate may produce a jitter in the estimation of the R-wave fiducial point, which alters the spectrum considerably. The optimal range is 250 to 500 Hz or perhaps even higher.*

- *Baseline and trend removal (if used) may affect the lower components in the spectrum. It is advisable to check the frequency response of the filter or the behavior of the regression algorithm and to verify that the spectral components of interest are not significantly affected.*
- *The choice of QRS fiducial point may be critical. It is necessary to use a well-tested algorithm (derivative plus threshold, template, correlation method) to locate a stable and noise-independent reference point. A fiducial point localized far within the QRS complex may also be influenced by varying ventricular conduction disturbances.*
- *Ectopic beats, arrhythmic events, missing data, and noise effects may alter the estimation of the PSD of HRV. Proper interpolation (or linear regression or similar algorithms) on preceding/successive beats on the HRV signal or on its autocorrelation function may decrease this error. Preferentially, short-term recordings that are free of ectopy, missing data, and noise should be used. In some circumstances, however, acceptance of only ectopic-free, short-term recordings may introduce significant selection bias. In such cases, proper interpolation should be used and the possibility of the results being influenced by ectopy should be considered.*
- *Frequency domain methods should be preferred to the time domain methods when short-term recordings are investigated. The recording should last for at least 10 times the wavelength of the lower frequency bound of the investigated component, and, in order to ensure the stability of the signal, should not be substantially extended. Thus, recording of approximately 1 minute is needed to assess the HF components of HRV, while approximately 2 minutes are needed to address the LF component. To standardize different studies investigating short-term HRV, 5-minute recordings of a stationary system are preferred unless the nature of the study dictates another design.*
- *The errors imposed by the imprecision of the NN interval sequence are known to affect substantially the results of statistical time domain and all frequency domain methods. It is known that casual editing of the RR interval data is sufficient for the approximate assessment of total HRV by the geometric methods, but it is not known how precise the editing should be to ensure correct results from other methods. Thus, when the statistical time domain and/or frequency domain methods are used, the manual editing of the RR data should be performed to a very high standard, ensuring correct identification and classification of every QRS complex. Automatic "filters" that exclude some intervals from the original RR sequence (for example, those differing by more than 20% from the previous interval) should not replace manual editing because they are known to behave unsatisfactorily and to have undesirable effects leading potentially to errors.*

In accordance to the recommendations described above, we decided to adopt different preprocessing methods based on the demands that HRV analysis has. The first aspect taken into account was the pre-processing performed on IBIs before we compute features. As often reported in the literature for HRV analysis [79]–[81], one of the most important steps is IBI correction, which prevents artefacts due to ectopic beats or motion from affecting features computation. Especially in case of video-based PPG signals, the interval correction is important since these

signals are very sensitive to motion artefacts that may cause unreliable HRV feature extraction. However, the published literature does not contain detailed reviews of relevant IBI correction methods and their impact on HR variability analyses. As a result, we have introduced different HRV signal pre-processing and IBI methods and tested for the artefact correction. These methods can have different effects on HRV measures. The effects of editing are dependent on the study setting, editing method, parameters used to assess HRV, type of study population, and the length of RR interval time series [82].

An example of consecutive heart intervals, detected with the video-based and the ECG-based method is demonstrated in Figure 37. The two segments of data, before correction, are very similar but result in different values due to ectopic beats that disrupt the computation and produce higher variation.

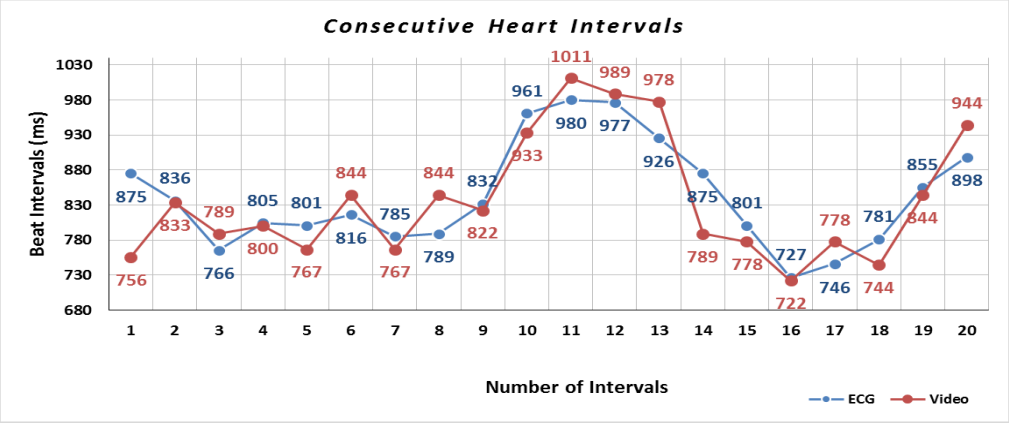


Figure 37: *Consecutive heart beat intervals detected with the video-based and the ECG-based method before artefact correction (a segment of one subject's data)*

In a large proportion of the available literature concerning HRV estimation, the ECG data is pre-processed by hand, with experts performing artefact rejection and noise removal [78], but this approach is not foreseen in our scenarios. Due to the inherent limitations of the above approach, especially for long term-ECG recordings, a common practice for IBI correction is to automatically adjust the inter-beat tachogram in order to remove the effect of abnormal beats on the HRV estimate. In achieving this objective, in our study the IBI tachogram was filtered using an impulse rejection filter for reducing the impact of artefact on the estimated PSD of IBI series. The algorithmic process implemented comprises two steps: a) for each sample a decision as to whether the signal is corrupted is taken, based on the formula of Equation 5 [76] which includes a user-specified threshold τ and b) if corruption is identified, the artefact is replaced by the median, else the signal is left unaltered.

$$D(n) = \frac{|x(n) - med\{x_n\}|}{1.483 med\{|x(n) - med\{x_n\}|\}} \quad \textbf{(Equation 5)}$$

if $D(n) < \tau$ ***then not ectopic; else ectopic***

In our case, following extensive experimentation, it was estimated that for HR measurement from the IBI series a threshold value of 0.5 gave the best results. The selection of an optimal threshold is always a challenge due to the variability in signal quality. A very low threshold makes the filter more sensitive, thus in case of good quality data it leads to overcorrection, as shown in Figure 38.

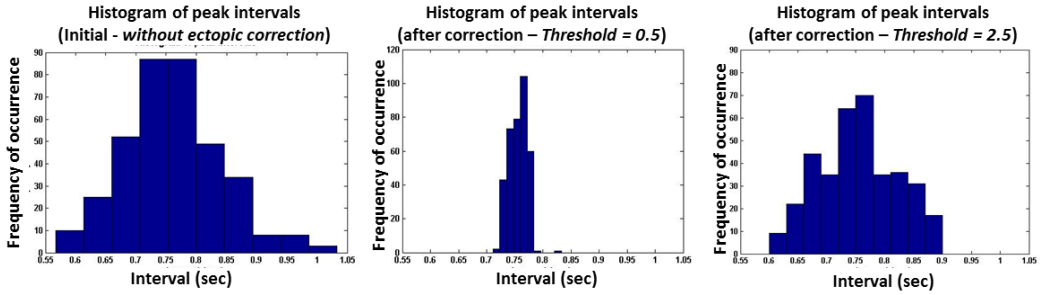


Figure 38: *The value of the relevant peak intervals with and without ectopic detection and correction, demonstrating the influence of the threshold value*

Given all the observed sources of artefacts we have decided to test our generic processing methodology in selected video segments. These were chosen, following visual inspection, to be segments with no apparent artefacts present. The results obtained, shown in Table 19 and Table 20, demonstrate that when a relatively good quality data-segment is selected and a threshold equal to 2.5 is used for the impulse rejection filter, the extracted estimates, both video and ECG, were in better agreement to those obtained when no artefact or ectopic beat correction was performed.

Table 19: *Overcorrection on mean intervals and HRV frequency domain results (good quality signal)*

Duration	Threshold	Mean Intervals (ms)		Frequency domain analysis									
				LF norm (n.u.)		HF norm (n.u)		LF power (ms²)		HF power (ms²)		LF/HF ratio	
		video	*ECG*	*video*	*ECG*	*video*	*ECG*	*video*	*ECG*	*video*	*ECG*	*video*	*ECG*
211 sec	-	767	763	70.728	61.174	89.272	38.826	166.65	103.56	1386.69	65.73	0.12	1.576
	0.5	755	766	57.07	69.076	42.93	30.924	82.25	33.81	61.88	15.13	1.329	2.234
	2.5	759	763	15.973	61.341	84.027	38.659	224.89	100.28	1183.03	63.2	0.19	1.587
280 sec	-	759	759	67.824	69.826	32.176	30.174	1968.9	1911.4	934.07	825.97	2.108	2.314
	0.5	754	752	63.353	64.027	36.647	35.973	172.96	131.4	100.05	73.83	1.729	1.78
	2.5	757	758	67.526	66.809	32.474	33.191	1741.9	1683.2	837.69	836.24	2.079	2.013

Contrarily, the very low threshold value (i.e. 0.5) produced significantly worse outcomes as a result of overcorrection. The results presented in Table 19 and Table 20 show also that the estimations extracted from ECG without ectopic correction are closer to those extracted from video

when using the 0.5 threshold value. In general, in time domain analysis of short-term data, both SDNN and RMSSD extracted from video and ECG were least affected by editing while in the frequency domain analysis, editing resulted in different error for both LF and HF spectral components extracted from the ECG. Thus, different signals may require different threshold values for an accurate estimation.

Table 20: *Overcorrection on HRV time domain results from good quality signal*

Duration	Threshold	Time domain analysis			
		SDNN (ms)		RMSSD (ms)	
		video	*ECG*	*video*	*ECG*
211 sec	-	76.8	37	108.5	11.8
	0.5	22.2	28.2	25.2	6.3
	2.5	63.4	36.7	88.5	11.7
280 sec	-	73	67.9	61.2	43.4
	0.5	25.2	22.3	17.2	13.9
	2.5	69	65.6	57.6	42.9

As previously reported, the PPG based systems are very sensitive to motion and we have to consider issues caused by these artefacts which last for a certain amount of time instead of only considering issues caused from single ectopic beats. In any case, it should be noted that although there are still open issues and optimizations that can take place, the results obtained have significantly improved, and were more precise.

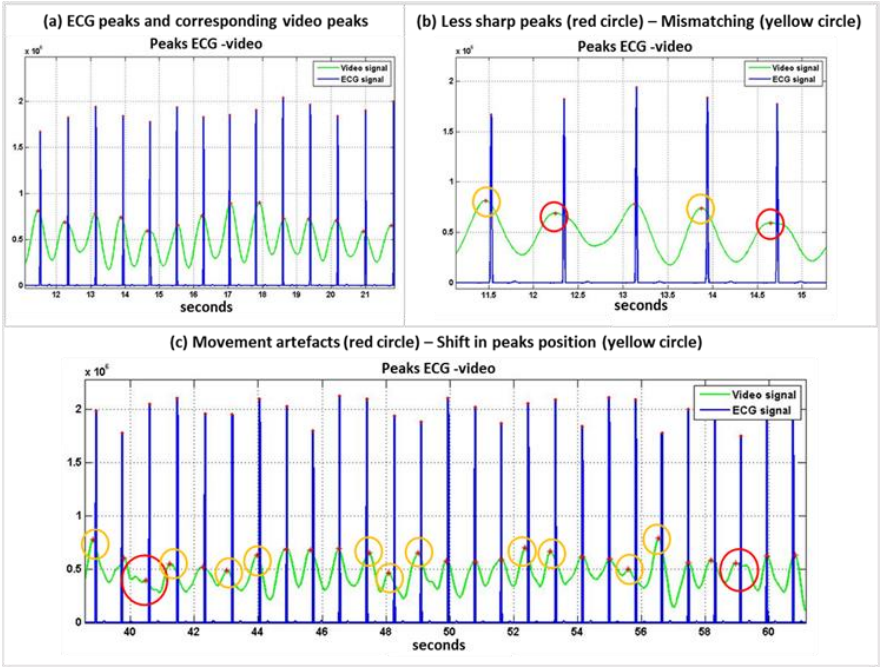


Figure 39: *Video and ECG signal peaks*

Focusing on the open issues, we have observed that the HR results were in good agreement due to the fact that the algorithm correctly detects the number of peaks even in cases were the signal is not so good and there are timewise peak-mismatching (Figure 39c). Although HR estimation is correct (Figure 39a), this does not hold for the HRV estimation, suggesting problems in ectopic beat detection and even more in the proper positioning of the fiducial point on each beat. The main problem relating to the HRV estimate, is the fact that, in the reconstructed BVP signal, there is a relative timewise shifting in the position of the peaks, due to movement artefacts (Figure 39b and Figure 39c). These artefacts deteriorate the accuracy of the HRV analysis and produces, somewhat, unreliable results for the fiducial points (peaks in our case). In addition, as seen in Figure 39b, the peaks in the video signal are less sharp, which implies that the precision at which the IBIs can be measured is also affected.

6.3 Respiratory Rate Estimation Analysis

Our final sign of interest is the respiration rate. Unfortunately the SEMEOTICONS dataset generated does not contain a reference signal, through the use of a respiratory belt for example and the low quality of the single-lead ECG did not allow a reliable evaluation of the respiratory signal from the amplitude modulation of the QRS peaks. As explained in Chapter 5 the respiratory rate was estimated from the center frequency of the highest peak between 0.15-0.4 Hz of the HRV power spectrum. The estimates are shown in Table 21.

Table 21: *The respiration rate results from 5-minutes video recordings*

Signal	Respiration Rate			
	ECG	***Video***	**% error**	**Visual inspection**
01	18	17	5.56	13-18
02	15	12	20.00	14-16
03	11	9	18.18	18-21
04	20	13	35.00	16-21
05	20	23	15.00	18-23
Average % error			18.75	

The results obtained are satisfactory overall, although there are deviations between the estimates calculated from the ECG and the HRV estimates, in some cases significant. Since, the measurements derived from the HF-Frequency band of HRV from both signals (ECG and video) were used, our estimates depend on the accuracy of the HRV estimation.

7 Discussion

The results presented in this Thesis regarding the contactless measurement of HR using facial videos, demonstrate that accurate estimation is possible. The reported outcomes, applying the computational and analytical pipeline fully described in this manuscript, indicate accuracies of the order of 97-99%, when compared to the ground truth measurement, i.e. estimation of HR from the ECG signal.

The signal pre-processing study indicated that, in order to improve performance and results, the most appropriate method for forming the raw signal traces was the calculation of the pixel values, excluding outliers. Concerning the most suitable method to remove oscillations in the signal due to external factors such as movement was the detrending based on a *SP approach* and the most appropriate source separation algorithm for motion artefact removal was the ICA based on JADE.

Our results regarding other factors influencing the estimations indicate that, although different areas of the face can be used as the ROI for the estimation of HR, the optimal ROI appears to be the whole face. Specifically, we have observed optimal results when using the full width of the face reduced by 40%. This result reinforces the findings of prior reported studies [52]. The reasons for this, we believe, are twofold. Firstly, the fact that the PPG signal strength varies over different regions in the face and depends on intensity of incident light I(x,y) in the region and on the blood perfusion underneath the skin [52]. As a result, if a person moves in front of camera, e.g. tilting, smiling, shifting, talking etc., a specific point may not be reliable, but one needs to track, at the very least, the whole face to extract the PPG signal. Secondly, even slight rotation of person's face relative to camera leads to change in the direction of incident and reflected light. Such small changes in light direction usually lead to large changes in skin surface reflectance. This was observed to be the case in the SEMEOTICONS reference dataset, in which participants performed some real-life activities in which the presence of rigid and non- rigid motion was evident. The above reasons explain the deterioration of the accuracy observed between trial 3 (RMSE 2.03), trial 9 (RMSE 2.21) and trial 1 (RMSE 2.85) in which the subjects demonstrated the least motion and trial 2 (RMSE 11.62) and trial 8 (RMSE 13.37), since in these trials a higher level of different type of motion was observed. These results derive from the whole dataset and refer to the max

selection method. Similar observations were found in the case of the number of BVP peaks and $60/\overline{IBI}$ methods.

In addition, our work reinforces other people's findings in relation to the issue of selection of the most suitable component after applying ICA-based BSS techniques. In the context of the present work we performed a detailed investigation of component selection algorithms by examining the influence of the component selection on later processing steps and results, i.e. in assessing the HR error after applying a certain estimation method [54]. In specific, our results indicate that, in some cases, there can be more than one component that can be used for accurate HR estimation, implying also that an adequate source separation is quite hard to achieve. Each component could estimate the HR better than others on different occasions and there were also cases that the ICA based on JADE was not able to separate properly the sources especially in cases where different noise factors co-exist meaning that the number of signal sources is unknown and may be greater than the number of recordings. In such cases different methods could be also applied. Our conclusions in relation to the most suitable component, after applying ICA-based BSS techniques, is that for the automated selection of the component exhibiting the maximum signal-to-noise ratio, even when the best ICA output component is selected by the manual selection method, an error from 1.62% to 4.66% still exists.

In relation to the accuracy of HR measurement, Table 22 summarizes our results as compared to some of the most relevant and recent published studies. Poh et al. [40] reports the minimum RMSE of 1.24 bpm using an alternative to the FFT (BVP signal processing in the time domain) approach. Using this approach the RR series is extracted after the peak detection in the BVP signal, the RRmean is evaluated and the HR is measured as HR = 60/RRmean. This approach can outperform the FFT approach with accurate peak detection, achieving a better frequency resolution than the one obtained using FFT which is df = fs / N (where N is the number of samples and fs is the sampling frequency). This approach is also necessary in order to perform HRV analysis, which was the subject of our later work. For such a purpose, a proper selection of the channel with the best replica of the BVP signal is paramount. In any case, the size of our dataset and the fact that the ground truth reference is provided by an ECG measuring device provides, in our view, conclusive evidence that highly accurate measurement of HR in real life conditions is possible.

Regarding the HRV estimate, the work presented in this thesis indicates that it is indeed possible. On the other hand we observed that the noise/artefacts in some cases proved to be dominant over the actual signal, thereby causing contamination by altering the original signal and masking the heart signal morphology. Prior to the HRV analysis, these interferences need to be minimized in the data acquisition stage, which is much more critical than in case of HR computation.

Table 22: *Summary of the estimated HR accuracy from the most relevant studies*

Study	*Number of subjects*	*Number of recordings*	*Reference*	*RMSE*
Poh et al. (2010)	12	24	finger BVP sensor	2.29
Poh et al. (2011)	12	12	finger BVP sensor	1.24
Monkaresi et al. (2014)	10	10	ECG	3.64
Our method All trials (*HR from IBI*)	24	247	ECG	2.87
Our method All trials (*HR from IBI*)	22	229	ECG	1.98

Furthermore, any noise within the obtained signal must be removed in order to accurately locate the BVP peaks with a robust detector algorithm. In the process of data acquisition, a good signal quality and elimination of noise is of essential importance for further HRV analysis. The data acquisition itself can be optimized by correct positioning of the subject in front of the camera (limited movements and expressions) and by applying several conditioning steps to the acquired raw signal without forgetting that, in non-contact PPG method, changes in ambient light and automatic brightness adjustment in camera or software have also a significant influence on the baseline of the signal [39].

After the process of data acquisition is completed, the BVP peaks are detected and the HRV is analyzed by applying appropriate mathematical algorithms over the obtained data. Moreover, in order to be able to perform effective measurement results and consequently recognize possible pointers of a certain condition, a study of measured signals may have to be carried out on measured data obtained over a longer period of time. In our study, a 5 minutes recording was recognized as appropriate standard option for further data analysis as this duration is the recommended for reliable short-term HRV analysis. However, our analysis allow us to provide evidence only for the feasibility and not for the reliability of our method. The HRV measures were found to be unreliable that may be explained in part, by the data collection protocol. These results must be considered in the light of several limitations such as: a) the small sample size, b) the fact that in our dataset the

individual subjects were recorded in the same environment but under different conditions and c) the fact that the recording protocol of the SEMEOTICONS dataset was not designed specifically for HRV analysis and this dataset was selected since it was the only available dataset that provided ECG and video recordings acquired in parallel for 5 minutes. Therefore, in order to thoroughly study the reliability and the set of issues initially established, a dataset compatible with the basic requirement set is necessary.

Finally, concerning the respiration rate, since the SEMEOTICONS dataset does not contains a reference signal (e.g. respiratory belt), the measurements derived from the HF-Frequency band of HRV from both signals (ECG and video) were used. Thus, the respiration rate values depend on the accuracy of the HRV estimation.

8 Conclusions

In conclusion, the work reported in this thesis proves that highly accurate contactless measurement of HR in real life conditions is possible, and hence enriches existing knowledge regarding the efficacy of contactless measurement of HR. This increased confidence provides the required evidence for the, potentially, commercial exploitation of such systems and their adoption in non-clinical settings and especially for wellness monitoring and management in pervasive eHealth environments.

On the other hand, the same confidence does not exist regarding the HRV analysis, since satisfactory results have been achieved only on videos where the subject movements were almost absent, thus imposing strict requirements for the subject during the whole acquisition phase. Although, positive indications on the feasibility of the approach were obtained, the reported inefficiencies in relation to a) the experimental setup, b) artefacts present in the recorded signals and c) quality of the available dataset (duration, motion artefacts) do not allow us for an experimentally validated estimation of the accuracy of HRV.

In all cases, there are still open issues that need to be successfully addressed in the future. It is known that the accuracy of the estimation is significantly influenced when dark colored subjects are involved [52]. This remains an issue that requires attention. Also, our results reaffirm the importance of the artefacts and errors introduced as a result of very high head motion (and head motion in general for the HRV). Several other published studies such as [12], [49], [52] have identified this issue. Our results, on the MAHNOB-HCI dataset, reaffirm the influence of a subject's motion on the accuracy of HR measurement. However, although this is true, moderate motion and non-extreme changes in facial expressions of the subject are present in our dataset and are well tolerated by our analysis pipeline for the HR estimation. What is not yet known is the quantitative relationship between head motion and the error in HR estimation. Applying ICA-based methods separately in different tracked regions of the face, as opposed to the whole face, can possibly lead to further improvement of PPG estimation under motion scenarios. Both these issues demand additional experimentation, such as the use of more than one synchronized camera or a five band camera [51], in the attempt to increase the number of signals in input to the ICA-based BSS algorithm for a more appropriate estimation of the 3 independent source signals (head movement, variation in the ambient illumination and BVP) of our interest, and this will be the

focus of our future work. Moreover, recently studies have been performed on optimal ways for reducing the movement artefacts [83] that are reported to have better performance than the ICA-based BSS methods. This provides an additional direction for future research and potential further enhancements of our algorithms.

Specifically with respect to HRV estimation, and in an attempt to further increase the accuracy and robustness of our estimation, several alternatives need to be explored. Namely, in most instances both the non-parametric and parametric methods for the calculation of PSD provide comparable results as reported in the Task Force of The European Society of Cardiology and the North American Society for Pacing and Electrophysiology guideline [27]. However, it is unclear which spectral methods are most applicable to HRV frequency-domain analysis. The report recommends using both parametric and nonparametric assessments when evaluating frequency domain HRV measures. While the nonparametric methods, such as the ones used in this work, have the advantage of algorithmic simplicity and rapidity, the parametric methods produce smoother spectral components that can be distinguished more easily, and if the order of the model is correctly chosen, they can allow an accurate estimation of the PSD over very short windows. It is also known that changes in ambient light and automatic brightness adjustment in camera or software have also a significant influence on the baseline of the signal [39]. These artefacts were observed in our dataset. Hence a more carefully planned experimental setup and data acquisition will be required, in order to conclusively study the effect of these parameters on our estimates of both HR and HRV.

References

[1] K. Steinitz and E. Belt, “Leonardo da Vinci’s Trattato della pittura (Treatise on painting): A biliography of the printed editions, 1651-1956, based on the complete collection in the Elmber Belt Library of Vinciana, preceded by a study of its sources and illustrations,” vol. 5, 1958.

[2] R. Descartes, *Les passions de l’âme (1649). oeuvres et lettres*. Paris: Gallimard,«Pléiade 739, 1953.

[3] C. Darwin, *The Expression of the Emotions in Man and Animals*. University of Chicago Press, 1965.

[4] F. Chiarugi, G. Iatraki, E. Christinaki, D. Manousos, G. Giannakakis, M. Pediaditis, A. Pampouchidou, K. Marias, and M. Tsiknakis, “Facial signs and psycho-physical status estimation for well-being assessment,” in *7th International Conference on Health Informatics (HEALTHINF)*, 2014, pp. 555–562.

[5] S. Ulyanov and V. Tuchin, “Pulse-wave monitoring by means of focused laser beams scattered by skin surface and membranes,” *OE/LASE’93 Opt. Electro-Optics, Laser Appl. Sci. Eng. Int. Soc. Opt. Photonics*, pp. 160–167, 1993.

[6] E. F. Greneker, “Radar sensing of heartbeat and respiration at a distance with applications of the technology.” IET Digital Library, pp. 150–154, 01-Jan-1997.

[7] M. Garbey and N. Sun, “Contact-free measurement of cardiac pulse based on the analysis of thermal imagery,” *Biomed. Eng. IEEE Trans.*, vol. 58, no. 4, pp. 1418–1426, 2007.

[8] P. Pelegris, K. Banitsas, T. Orbach, and K. Marias, “A novel method to detect heart beat rate using a mobile phone,” in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 5488–5491.

[9] U. Morbiducci and L. Scalise, “Optical vibrocardiography: A novel tool for the optical monitoring of cardiac activity,” *Ann. Biomed. Eng.*, vol. 35, no. 1, pp. 45–58, 2007.

[10] H.-Y. Wu, M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman, “Eulerian video magnification for revealing subtle changes in the world,” *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–8, Jul. 2012.

[11] C. Takano and Y. Ohta, “Heart rate measurement based on a time-lapse image,” *Med. Eng. Phys.*, vol. 29, no. 8, pp. 853–857, 2007.

[12] M.-Z. Poh, D. J. McDuff, and R. W. Picard, “Non-contact, automated cardiac pulse measurements using video imaging and blind source separation,” *Opt. Express*, vol. 18, no. 10, pp. 10762–10774, 2010.

[13] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu, “Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise,” *J. Biomed. Opt.*, vol. 16, no. 7, p. 077010, Jul. 2011.

[14] E. Christinaki, G. Giannakakis, F. Chiarugi, M. Pediaditis, G. Iatraki, M. Dimitris, K. Marias, and M. Tsiknakis, “Comparison of Blind Source Separation Algorithms for Optical

Heart Rate Monitoring," in *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*, 2014, pp. 339–342.

[15] A. Alwan, *Global status report on noncommunicable diseases 2010*. World Health Organization, 2011.

[16] T. F. of the E. S. of Cardiology, "Heart rate variability standards of measurement, physiological interpretation, and clinical use," *Eur Hear. J*, vol. 17, pp. 354–381, 1996.

[17] J. Sztajzel, "Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system," *Swiss Med. Wkly.*, vol. 134, pp. 514–522, 2004.

[18] R. E. Kleiger, J. P. Miller, J. T. Bigger, and A. J. Moss, "Decreased heart rate variability and its association with increased mortality after acute myocardial infarction," *Am. J. Cardiol.*, vol. 59, no. 4, pp. 256–262, Feb. 1987.

[19] M. T. La Rovere, "Short-Term Heart Rate Variability Strongly Predicts Sudden Cardiac Death in Chronic Heart Failure Patients," *Circulation*, vol. 107, no. 4, pp. 565–570, Jan. 2003.

[20] M. Pagani, G. Malfatto, S. Pierini, R. Casati, A. M. Masu, M. Poli, S. Guzzetti, F. Lombardi, S. Cerutti, and A. Malliani, "Spectral analysis of heart rate variability in the assessment of autonomic diabetic neuropathy," *J. Auton. Nerv. Syst.*, vol. 23, no. 2, pp. 143–153, Aug. 1988.

[21] F. Roche, J.-M. Gaspoz, I. Court-Fortune, P. Minini, V. Pichot, D. Duverney, F. Costes, J.-R. Lacour, and J.-C. Barthelemy, "Screening of Obstructive Sleep Apnea Syndrome by Heart Rate Variability Analysis," *Circulation*, vol. 100, no. 13, pp. 1411–1415, Sep. 1999.

[22] B. Appelhans and L. Luecken, "Heart rate variability as an index of regulated emotional responding.," *Rev. Gen. Psychol.*, vol. 10, no. 3, pp. 229–240, 2006.

[23] A. H. Kemp and D. S. Quintana, "The relationship between mental and physical health: insights from the study of heart rate variability.," *Int. J. Psychophysiol.*, vol. 89, no. 3, pp. 288–96, Sep. 2013.

[24] L. Vanderlei and C. Pastre, "Basic notions of heart rate variability and its clinical applicability," *Rev. Bras. Cir. Cardiovasc.*, vol. 24, no. 2, pp. 205–217, 2009.

[25] R. E. Kleiger, P. K. Stein, and J. T. Bigger, "Heart rate variability: measurement and clinical utility.," *Ann. Noninvasive Electrocardiol.*, vol. 10, no. 1, pp. 88–101, Jan. 2005.

[26] A. H. Kemp, D. S. Quintana, M. A. Gray, K. L. Felmingham, K. Brown, and J. M. Gatt, "Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis.," *Biol. Psychiatry*, vol. 67, no. 11, pp. 1067–74, Jun. 2010.

[27] 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, Mar. 1996.

[28] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophys. Space Sci.*, vol. 39, no. 2, pp. 447–462, 1976.

[29] R. J. Martin and C. A. Openshaw, "Autoregressive modelling in vector spaces: An

application to narrow-bandwidth spectral estimation," *Signal Processing*, vol. 50, no. 3, pp. 189–194, May 1996.

[30] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, 2001.

[31] E. Labbé, N. Schmidt, J. Babin, and M. Pharr, "Coping with stress: the effectiveness of different types of music," *Appl. Psychophysiol. Biofeedback*, vol. 32, no. 3–4, pp. 163–168, 2007.

[32] N. Gupta, "ECG and wearable computing for drowsiness detection," California State University, Northridge, 2014.

[33] C. Kossmann, "The normal electrocardiogram," *Circulation*, vol. 8, no. 6, pp. 920–936, 1953.

[34] J. Allen, "Photoplethysmography and its application in clinical physiological measurement.," *Physiol. Meas.*, vol. 28, no. 3, pp. R1–39, Mar. 2007.

[35] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals.," *Curr. Cardiol. Rev.*, vol. 8, no. 1, pp. 14–25, Feb. 2012.

[36] E. Gil, M. Orini, and R. Bailón, "Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions," *Physiol. Meas.*, vol. 31, no. 9, p. 1271, 2010.

[37] J. Zheng and S. Hu, "The preliminary investigation of imaging photoplethysmographic system," *J. Phys. Conf. Ser.*, vol. 85, p. 012031, 2007.

[38] S. Hu, J. Zheng, V. Chouliaras, and R. Summers, "Feasibility of imaging photoplethysmography," *Biomed. Eng. Informatics New Dev. Futur. - Proc. 1st Int. Conf. Biomed. Eng. Informatics, BMEI 2008*, vol. 2, pp. 72–75, 2008.

[39] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[40] M. Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, 2011.

[41] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam - A non-contact method for evaluating cardiac activity," in *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2011, pp. 405–410.

[42] S. Kwon, H. Kim, and K. S. Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2012, pp. 2174–2177.

[43] Y. Sun, C. Papin, V. Azorin-Peris, R. Kalawsky, S. Greenwald, and S. Hu, "Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam," *J. Biomed. Opt.*, vol. 17, no. 3, p. 037005, 2012.

[44] L. Wei, Y. Tian, Y. Wang, T. Ebrahimi, and T. Huang, "Automatic webcam-based human heart rate measurements using laplacian eigenmap," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7725 LNCS, no. PART 2, pp. 281–292, 2013.

[45] G. R. Tsouri, S. Kyal, S. Dianat, and L. K. Mestha, "Constrained independent component analysis approach to nonobtrusive pulse rate measurements," *J. Biomed. Opt.*, vol. 17, no. 7, pp. 0770111–0770114, 2012.

[46] D. Datcu, M. Cidota, S. Lukosch, and L. Rothkrantz, "Noncontact Automatic Heart Rate Analysis in Visible Spectrum by Specific Face Regions," in *Proceedings of the 14th International Conference on Computer Systems and Technologies*, ACM, 2013, pp. 120–127.

[47] G. Lempe, S. Zaunseder, S. Tom Wirthgen, T. Zipser, and H. Malberg, "ROI selection for remote photoplethysmography," in *Bildverarbeitung für die Medizin 2013*, Springer Berlin Heidelberg, 2013, pp. 99–103.

[48] H. Monkaresi, R. A. Calvo, and H. Yan, "A Machine Learning Approach to Improve Contactless Heart Rate Monitoring Using a Webcam," *Biomed. Heal. Informatics, IEEE J.*, vol. 18, no. 4, pp. 1153–1160, 2013.

[49] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote Heart Rate Measurement From Face Videos Under Realistic Situations," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE*, 2014, pp. 4264–4271.

[50] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.

[51] D. McDuff, S. Gontarek, and R. W. Picard, "Improvements in Remote Cardio-Pulmonary Measurement Using a Five Band Digital Camera," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2593–2601, 2014.

[52] M. Kumar, A. Veeraraghavan, and A. Sabharval, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomed. Opt. Express*, vol. 6, no. 5, pp. 1565–1588, 2015.

[53] K. Mannapperuma, B. D. Holton, P. J. Lesniewski, and J. C. Thomas, "Performance limits of ICA-based heart rate identification techniques in imaging photoplethysmography," *Physiol. Meas.*, vol. 36, no. 1, pp. 67–83, 2015.

[54] D. Wedekind and F. Gaetjen, "Automated Identification of Cardiac Signals after Blind Source Separation for Camera-Based Photoplethysmography," in *Electronics and Nanotechnology (ELNANO), 2015 IEEE 35th International Conference on*, 2015, pp. 422–427.

[55] D. J. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford, "A survey of remote optical photoplethysmographic imaging methods.," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 6398–6404.

[56] H. Liu, Y. Wang, and L. Wang, "A review of non-contact, low-cost physiological information measurement based on photoplethysmographic imaging.," in *Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2012, pp. 2088–2091.

[57] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomed. Signal Process. Control*, vol. 13, pp. 102–112, Sep. 2014.

[58] J. R. Stroop, "Studies of interference in serial verbal reactions," *J. Exp. Psychol.*, vol. 18, no. 6, pp. 643–662, 1935.

[59] M. P. Tarvainen, P. O. Ranta-Aho, and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis.," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 2, pp. 172–5, Feb. 2002.

[60] F. M. Pouzols and A. Lendasse, "Effect of different detrending approaches on computational intelligence models of time series," in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 2010, pp. 1–8.

[61] K. T. Sweeney, T. E. Ward, and S. F. McLoone, "Artifact removal in physiological signals-practices and possibilities.," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 488–500, May 2012.

[62] G. Naik and D. Kumar, "An overview of independent component analysis and its applications," *Informatica*, 2011.

[63] J. Iriarte, E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri, and J. Artieda, "Independent Component Analysis as a Tool to Eliminate Artifacts in EEG: A Quantitative Study," *J. Clin. Neurophysiol.*, vol. 20, no. 4, pp. 249–257, Jul. 2003.

[64] M. Antonakakis, G. Giannakakis, M. Tsiknakis, S. Micheloyannis, and M. Zervakis, "Synchronization coupling investigation using ICA cluster analysis in resting MEG signals in reading difficulties," in *13th IEEE International Conference on BioInformatics and BioEngineering*, 2013, pp. 1–5.

[65] T. D. Lagerlund, F. W. Sharbrough, and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition.," *J. Clin. Neurophysiol.*, vol. 14, no. 1, pp. 73–82, Jan. 1997.

[66] T. He, G. Clifford, and L. Tarassenko, "Application of independent component analysis in removing artefacts from the electrocardiogram," *Neural Comput. Appl.*, vol. 15, no. 2, pp. 105–116, Nov. 2005.

[67] H. Zhang and L.-Q. Zhang, "ECG analysis based on PCA and Support Vector Machines," in *2005 International Conference on Neural Networks and Brain*, vol. 2, pp. 743–747.

[68] N. Correa, T. Adali, and V. D. Calhoun, "Performance of blind source separation algorithms for fMRI analysis using a group ICA method.," *Magn. Reson. Imaging*, vol. 25, no. 5, pp. 684–94, Jun. 2007.

[69] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[70] A. Hyvärinen and E. Oja, “A fast fixed-point algorithm for Independent Component Analysis,” *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, Oct. 1997.

[71] J. F. Cardoso and A. Souloumiac, “Blind beamforming for non-gaussian signals,” *IEE Proc. F Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, 1993.

[72] P. Georgiev and A. Cichocki, “Blind source separation via symmetric eigenvalue decomposition,” in *Proceedings of the Sixth International Symposium on Signal Processing and its Applications (Cat.No.01EX467)*, 2001, vol. 1, pp. 17–20.

[73] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, vol. 1, pp. I–511–I–518.

[74] C. Tomasi and T. Kanade, “Detection and Tracking of Point Features,” *Int. J. Comput. Vis.*, no. 7597, 1991.

[75] R. Schafer, “What Is a Savitzky-Golay Filter? [Lecture Notes],” *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011.

[76] J. McNames, T. Thong, and M. Aboy, “Impulse rejection filter for artifact removal in spectral analysis of biomedical signals.,” in *Engineering in Medicine and Biology Society, 2004. IEMBS’04. 26th Annual International Conference of the IEEE*, 2004, vol. 1, pp. 145–148.

[77] D. Nunan, G. R. H. Sandercock, and D. A. Brodie, “A Quantitative Systematic Review of Normal Values for Short-Term Heart Rate Variability in Healthy Adults,” *Pacing Clin. Electrophysiol.*, vol. 33, no. 11, pp. 1407–1417, 2010.

[78] G. Clifford, “Signal processing methods for heart rate variability,” University of Oxford, 2002.

[79] M. N. Cheung, “Detection of and Recovery from Errors in Cardiac Interbeat Intervals,” *Psychophysiology*, vol. 18, no. 3, pp. 341–346, May 1981.

[80] G. G. Berntson and J. R. Stowell, “ECG artifacts and heart period variability: Don’t miss a beat!,” *Psychophysiology*, vol. 35, no. 1, pp. 127–132, Jan. 1998.

[81] G. G. Berntson, K. S. Quigley, J. F. Jang, and S. T. Boysen, “An Approach to Artifact Identification: Application to Heart Period Data,” *Psychophysiology*, vol. 27, no. 5, pp. 586–598, Sep. 1990.

[82] M. A. Peltola, “Role of editing of R-R intervals in the analysis of heart rate variability.,” *Front. Physiol.*, vol. 3, p. 148, Jan. 2012.

[83] G. de Haan and A. van Leest, “Improved motion robustness of remote-PPG by using the blood volume pulse signature.,” *Physiol. Meas.*, vol. 35, no. 9, pp. 1913–1926, Aug. 2014.