TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

ALEKSI HAAVIKKO

EVALUATION OF PERFORMANCE OF AN OPTICAL HEART RATE SENSOR

Master of Science Thesis

Examiners: Professor Jukka Lekkala
Professor Ilkka Korhonen
Examiners and topic approved by the Council of the Faculty of Engineering Sciences on 3 September 2014

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master’s Degree Programme in Automation Technology
HAAVIKKO, ALEKSI: Evaluation of performance of an optical heart rate sensor
Master of Science Thesis, 50 pages, 11 Appendix pages
December 2014
Major: Measurement Technology
Examiners: Professor Jukka Lekkala, Professor Ilkka Korhonen

Keywords: optical heart rate measurement, photoplethysmography, electrocardiography

Older technologies, which might have been the golden standard in the industry for years, are rapidly becoming available to a wider audience as manufacturing methods become easier and cheaper. Companies are able to provide every consumer the same devices which have been the privilege of only the professional field. This has also been the case with fitness wearables, of which one subclass is the optical heart rate sensors. The goal of this thesis was to evaluate the performance of one such device, namely the PulseOn wrist device.

The device utilizes photoplethysmography (PPG) in acquiring the heart rate signal. PPG has been used in clinical settings for oxygen saturation level determination, but the technology can also provide other figures from the cardiovascular system, such as heart rate. The measurement method is based on the detection of light, which is emitted into the skin and then interacts with the tissue. The composition of the blood vessels changes in synch with the beating of the heart, and so does the intensity of the detected light.

The PulseOn device was tested in controlled laboratory conditions with 20 subjects. The measurement protocol included periods of rest and activities of varying intensities. A reference measurement was made simultaneously with a Polar heart rate belt, and also two other devices were used to record data for later assessments.

The results were analysed in MATLAB, and values for heart rate reading reliability and measurement errors were calculated. For example, the correlation of the PulseOn device against the Polar belt was found to be approximately 96 %, the amount of readings that were within 10 % of the values given by the heart rate belt was 90.4 %, and the average value of the absolute errors between the two devices was 4.76 beats per minute.

Even though the PulseOn device was still in its development phase at the time of the measurements, it showed satisfactory results, and that it could be used in the heart rate measurements of everyday fitness activities.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Automaatiotekniikan koulutusohjelma
HAAVIKKO, ALEKSI: Optisen sykemittarin suorituskyvyn arviointi
Diplomityö, 50 sivua, 11 liitesivua
Joulukuu 2014
Pääaine: Mittaustekniikka
Tarkastajat: professori Jukka Lekkala, professori Ilkka Korhonen

Avainsanat: optinen sykkeenmittaus, fotopletysmografia, elektrokardiografia

Vanhat teknologiat, jotka ovat saattaneet olla alan kultainen standardi vuosien ajan, ovat nopeasti tulossa saataville laajemmalle yleisölle valmistusmenetelmien tullessa helpommiksi ja halvemmiksi. Yritykset voivat tarjota jokaiselle kuluttajalle samoja laitteita jotka ovat olleet vain ammattilaiskentän etuoikeus. Tämä on myös tapahtunut puettavien hyvinvointilaitteiden kohdalla, joiden yksi alaluokka ovat optiset sykesensorit. Tämän työn tavoitteena oli arvioida yhden tällaisen laitteen suorituskykyä, nimenomaisesti PulseOn rannelaitteen.

Laite käyttää hyväkseen fotopletysmografiaa (PPG) havaitakseen sykesignaalin. PPG:tä on käytetty sairaalaolosuhteissa happisaturaatiotason määrittämiseen, mutta teknologialla on mahdollista saada myös muita lukemia verenkiertoelimistöstä, kuten syketaajuus. Mittausmenetelmä perustuu valon aistimiseen, joka lähetetään iholle ja on sitten vuorovaikutuksessa kudoksen kanssa. Verisuonien koostumus vaihtuu synkronoidusti sydämen sykkeen kanssa, ja samoin vaihtuu myös aistitun valon voimakkuus.

PulseOn-laitetta testattiin kontrolloiduissa laboratorio-oloissa 20 koehenkilöllä. Mittausprotokolla sisälsi lepojaksoja ja vaihtelevaintensiteettisiä aktiviteetteja. Referenssimittaus suoritettiin samanaikaisesti Polarin sykevyöllä, ja myös kahdella muulla laitteella tallennettiin dataa myöhempää arviointia varten.

Tulokset analysoitiin MATLAB:ssa, ja arvoja laskettiin sykelukeman luotettavuudelle ja mittausvirheille. Esimerkiksi PulseOn-laitteen korrelaatio Polariin nähden oli noin 96 %, sykelukemien määrä, jotka olivat 10 % sisällä sykevyön lukemasta, oli 90.4 %, ja laitteiden välisten absoluuttisten virheiden keskiarvo oli 4.76 lyöntiä minuutissa.

Vaikka PulseOn-laite oli vielä kehitysvaiheessa mittausten aikaan, sillä saatiin tyydyttäviä tuloksia, ja laite osoitti että sitä voidaan käyttää sykkeen mittaamiseen jokapäiväisissä kuntoiluaktiviteeteissa.

PREFACE

This thesis was done for PulseOn Oy, and the work was carried out at the Department of Signal Processing at Tampere University of Technology. The measurements took place during the summer of 2014 at the VTT laboratories in Hervanta, and the written part was done later during the autumn.

I would like to thank both of my examiners, professor Jukka Lekkala from the Department of Automation Science and Engineering and professor Ilkka Korhonen from the Department of Signal Processing and PulseOn. Professor Korhonen offered me the topic after I got involved in earlier measurements for the PulseOn device already during 2013, and guided me afterwards in the creation of this thesis with many helpful advice and comments.

I would especially like to thank doctoral student Jakub Parak, who worked in the same room with me last year and got me involved in the measurements before I even knew about PulseOn. He has also helped me with numerous smaller and more significant problems I encountered during my work.

Lastly I want to express my love and gratitude to my wife Anna for all the encouraging words and motivation which kept me going to reach my goals. And my final motivator, our little newborn Silja, this is also for you to see later what Dad did while waiting for your arrival. You are so dear.

In Tampere, Finland, on 17 November 2014

Aleksi Haavikko

TABLE OF CONTENTS

ABBREVIATIONS

AC	alternating current
AD	(average) absolute deviation
BMI	body-mass-index
BP	blood pressure
bpm	beats per minute
CV	coefficient of variation
DC	direct current
ECG, EKG	electrocardiogram, electrocardiography
GPS	global positioning system
HR	heart rate
HR_i	instantaneous heart rate
LAN	local area network
LED	light-emitting diode
MAD	mean absolute deviation
MAE	mean absolute error
MAPE	mean absolute percentage error
ME	mean error
MSE	mean squared error
NRMSD	normalized root-mean-square deviation
PPG	photoplethysmography
PVC	premature ventricular contraction
RMSD	root-mean-square deviation
SAE	sum of absolute errors
SEE	standard error of estimate
SSE	sum of squared errors

1 INTRODUCTION

Technology, including health technology, has taken many strides forward in the past century, and at the same time it has been brought ever closer to our everyday lives. Consumers are nowadays using the same gadgets that were first designed for clinical or military use. Probably the best-known example of this is the global positioning system, GPS. This military technology is used every day by the layman hiker or geocacher of the 2010s.

The same kind of movement can now be seen in devices which monitor physiological measurands. The term quantified self (Quantified Self, 2014) has been coined to describe people who want to implement a kind of a bio-feedback system to their bodies by measuring everything possible. Starting with the usual height and weight, people can nowadays buy devices to record their glucose levels, blood pressure and heart rate, just to name a few, and do it all continuously through day and night. Sleeping patterns are analyzed in the morning to see where the quality of sleep hasn't been the most beneficial, and actions can be taken to counter those defects in sleeping posture or environment. Every food can be photographed to later assess the calorie intake of the meal and adjust the diet accordingly, and the flow of everyday life is scheduled to be as optimal and care-free as possible with various calendar applications.

Of course, all of this aims to develop the individual, both physically and mentally, to have a better way of life or to manage easier through normal daily routines. One way to achieve this goal is to be physically in a top-notch condition. That is why most people go running or cycling every other day; not necessarily because they love the struggle and sweat exercising brings, but because it is good for their health and general well-being. Some individuals might be just training with a specific goal in mind, for example a marathon. Whatever the purpose, there is always some method by which to make the training more organized and fruitful, and that is where the quantification comes in.

The old saying goes that if you can't measure something, you can't control it, and that means you can't improve it. And since your performance during exercising is usually something you want to improve, the most effective would be to measure your performance, both during and after the exercise itself. For this purpose companies like Polar (Polar Electro, 2014) have provided heart rate monitors for runners since the 80s, and also other Finnish companies have followed, for example Suunto (2014). What used to be the privilege of clinical practitioners and professional trainers became available to all consumers wanting to see their heart rate online and track it from training to training. Since the early days of heart rate sensors, and for a good time until the end of this millennium's first decade, these devices have required the use of a heart rate belt worn around the thorax. Although this belt has been the golden standard for joggers and cyclists for years, it certainly is not the most comfortable thing to wear, let alone to keep in place during sweating and rigorous movement. For a certain body shape, and especially for women, it can be extremely difficult to have the belt stay in one place, and that, of course, leads to inaccuracies in the readings.

Different companies have started to notice this urge of people to be more aware of their diet and exercise schedule and activities in general. One way to achieve this is to monitor and track as much information as possible, like the quantified self –movement has been doing for some years already. What better way to incorporate measurement devices into peoples' lives than attaching them to appliances we nowadays carry with ourselves everywhere all the time. Probably the easiest transition to a regular exerciser is to have these sensors in the wrist device that they have carried with themselves on the jogs for years. These wrist devices can even be made into smartwatches with e-mail and calling capabilities, but there is only so much capacity in one small apparatus, that usually just some qualities of the device can be optimized for the price of the others' performance. This has been sadly true with Samsung's (Samsung, 2014) latest Gear series smartwatches. It is usually better to do one thing correctly, than to try to do many things half-way at the same time. It may be easy to merge a phone's microprocessor and some sensors in a tiny frame of a wrist-worn device, and measuring reflected light on a photodetector to sense heart beat may sound like a simple task. In reality, though, the human physiology is a complex, comprehensive process, and measuring its signals, let alone making some sense of them for the average consumer, is far from a menial task.

When it comes to exercising, heart rate usually tells more than enough about the intensity and variability of the work-out. However, to put the capabilities of heart rate sensing into one comfortable package, that would be easy to use and still give meaningful insights into your training, is not just a simple job. There are many things to consider in the measurement signal and all the different functional or outside sources of noise and errors.

The objective of this thesis was to the evaluate performance of a new optical heart rate monitor, PulseOn (2014). In addition, this thesis aims to point out some of those difficulties that can arise when doing an optical heart rate measurement, and present methods to assess the gravity and quality of those measuring defects.

The structure of the thesis is as follows: first, we will go through the fundamental physics and underlying physiology of the human cardiovascular system that make the optical heart rate measurement possible. Then the principles of performance estimation are presented along with the importance of noise and the role of a reference device in its cancellation. In chapter four different devices for optical heart rate measurement are introduced. The methods in the performance evaluation recordings done for this thesis are portrayed in chapter five and the following chapter presents the results of these measurements. Finally, the results are compared and discussed, and a summary is given.

2 OPTICAL MEASUREMENT OF HEART BEAT

The most common way to measure the beating frequency of the heart is to do it electrically. The golden standard, and the most used method in clinical practice, is the electrocardiogram ECG. This recording detects the changing voltages on the body surface in different locations, and deducts the direction and magnitude of the heart vector. This electric dipole is the result of the heart muscle's de- and repolarization that happens during beating of the heart. The frequently used 12-lead ECG system, uses electrodes on both arms and the left leg, and also six on the chest, the locations of which is portrayed in Figure 1. By measuring voltages between different points one can determine the part of the heart's electrical vector that is parallel to the lead pair. Pairs are formed from each combination of two limb electrodes, from each limb and the average of the other two, and each chest electrode is paired with the average of all the limb potentials. This average is also called the Wilson central terminal. Three distinct sections can be seen in one heart beat in the electrocardiogram, which are all caused by different depolarization and repolarization phases of the heart's atriums and ventricles. These are the P-wave, the QRS-complex and the T-wave.

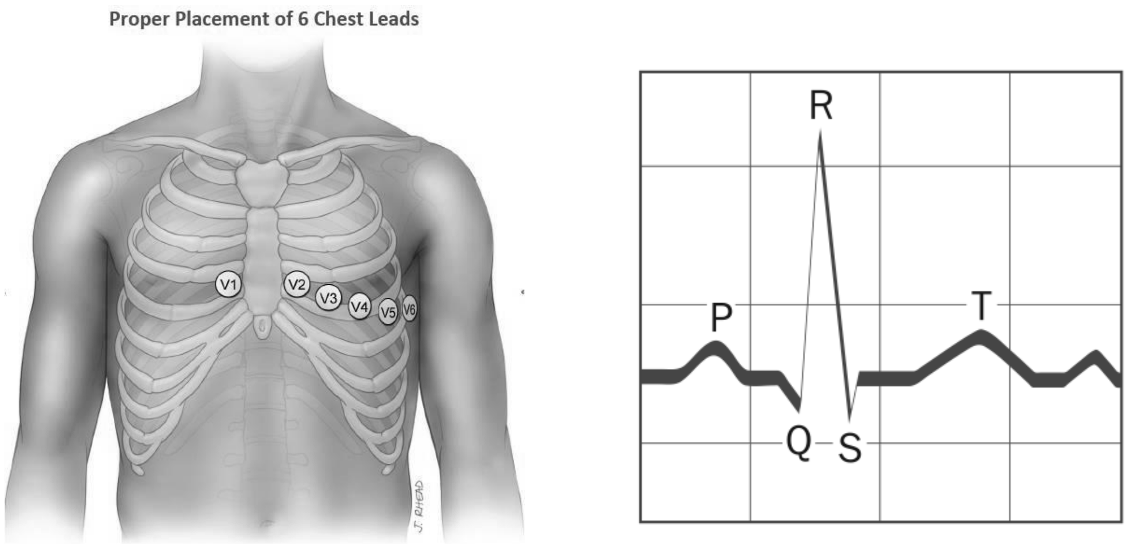


Figure 1. Placement of the chest electrodes (left; Eccles Health Sciences Library, 2014) and the different segments of the ECG signal (right; LearntheHeart.com, 2014)

Heart rate is usually defined as an average beat count over a certain time window. This is expressed in units of beats per minute or bpm. Instantaneous heart rate can be deducted from two consecutive QRS-complexes, however, with equation (1)

$$HR_i = \frac{1}{RR} \times 60[bpm], \tag{1}$$

where RR is the time interval between the two R-spikes in seconds. Electrical measurement of heart rate is the most effortless practice nowadays, since it has been developed for over a hundred years and gives extremely accurate results. ECG has its downsides too; the electrodes require a firm contact to the skin, with materials that are usually at least irritating to the skin in continued use, if not even allergenic. In a clinical setting the electrodes of the 12-lead ECG require a lot of wires, which make the movement of the patient difficult if not impossible. Normally this is satisfactory, because the patient is supposed to stay in the bed and hold still. Capacitive measurement systems, that can be integrated into the bed sheet for example (Vehkaoja et al., 2014), have been made to remove the wires from attaching to the patient's skin, and with these the person can sleep normally in their own bed through the entire night.

Even capacitive measurement cannot handle the recording of the heart beat through the whole day, especially during exercise. It is very sensitive to movement artefacts, and these are probably the most important aspect affecting heart rate measurement designed for workout monitoring. Heart rate belts have been doing a decent work in this for years already, but these systems also have the same restrictions as clinical ECG, although in a somewhat different form. The belt has to be worn on the chest, pushing firmly to the skin. It is normally moistened to make the impedance between the electrode-skin-interface more ideal. The changing of this impedance affects the measurement signal and is a major factor in the measurement error. If the needed skin contact is not achieved or there is not enough friction or tightness to hold the belt in place, the moving of the belt can cause not only error readings in the heart rate measurement, but also irritation on the skin. Some people may find the belt to be uncomfortable to wear altogether, and it is obviously one more additional device to carry with you during your training along with the wrist unit. This also brings up problems in wireless connection between the two devices.

The optical measurement of heart beat means to tackle many of these obstacles all at once. The measuring electronics can be all integrated into a small wrist device, because the measurement is done by shining light inside of the skin on the wrist with small LEDs. No additional wires or gear is needed, and all the filtering algorithms and signal processing is done on the same chip. The basics of this measurement principle are presented in this chapter.

2.1 Blood flow in the veins during heart beat

The human heart pumps continuously to deliver enough oxygenic blood to different tissues around the body. Proper blood pressure and heart rate guarantee that every single capillary receives adequate blood supply to be delivered to the surrounding tissue. The maximum blood pressure in the circulation system is in the aorta right after the heart has pumped one stroke volume of blood out. As the blood flow continues towards the tissues of the body the pressure in the arteries decreases gradually and at the capillaries there is a distinct drop when the arteries split into several smaller ones. This guarantees that the blood has enough time to change gases at the destination site, before gathering again in to larger veins of deoxygenated blood. In the veins the pressure is almost non-existent, so that there is a pressure gradient large enough to keep the blood flow at a required level.

As the heart contracts periodically, so does the arterial blood pressure fluctuate with some delay. This synchronization can be seen in Figure 2 below. The highest pressure during this cycle is called the systolic, and the lowest is the diastolic blood pressure. The difference between these two values is called the pulse pressure and it is usually about a quarter of the systolic blood pressure. This pressure wave front travels faster in the arteries' walls than in actual blood, but the same fluctuation can be seen in the flow volume. The elasticity of arteries allows them to expand as a larger volume of blood passes through and then contract back to their original size. The elastic fibers that allow this to happen are also responsible for keeping up the pressure gradient initiated in the ventricles.

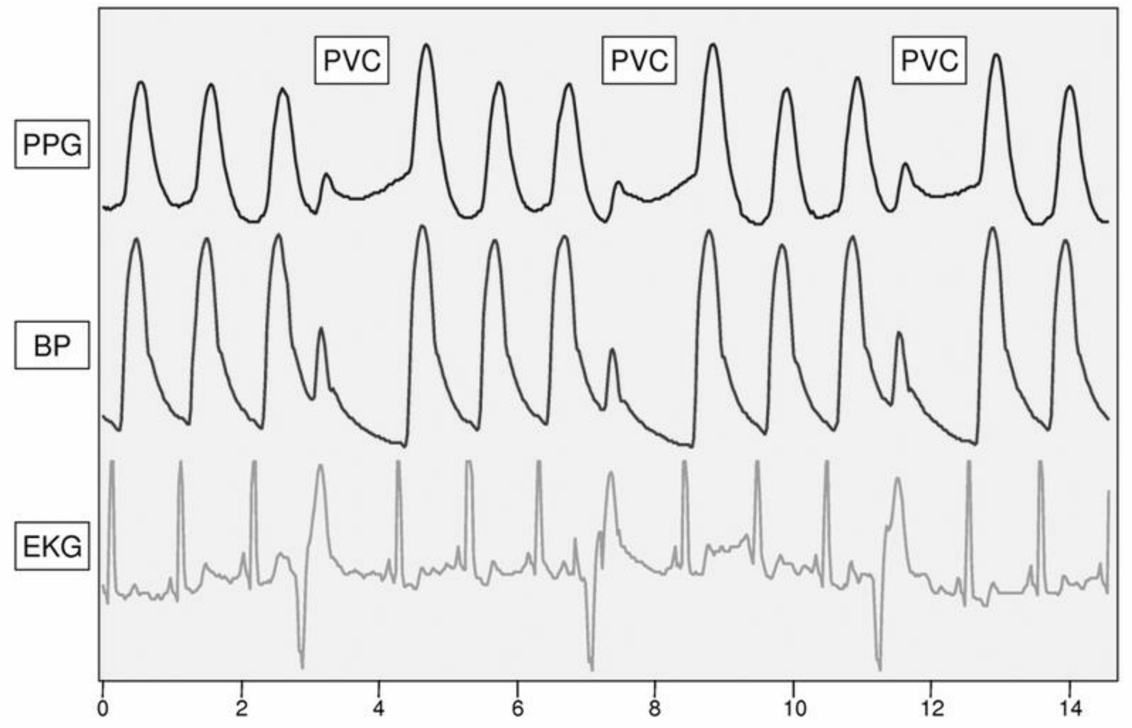


Figure 2. Simultaneous plots of a photoplethysmograph, blood pressure and EKG lead with a few cases of a premature ventricular contraction (PVC) (Spl4, 2006)

Many factors affect the flow of blood and the blood pressure. The amount of blood that is pumped from the heart to the systemic circulation is dependent on the stroke volume of each heart beat and the heart rate, that together make the cardiac output, expressed in volume units per time unit. This blood volume flow is being resisted by the arteries, which have several factors that contribute to the flow resistance. The most important one is the compliance of the artery walls. This is a measure of the elasticity that allows the blood vessels to expand and recoil, and it decreases with age. As the arteries get stiffer and stiffer, the blood pressure increases which in turn demands more activity from the heart. Bad cholesterol can also build up inside the vessel walls and constrict the blood flow, resulting in higher resistance and therefore increased blood pressure.

Blood vessel’s resistance to blood flow can vary from person to person because of different lengths of the vessels and their diameter. These two variables together with the viscosity of the fluid, blood in this case, are the components that make up the overall resistance of any cavity that has fluid flowing in it. In the systemic circulation only the diameter of the blood vessel can change quickly while the other two are somewhat constant during a short inspection interval. Signals from the neural system can change the vascular tone rather rapidly by controlling the smooth muscle fibers surrounding the vessels, which determines the diameter of these vessels. The length of the vessels is, of course, increasing during childhood, but for adults it stays the same in normal conditions. The viscosity of blood may also change due to the illnesses of the blood cells themselves or the liver, which creates most of the plasma proteins in the blood. These changes don’t usually happen too quickly, though.

Even if there wouldn’t be any neural or chemical actions affecting the blood vessel diameter, their overall cross-sectional area is not uniform along the systemic circulation. The area of the individual arteries is much larger than those of the smaller arterioles, but there are many times more of these arterioles all around the body, so that the sum of all the areas of these arterioles is much larger than the cross-section of the arteries. While the area is much larger, the overall resistance to the blood flow is also greater when getting closer to the capillaries because of the smaller individual cross-section areas. That is why the blood pressure drops significantly faster when the flow reaches the smaller vessels although it decreases at least a fraction everywhere throughout the blood circulation. The pressure drop also means a slower flow speed of the blood, which in turn assures that the tissue, where the capillaries are, gets enough oxygen from the blood because the tissue and the blood have adequate time to change gases between themselves.

2.2 Optical properties of tissue

Light passing through a substance can be subjected to various phenomena depending on the characteristics of the matter. Factors affecting the light include, for example, the density and the color of the substance. Denser matter absorbs and reflects light more than sparser matter. Also, a darker material absorbs more photons than a lighter one. Regarding the optical properties of human tissue, and specifically the skin and blood veins, these qualities are realized in the form of skin tone and different inhomogeneities within the tissue.

In their review, Anderson and Parrish (1981) summarize the optical characteristics of all the layers of human skin starting from the outer-most stratum corneum, and going through epidermis and dermis. When considering a nearly perpendicular light coming to the skin, the differences in the refractive index of the outside air and the stratum corneum cause a small fraction of the light to be scattered back. This is called the regular reflectance. When the light penetrates the first layer, it can be subjected to either further scattering in any direction, or to absorption. These two phenomena determine the amount of light that gets to penetrate the deeper layers of the skin, and finally other tissues. Scattering happens because of inhomogeneities in the tissue, and the amount of scattering is determined on the physical size and shape of the inhomogeneity and the difference in the refractive index that this portion of the skin has with the surrounding area. The strongest scattering occurs when the inhomogeneity is about the size of the wavelength of the light, and this scattering is directed mostly forward. Smaller or larger inhomogeneities have less of an impact on the scattering, and for small defects the scattering profile is more isotropic.

In reality, when a beam of light is subjected to the skin, the photons that penetrate the skin scatter multiple times within the tissue, and the overall distribution becomes highly isotropic, or diffuse. This type of electromagnetic radiation can be shown to travel a total distance of 2 dx, when considering an infinitesimal space of size dx. This makes theoretical calculations more straightforward and is assumed in the modeling by Anderson and Parrish (1981). They use a highly simplified model, called the Kubelka-Munk model. With this they define two factors, the back-scattering and absorption coefficients, from two simple differential equations given below.

$$dI = (-KI - SI + SJ)dx \tag{2}$$

$$-dJ = (-KJ - SJ + SI)dx, \tag{3}$$

where I and J are the inward and outward fluxes of light, respectively, K is the absorption coefficient, S is the back-scattering coefficient, and dx is the thickness of a small layer of skin. For example, equation (2) states that over a distance dx, the inward flux of optical radiation is decreased by the amount of light back-scattered and absorbed in that space, and increased by the intensity that is back-scattered of an outward flux moving in the opposite direction. Equation (3) gives the identical change in flux for the outward moving light. Remittance is the portion of the light going inside the tissue that is scattered back or $R = J_0/I_0$ expressed as an equation. Transmittance, in turn, is the ratio of the inward light that is transmitted through the whole tissue to the other side, that is $T = I_D/I_0$. If the tissue is thick enough, the transmittance approaches zero. This is normally true when considering the human skin from the surface to the dermis and with wavelengths of less than 600 nm. Using this information and by integrating equations (2) and (3), one can derive a simple equation for K and S as

$$\frac{K}{S} = \frac{(R-1)^2}{2R}. \tag{4}$$

In the skin, the structures that define the amount of scattering from it are different from the chromophores that are largely responsible for the absorption of the light. The scattering characteristics are also quite stable in normal conditions unless some notable change occurs. This means that the scattering coefficient can be thought of as a constant in equation (4), so that the absorption coefficient depends only on the remittance, and it is changing rapidly because of the continuous alterations in the density and distribution of hemoglobin, bilirubin and melanin. Also, as the stratum corneum and the epidermis are mostly thin enough, one can deduct from the equations that their contribution to the remittance is minimal.

The role of melanin is important when considering the optical properties of the human skin. The amount of melanin determines the color of the skin, as it is the main absorber of light in the visible spectrum. The transmittance of skin can vary multiple-fold between fair- and dark-skinned individuals. However, melanin doesn't absorb wavelengths uniformly. It actually absorbs shorter wavelengths better, so that at the longer infrared wavelengths the absorption of light is almost non-existent. In their study, Fallow et al. (2013) investigated the influence of the different skin types and wavelengths of light on the light reflectance from the wrist. They had 23 subjects with varying skin colors and they used four wavelengths of light; blue, green, red and infrared. In the study, they concluded that at rest green light had the best modulation factor, and in exercise conditions either blue or green had the highest signal-to-noise ratio, depending on the skin type. Fallow et al. also noted that the darkest skin type had the poorest signal quality when compared to the lighter skin types and that there was no significant relation between the skin types while resting and doing exercise. This, they deducted, was because melanin affects the light in the epidermal layer of the skin where there are no blood vessels, and so it is a static factor that has the same effect no matter what the conditions are. The wavelength dependence on light interactions in tissue can be seen in Figure 3 below.

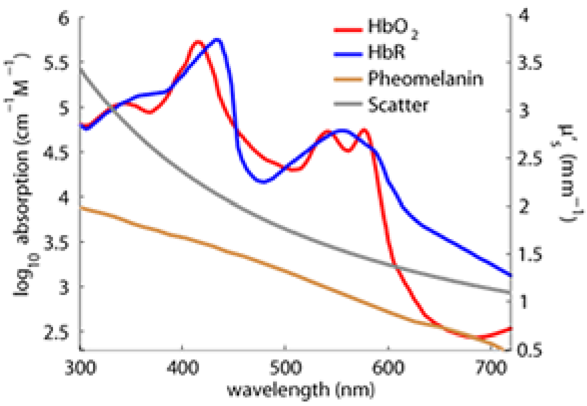


Figure 3. Absorption and scattering of light from different tissue constituents as a function of wavelength (Hillman Lab, 2012)

2.3 Basic principle of photoplethysmography

Plethysmography is the measurement of changing volumes. Photoplethysmography (PPG) utilizes light to detect these differences in volume. A beam of light is shone from a light source at a specific intensity and wavelength. While travelling in the tissues, the light is subjected to various optical phenomena like scattering, absorption and reflection. In a subject medium that has stable consistency, the optical properties, like absorption coefficient, are somewhat constant. In the measurement signal, this accounts for the stable or DC-level component, seen in Figure 4. However, human tissue is highly vascularized, and its composition and physical shape change slightly all the time, mostly because of blood flow. These changes are periodical and synchronized with the beating of the heart, and can be seen as the changing AC-level component in the measured photoplethysmography signal.

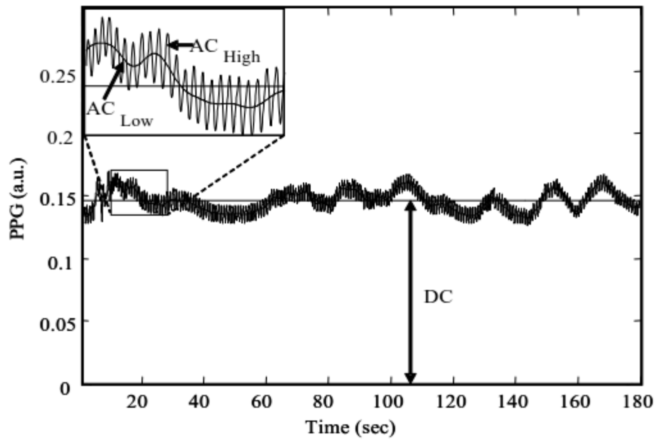


Figure 4. A photoplethysmography signal showing the DC and AC components (Huang et al., 2011)

When measured from the wrist, the pulsation of blood can be seen some tens of milliseconds after the actual beating of the heart muscle. Two distinct parts can be seen in the signal; the more rapidly increasing, rising part also known as the anacrotic phase, and the slower descending part, or the catacrotic phase. These phases correspond to the systole and diastole of the blood pressure. The shape of the pulse signal also depends on the location of the measurement site. Closer to the aorta the rising part can be much steeper than in the more distal parts of the body where the blood pressure has already dropped significantly, and it can start to resemble the slope of the decreasing phase. Because of this it can more difficult to say where the pulse is happening in time exactly. The height of the pulse also gets lower, which makes it tedious to recognize the heart rate without sophisticated signal analysis algorithms. On the other hand, when the distance from the heart gets greater, the area of the cross-section of individual vessels also gets smaller, which results in changes in the volume that are relatively much larger than those more proximal to the heart.

Nowadays, the light source of the usual photoplethysmography device is a light-emitting diode, LED. They are relatively cheap and easy to get in today's semiconductor markets, and can be made to serve any need with different wavelengths of the emitted light. It has been suggested that in certain conditions a green light might be more beneficial than, for example, an infrared light (Maeda et al., 2008). A sensitive photodetector, that matches the spectrum of the LED, converts the detected light into an electrical signal that can be further filtered and amplified in a measurement circuit.

Two main modes can be characterized in photoplethysmography measurements: the transmission and reflection modes. Transmission means that the light that is picked up travels through the entirety of the tissue under measurement, and the detector site is on the opposite side of the tissue. This mode only senses the light that is moving straight forward without any scattering or absorbance. An example of the transmission mode can be seen in the oxygen saturation meter of Figure 5. Of course, the transmission mode can also pick up light that has been reflected from some other site adjacent to the actual measurement site, or light that is reflected several times, but this depends largely on the tissue being measured. In reflection mode, the majority of the light's intensity detected is, optimally, reflected only from the capillaries that are supposed to be measured with the device.

Figure 5. A blood oxygen saturation meter using transmission mode of PPG (Rama, 2005)

2.4 Interference affecting the measurement

Because photoplethysmography is based on the measurement of light intensity, the most harmful effector of interference is that of the light that is coming from other sources than the intended measurement LED. Normal artificial light in usual office conditions doesn't necessarily amount to a lot of noise in terms of intensity since it is somewhere around a few thousand lumens. Moving inside from one type of artificial lighting to another, however, can cause irregular interference that can be hard to compensate, though typically this should be seen in the 50 Hz frequency. Sunlight, however, can be much brighter, especially in the summer with intensities reaching over hundred thousand lumen. It may be somewhat easier to remove its effect because of quite constant intensity, but obviously moving under trees or such can make sunlight as difficult as, or even trickier to account for than artificial light when it comes to noise cancellation.

Some notable sources of interference come from different movement artifacts, and systems to reduce these have been widely studied (Hayes & Smith, 1998; Foo & Wilson, 2006). Most types of movement of the device or the subject medium should express a fault in the measurement signal. Tilting or sliding of the measurement device alters the path that light travels, and changes the composition of the pathway it is taking. Changes in pressure on the skin also have an effect on the signal. Even if the device would stay in its place related to the subject skin, there can be movement noise if the system is moving in a brightly lit area. Depending on the ambient light's intensity, it can penetrate the skin from all directions and through the whole body, and add to the amount of sensed light on the sensor.

Photoplethysmography is also affected by various sources of interference as any other electrical measurement circuit normally is. The most common 50 Hz noise from other electrical appliances and wires can couple also to this type of measurement though this is nowadays normally accounted for automatically.

2.5 Measurement methods

Allen (2007) has given a topical review to the novel clinical applications of PPG, and Tamura et al. (2014) have reviewed different wearable photoplethysmography sensors of the past and present. Already in 1935, German physician Karl Matthes developed the first ever device to measure blood's oxygen saturation. This device used the same method as the most temporary clinical apparatus, the transmission mode. In this mode light is shone through the tissue and the amount of oxygen in the blood is proportional to the absorbed light on the way. This mode requires the tissue to be thin so that at least some light is emitted through it, so for example the fingertip is widely used for this kind of measurement. The alternative reflectance mode doesn't have this requirement, and so it has more possibilities in the placement of the sensor. The light only needs to travel deep enough to reach capillaries, and is then reflected back at the sensing unit.

For the transmission mode, the light of the longer wavelengths is more suitable because it is absorbed less in the tissue and can penetrate through to the other side. This is used largely in pulse oximetry, where information of the constitution of blood is needed. Partly because of these same characteristics, this red spectrum of visible light is not equally suitable for reflection mode photoplethysmography. More scattering happens for example with a green light, and so the detected signal level is higher (Maeda et al., 2008).

3 ESTIMATION OF THE PERFORMANCE OF HEART RATE MEASUREMENT

When estimating the performance of a particular process, one can use a variety of metrics to assess different aspects of the process. In a production facility, for example, the performance might mean the amount of products manufactured in a day, or the use of raw materials and time for a given task. In determining the performance of a heart rate measurement, performance usually means how close to the truth the readings are. This can be looked at one reading at a time or as a whole across the whole measurement as explained later.

In the following chapter, we will go through some performance metrics regarding heart rate measurement, and how different kinds of errors and deviations can be calculated from these. These methods have been largely covered by Morris & Langari (2012), for example. One section will also explain how different sources of interference affect the signal, and how these errors can be identified and removed. In the last chapter, we will go through the role of a reference measurement signal.

3.1 Accuracy and reliability

Measurement error may be defined as a difference between the true value x and its estimated value $\hat{x}$ given by the measurement as

$$e = x - \hat{x}\,. \tag{5}$$

In measurement technology, the accuracy and precision of a measurement are two different types of metrics and can have more value in different contexts. Accuracy is defined as “the extent to which the results of a calculation or the readings of an instrument approach the true values of the calculated or measured quantities, and are free from error (McGraw-Hill Global Education Holdings, 2014)”. Precision, on the other hand, is defined as “the measure of the range of values of a set of measurements; indicates reproducibility of the observations (McGraw-Hill Global Education Holdings, 2014)”. Precise measurement values at different times can be close to each other value-wise, but that doesn’t mean that they are necessarily accurate, that is, close to the truth. They have a certain bias that is off from the actual value of the measurand. On the other hand, more accurate results in a measurement might be less precise. Which type of performance is more desired depends entirely on the requirements of the measurement.

In heart rate measurement it is usually more favorable to get heart rate values that are as close to the real heart rate as possible. Since the heart rate is normally expressed as beats per minute, it means that number value is just a momentary estimation of what the heart rate would approximately be over the course of one minute period. This allows for somewhat looser quality of precision, because the completely true heart rate changes every second and beat to beat. So, it is more desirable to have results that are more accurate than precise. In practice this could mean, for example, that for a runner it is much more valuable to know that they are in the right training intensity zone determined by some heart rate boundaries, than it is to have exactly the same heart rate during the whole run. Especially so if the so called precise heart rate would actually be ten beats off the true value across the whole run.

The reliability of a measurement can mean a few things depending on the situation. It can be understood to mean the same as precision in some cases. In psychometrics it can be used as a measure to evaluate the agreement between two different raters, as so called inter-rater reliability. This can be applied even to nominal data that represents a patient's condition on some qualitative scale from healthy through slightly ill to sick, for example. One common interpretation is also reliability meant as the amount of consistency in one measurement method when done at two different time instants. This measure is used, for example, in measurement device calibration, where the readouts of a device between calibrations can start to deviate slowly from the optimal due to wearing of the mechanical composition of the device. This is also called the test-retest reliability.

Another type of reliability is called inter-method reliability, where two different methods of measurement are being compared. This can be thought to be the type of reliability we are going to focus in this thesis. We are comparing two solutions that are based on different technologies, and want to know how much they agree on the value of the physical quantity being measured, that is, the heart rate. Of course, in our case we assume that the other method, also called the reference device, is accurate all the time. We then want to get just the reliability of the first device's readings compared to this reference.

In this case, the reliability is a measure of how close our device under evaluation is to the true value, or more specifically, how often it is close enough to be said to be accurate. The definition of accuracy with this interpretation can be a predefined percentile zone above and beyond the real value. For example, we can say that the measurement is accurate enough if the value is within five percent of the true value. Then we calculate the portion of all the measurement-real value –pairs that fulfill this definition of accuracy, and express this ratio as percentages out of one hundred. This value then tells us how often our measurement device can be said to give reliable results. The accepted neighborhood around the true value can also be expressed in absolute units, but it depends on the application whether this approach is justified performance-wise. It is obviously more forgiving than the relative boundaries, but on the other hand it is also a more static option for calculation purposes, for example.

3.2 Estimation principles and error variables

The nature of the measurement determines which type of performance measure suits best to describe the errors in the measurement. The simplest method of determining the performance of a measurement reading is to express the amount of its difference to the actual value of the variable. This absolute error can be transformed to a more describing relative error by expressing its percentage of the actual value. Absolute error is useful when we are particularly interested in the value of the error, and relative error is used when the ratio of the measurement result and error stays quite constant. If the measurement range can be adjusted, and this might affect the result, a comparison error can be calculated. This error is the ratio of the absolute error to the particular measurement range. These errors can be calculated for all value-measurement pairs separately, and are not normally that useful by themselves.

Various kinds of metrics for measurement results can be derived from traditional statistics. Basic average and standard deviation are just the simplest forms of statistical values for a dataset and can be used for a satisfactory assessment. There are, however, many more values that can be derived and used to describe measurement results. Many of these are introduced in the online book by Lane (2014).

The simplest form to use is the mean error (ME), which is defined as

$$ME = \frac{\sum(x-\hat{x})}{n}, \tag{6}$$

where x is the true value, $\hat{x}$ is the measured value, and n is the number of measurement points. Using the absolute errors, one can derive a mean absolute error (MAE), which is similarly to ME defined as

$$MAE = \frac{\sum|x-\hat{x}|}{n}. \tag{7}$$

The mean squared error (MSE) is the average of the absolute errors squared, given by equation

$$MSE = \frac{\sum(|x-\hat{x}|^2)}{n}. \tag{8}$$

These two latter measures don't take into account the direction of the error, which may not be relevant in all types of measurements. This also means that errors of the same magnitude but opposite direction won't cancel each other out. The mean absolute percentage error (MAPE) is of the same form as MAE, but the absolute errors are given as a percentage of the actual value as

$$MAPE = \frac{\sum(\frac{|x-\hat{x}|}{x} \cdot 100\%)}{n}. \tag{9}$$

The amount that the values of a certain measurement set deviate from the recording's mean is called the standard deviation. It is calculated with the equation

$$S = \sqrt{\frac{\sum(\hat{x}-\mu)^2}{n}}, \tag{10}$$

where μ is the mean of the measurement values, and $\hat{x}$ are now the individual measurements. This value has the advantage of having the exact same unit as the measurement results themselves. Depending on the case, if the aforementioned sample only represents a part of a larger population of measurements, we use n-1 in the division when calculating the value, n being the size of the sample. This kind of a more uncertain statistical value is called the sample standard deviation, and it is naturally somewhat higher than a normal standard deviation would be.

The dispersion of a statistical data set can more generally be defined by the average absolute deviation, which is the amount of deviation from a certain central value defined as

$$AD = \frac{\sum|\hat{x}-X_M|}{n}, \tag{11}$$

where $\hat{x}$ are the measured values, X_M is the central value, and n is the number of measurement points. The choice of the central value affects the result of the dispersion measure. The median of the data set is normally used as this choice gives the lowest average deviation for every data set; any other value can only be equal to it, never less. Also, the average absolute deviation from the mean of the set is always less or equal to its standard deviation. This deviation from the mean is usually what is meant when talking about mean absolute deviation (MAD) and it is considered a better descriptor of the dispersion than standard deviation because it connects better to actual values.

When talking about a model that forecasts or estimates the true value of some phenomena, the standard deviation of a modeled sample's difference to the actual value is also called the root-mean-square deviation (RMSD). It is calculated in the same manner as standard deviation. The normalized RMSD (NRMSD) can be derived by dividing the RMSD by the range of the true value, and the coefficient of variation (CV) is the ratio of the RMSD to the mean of the actual values. These measures are defined as the equations below.

$$RMSD = \sqrt{\frac{\sum(x-X)^2}{n}} \tag{12}$$

$$NRMSD = \frac{RMSD}{x_{max} - x_{min}} \tag{13}$$

$$CV = \frac{RMSD}{x_M} \tag{14}$$

In these equations x is again the true value, x_{max} and x_{min} being the maximum and minimum values, x_M is the mean of all the actual values, X is the value given by the estimator model and n is the number of data points.

RMSD can also be used to describe the differentiation between two variables even if neither one of them is really considered the true value. One can think of the measurement setup in this thesis, or in fact any measurement where two values are being compared, to be like this; even though we compare the optical device to the reference heart rate belt, we cannot be totally sure of the belt's accuracy, either. Nevertheless, the calculations for RMSD are still the same.

The sum of squared errors (SSE), which is also sometimes called the residual sum of squares, is the sum of all the squared errors of individual observations to their true counterparts, given by equation (15). Likewise, the sum of absolute errors (SAE), equation (16), sums the absolute values of these errors.

$$SSE = \sum (x - \hat{x})^2 \tag{15}$$

$$SAE = \sum |x - \hat{x}| \tag{16}$$

Standard error of the estimate (SEE) is the standard deviation of the measurement value from the true value, and is calculated as

$$SEE = \sqrt{\frac{SSE}{n}}, \tag{17}$$

where n is the number of measurement points. Standard error of the estimate is more used with regression analysis, but it can also be implemented in our reference-comparison measurement evaluations.

Regression analysis is a way to define the dependence of two data sets from each other. It is normally used in prediction and modeling to estimate how well a predictor or simulated model represents the truth. The most common method is to use linear regression, where it is usually presumed that two variables should be totally linearly correlating with each other. If the true value is increased by a certain amount, then also the estimator of this value should show an increase of the same proportion. The amount of error in the estimator from this linearity is given by the coefficient of determination, denoted by r^2. This value can range from zero to one, where one means that the model fits perfectly the evaluated system, and a lower value means that the goodness of fit is not high or even non-existent. The coefficient of correlation, which is the square root of r^2, is given by the equation

$$r = \frac{cov(X,Y)}{s_X s_Y}, \tag{18}$$

where cov(X,Y) is the covariance of the data sets, and s_X and s_Y are the standard deviations. Computational engines can give these parameters for data sets with simple commands, but they can be calculated by hand from the original data, too.

3.3 Noise sources affecting the measurement result

3.3.1 General noise sources and their prevention

Noise problems have three recognizable components to them (Aumala, 2002); the source of the noise, the coupling of the noise and the device that the noise is affecting. Usually, it is best to begin eliminating the effect of the noise from as close to the source as possible. By switching off the disturbing component, one doesn’t need to worry about the latter parts of the problem. Sometimes this can be difficult or even impossible, though, because the function of the system that produces this error mechanism in question might be vital to the whole measurement system. For example, in our case of photoplethysmography, removing the sources of ambient light and movement artefacts while running outside is virtually impossible. In fact, they are a key factor in our realistic measurement problem, and removing these errors altogether would result in a trivial measurement case.

The coupling of the noise signal from outside sources can be many times prevented with careful planning. Solutions for installment and wiring can be simple to apply in practice, but may require thorough theoretical understanding of the fundamental physics behind the sensing system at hand. Coupling can be galvanic, inductive, capacitive, or happen straight through radiation. Again, it is tedious to try to make our sensors choose which kind of signals to pick up in our optical heart rate sensor. One option could be to use other wavelengths of light than the visible spectrum, for example, but this wouldn’t solve the entire problem since there are also numerous sources of infrared and ultraviolet radiation.

If the noise source cannot be cancelled out, nor can its coupling be prevented, one can still try to filter its effect. This is basically the idea behind these types of biomeasurements, where the desired signal has such low amplitude compared with other signal sources. The noise can be compensated by taking a separate reading of it and removing this from the output signal. For instance, in most wrist devices for health and fitness nowadays there are already accelerometers to measure movement, and the signal from these sensors can then be used to cancel out gross movement artefacts from optical signal pathways. If all of the earlier mentioned methods fail, one has to have a measurement device that can withstand these errors, and still give a reasonable reading even with maximal noise. This might only be achievable in cases where the wanted signal is clearly stronger than any outside signal, and can be seen as clear spikes from the background noise, for example.

3.3.2 Noise sources in PPG measurement

As mentioned in the earlier chapters, photoplethysmography suffers mainly from noise that is generated by ambient lighting conditions and movement artefacts. The coupling of electrical noise, if significant, is normally prevented by design and instrumentation choices. The human tissue is a complex, non-uniform medium and it gives rise to the more difficult signal noise problems. Ambient light plays an important role in noise cancellation as it is present in various forms basically anywhere. In theory, though, its effects can be removed quite easily. A surrogate measurement can be used to simply reduce the ambient intensity from the desired measurement signal, assuming that the light just adds to the signal linearly.

In practice this can be more difficult. If the intensity is far above the measurement range of the optical sensors, which might be the case when considering sunlight, saturation occurs, and one can’t measure the actual amplitude of the ambient light. Also, the non-linear coupling of noise sources can pose an issue if we assume that cancellation can just be done by linearly deducting the signals from each other. This is true in the case of movement artefacts, where alterations in orientation or position can affect the whole measurement system in many complex ways and distort the PPG signal fundamentally.

3.4 Reference

The absolute values of a recording don’t have a meaning performance-wise, unless there is something to compare them to. For this purpose, there must be a different device with which to make another measurement. This reference should be accurate, or as close to accurate as is desired, and usually has its performance already validated to be trustworthy. One should, of course, keep in mind the capability and accuracy of the reference device in question. When estimating the total error in the measurement, one must remember that when we compare our device to the reading of the reference device, the error between these two is not the actual error of the device compared to the true value. The measured error is made up of the true error and the error of the reference device itself, and only by having an accurate reference are we able to deduce values of the accuracy of our device compared to the truth.

Every measurement is basically just an educated and sophisticated guess of the true value, no matter how reliable the measuring device is. This is especially true in the case of physiological measurements, where the largest source of error and uncertainty is the measurement subject itself, the human physiological system. Through careful calibration, however, one can ascertain that the measurement device has close-to-optimal accuracy. If there exists an unbreakable chain of calibration from the device to the international primary norms, the measurement is said to have traceability. These primary norms are fundamental definitions for the basic units like the length of one meter or weight of one kilogram. From the collection of these basic units, all the other related quantities can be formed. National calibration laboratories, which use the international norms to calibrate their devices, calibrate devices for the industry further down the calibration chain. (Aumala, 2002)

In the case of heart rate measurement, the most widely used golden standard is the electrocardiogram. However, the usual ECG cannot be easily used in more consumer-oriented cases because of the nature of the measurement environment and conditions. For a more long-term measurement, for example with patients suffering a heart disease that is randomly encountered during normal daily routines, portable devices called Holter monitors can be utilized. They simulate the usual 12-lead ECG with similar electrode positioning, but the leg and arm electrodes are usually much closer to the heart for extra comfort in movement. There may be fewer electrode placements, and even as few as only two might be sufficient. This, however, is done with the disadvantage of less accuracy and parallelism. Normally these kind of portable devices also have a limited capacity for the storing of the data, and thus the resolution of the recordings is also worse.

One kind of a portable heart beat sensor is that of the renowned Finnish wellness company based in Jyväskylä, Firstbeat Oy (2014). Its Bodyguard 2 –device, seen in Figure 6 has been used by many major companies and sports clubs across the globe, like Nokia and Liverpool FC just to name a few. They offer unique insights for the companies' and clubs' leaders into the wellbeing and performance of their subordinates with a specific form of heart rate measurement. The device has two easily attachable electrodes, which can be worn non-stop through a normal day. There is just a short wire joining the electrode pads, which are located around the right clavicle and lower on the left side of the thorax. The other end of the wire has a sensor unit that records the data and also signifies the successful recording of the heart beat by a blinking light. This main unit can be connected to a USB-port on the computer to transfer the heart rate data. The device gives the readings of the RR-intervals with one milliseconds accuracy, which corresponds to the variations of just about a few tenths of beats in a minute in the normal heart rate zones. Parak and Korhonen (2014a) have evaluated the device to be able to detect the heart beats with accuracy of 99.98 % during different activities.

Figure 6. The Firstbeat Bodyguard 2 (Firstbeat Technologies Oy, 2014)

Another type of reference for heart rate measurement, and the one used primarily in the performance assessment of this thesis, is the traditional heart rate belt. The belt is placed around the thorax of the subject, and large electrode pads placed on the inside of the belt measure the potential differences on the skin across the chest. Companies like Polar (2014) have produced these belts for consumers since the 80s, and Polar's RS800CX unit is the one used in the protocol measurements of this thesis. We found in earlier test measurements that while Firstbeat's Bodyguard gives mostly accurate heart rate readings, it is still more designed for less active, normal everyday routines instead of just sports or exercise. For example, during running the main unit of the device and the connecting wire bounce quite a bit if they are not properly attached or taped to the skin, and this may lead to movement artefacts that are too disturbing for the device to get a proper reading. Polar's belts, however, are specifically designed for activities, and should stay better in place if attached as instructed. Studies like that of Terbizan et al. (2002) have shown that these belts also correlate well with actual ECG.

4 DEVICES FOR OPTICAL HEART RATE MEASUREMENT

The trend in health and fitness devices right now is to make activity trackers, whatever that activity is. The quantified self –movement has become the word to describe the modern human who cares about their body and aims to develop it by observing its habits and functions, and then trying to find actions for improvement based on the observations. In the case of weight control, it means photographing your meals and estimating their calorie counts, for the busy business person prone to burnouts it is time management tools and logging your feelings for some sort of a diary. For the enthusiastic runner, this type of data recording has been ongoing for years already with heart rate belts and their wrist counterparts. During the past few years different movement trackers or sophisticated pedometers have been introduced also to the wider audience, but for the keen athlete acceleration sensor data just isn't enough. The heart rate has to be there.

More and more companies are now introducing heart beat measurement to their devices thanks to photoplethysmography. It is easy to implement even to a tiny wrist device, because basically all that you need is one LED, which shines light, and one other component which detects this light. The rest of the functionality can be fitted to the existing computing unit, but is in practice somewhat trickier to do. The next few sections introduce different technologies and brands, both new and old, which utilize this novel measurement method.

4.1 Mio Global

Mio Global (2014) started at the beginning of the 21st century, when Liz Dickinson wanted to have an easy way of following her training plan and calorie intake after giving birth to her third child. She wanted to have a care-free heart rate monitor without a belt, and the technology for this she found with the Philips Electronics in Netherland. The Mio Continuous Technology behind the devices is the product of this partnership and it uses two green LEDs with an electro-optical cell in between them.

The first device, Mio Alpha, had a patented calorie management system with an accurate optical heart rate sensor. Its bulky screen shows the continuous heart rate or optionally a training timer, heart rate zone or time. The device can connect to different applications and devices with Bluetooth Smart. The manufacturer claims to have a "99 % EKG accuracy, even while running at speeds of up to 14.4 miles per hour" though this is probably the correlation of the devices' readings. Parak and Korhonen (2014b) have tested the Alpha in laboratory conditions, comparing it with an ECG device. They found the reliability score for the heart rate readings differing less than 10 % to be about 87.5 %, and the mean error was -1.21 bpm on average. Cycling proved to be the most erroneous part of the measurements with a mean error of nearly -4 bpm. The latest device, Mio Link, has a smaller frame with no screen and shows training zones with 5 differently colored blinking LED lights. The Link can also use the ANT+ (Dynastream Innovations Inc., 2014) connection technology, which is still pretty popular among heart rate devices, together with Bluetooth Smart. Both of these devices and the accompanying Mio Go –application can be seen in Figure 7.

Figure 7. The Mio Link (left) and the earlier Alpha (right) (Mio Global, 2014)

Recently, Mio Global has been working on collaborations to bring the wrist heart rate detection available to a wider audience. Both TomTom (2013) and Adidas (2014) have integrated the Mio technology into their latest smart watches. TomTom has the Runner Cardio and Multi-Sport Cardio, which have all the usual TomTom functionality like GPS, but also the optical sensors of Mio for heart rate measurement. Adidas has released the miCoach Smart Run, which also acts as a music player, and you can synchronize all your training data with integrated wireless LAN to the miCoach web page.

4.2 Scosche

Scosche Industries (2014), founded in 1980, has mainly focused on consumer car electronics with installation kits for different brands. They started on the consumer health market with the myTrek band, which is now discontinued. They have continued this product category with the Rhythm+ and Smart, which utilize the patented PerformTek technology by Valencell. This US company's technology is also used by, for example, LG in their heart rate earphones and it is validated with a 12-lead ECG measurement.

The Scosche Rhythm+ seen in Figure 8 uses both Bluetooth Smart and ANT+ like the Mio Link and it is waterproof with an IP67 rating. The backside of the device has three photoemitters with different wavelengths of light and detector unit in the middle. Unlike the Mio Link for example, the Rhythm+ has a larger strap and is worn on the forearm instead of the wrist. This, of course, makes the device bulkier, but it should be more accurate and less susceptible to movement and error caused by the superficial bones in the wrist.

Figure 8. The Scosche Rhythm+ (Scosche Industries, 2014)

Parak and Korhonen (2014b) have also tested the performance of the Scosche Rhythm against an ECG. The portion of readings having less than a 10 % difference compared with the ECG was found to be 86.26 % throughout various activities, and the mean error was only 1.11 bpm. The device performed better during cycling than the Mio Alpha, but had more errors in the running part of the evaluation protocol.

4.3 Samsung

In 2013, Samsung (2014) released their take on the smartwatch market. The Samsung Galaxy Gear was a first experiment in the wearable technology field by the Korean company, but it left hoping for much more. So in 2014, as they revealed the latest Samsung Galaxy S5, along came two new entries on the smartwatch front.

As the S5 had the new implementation of an optical heart rate sensor, so did the continuations of the Gear series; the direct descendant of the original Gear, the Gear 2, and the slightly more compact and fitness oriented Gear Fit. Both of these devices, seen in Figure 9, have the same optical sensor with green LED lights and acceleration sensors for step counting and tracking other activity. The drawback of these watches from the consumer point of view is that they can only be connected to the Galaxy Gear Manager application, which you can install on some of the latest Galaxy series phones and tablets which have Android 4.3 or higher.

The devices may give acceptable results for your step count, but when it comes to heart rate detection, the performance isn’t that favorable (Stein, 2014). First of all, the smartwatches can’t or even won’t measure your heart rate if they detect that you are moving, so they are obsolete as real time exercise partners. The manual of the devices instruct to take the heart rate measurement sitting down in a calm and quiet place. No talking or breathing deeply is allowed, or it may disrupt the measurement. If you manage to fulfill these criterions, the sensors record the heart rate over a certain period of time, and you get an average reading of the heart beat during that time. After you have managed to record this value on the device, you may still have to wait quite a while to get the reading to your phone for later inspection, or there might be a chance that this synchronization doesn’t happen at all.

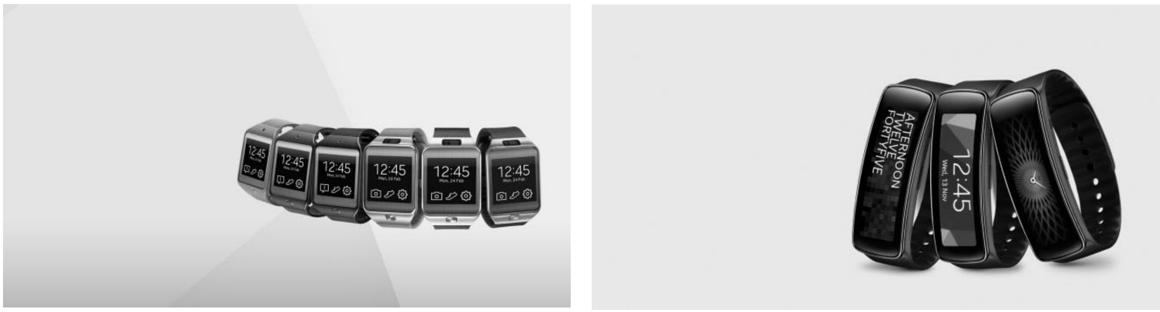


Figure 9. Promotional pictures of the Samsung Galaxy Gear 2 (left) and Gear fit (right) (Samsung, 2014)

4.4 PulseOn

PulseOn Oy (2014) is a Finnish spin-off from Nokia that started in the end of 2012. Based in Espoo, the company now has over ten employees, and they have secured funding of approximately three million euros, mostly from their lead investor Otar Margania, who is a banker and dean of the St. Petersburg State University. PulseOn had a pre-order campaign on the crowdfunding website Indiegogo, where anyone could purchase the device for 169 dollars to be shipped during September 2014.

PulseOn wanted to get rid of the usual heart rate chest belt with an easy solution for continuous heart rate monitoring. They felt that heart rate recording should be available to anyone, as it can give meaningful insights into training and recovery and therefore people’s wellness altogether. However, the normal graphs and figures produced by the common activity tracking applications don’t have much meaning to the average consumer. They just show the frequency of the beating heart as a number and that is it. PulseOn means to offer feedback for optimal exercising by combining their sensor solution with the heart rate variability analysis of Firstbeat Oy.

The device, seen in Figure 10, uses different wavelengths of light to get the best reading of the heart rate in every situation. This is also combined with intelligent signal processing algorithms to detect the true heart rate even when there is much interference from movement artefacts for example. The recorded exercise data can be sent to a mobile application after training via Bluetooth. From the application you can inspect more closely things like the intensity of the workout, calories burned, time to recovery or monthly progress of your fitness level. When you record a 20-minute moderate exercise with the phone’s GPS on, the application will calculate your maximal oxygen consumption, or VO2max, based on the speed and heart rate during the workout. The application can then tell you how effective each of your exercises are through the Training Effect number. This number goes from one to five, where the first half of the scale is for recovering or maintaining fitness exercises, and a number over three means the exercise is improving your fitness level.

Firstbeat’s analytics is used for the interpretation of the heart rate data from training intensity levels to fitness level estimation. The Training Effect and VO2max are based on their expertise on what your heart rate actually means for you individually.

Figure 10. The PulseOn wrist devices (left) and a screenshot of the mobile application (PulseOn Oy, 2014)

5 MEASUREMENT METHODS

The object of this thesis is to evaluate the performance of a novel optical heart rate sensor. This is done by conducting laboratory measurements on an adequate amount of test subjects wearing the device and a reference apparatus, to which the heart rate readings are compared to. The primary device which performance is assessed here is the PulseOn wrist heart rate monitor. The device is meant for easier heart rate monitoring for anyone, but the main focus group are young, active individuals, and the test subjects in these measurements represent this demographic quite well.

In the following chapter, the practicalities of the performance analysis measurements are presented. The settings and the actions during the measuring sessions were designed to be easily repeatable and tightly controlled, but still representative enough to truly estimate the performance in different conditions.

5.1 Measurement protocol

There were a few goals to achieve when designing the used measurement protocol. First, to mimic actual consumer user cases like going for a run, the protocol had to have exercising sessions with varying intensities and actions. A usual exercise starts with a slow warm-up and continues to increase speed gradually. After achieving maximum intensity for the exercise there is again a slower cool-down period before stopping completely. In addition to this pacing requirement, the warming up and build-up to higher speeds also ensures that the cardiovascular system of the subject has time to adjust to exercising without any danger of injuries caused by the sudden intervals of high intensity training.

Secondly, the protocol had to be carefully planned to be able to evaluate the reaction speed to changes in the system. When we know the exact time in the data when a certain activity starts or ends, we can try to assess the time it takes for the measurement device to respond to this change. Of course, this assessment has to be done keeping in mind that the subject’s heart rate also has some delay to adjusting to the changing load. If, for example, there is a short break in the activities when the subject can rest and their heart rate should decrease at least for a while, but we see no such drop in the measured heart rate, we can tell that the system has a too large time constant to be able to notice such a quick deviation.

Thirdly, there had to be different types of activities in the protocol. Because the device has built-in accelerometers to detect movement, it should be able to recognize when exercises with distinguishable limb movements are being performed. For example, the arms have quite differing movement patterns when comparing walking, running and cycling. During walking they swing in a larger arc and lower frequency than during running, and in cycling they don’t really move in relation to the torso, but there can be small, high frequency vibrations caused by the pedaling motion. This is also one reason why a wrist-worn device has advantages when compared with other body locations, but it has its flaws, too.

Before the actual protocol, the devices were set up, but also there was a three-minute waiting period after all the devices were turned on. During this period five squats were made at two-minute mark. This was partly to set the baseline at the resting heart rate, but also to time-synchronize some of the devices later. This was done from the acceleration signals, which present clear spikes when the squats are being done because of quick up-and-down movement.

The protocol itself, which is described in more detail in Table 1, starts in the sitting position because of the earlier mentioned baseline setting. After that there is a short warming up on an ergo bike, from which there is a transition to the treadmill after standing still for a minute. On the treadmill, the subject starts walking first at a tranquil pace. The inclination of the treadmill can be adjusted to simulate walking uphill and this was used to slowly increase the intensity of the exercise. The inclination was increased twice during both walking speeds, three and five kilometers per hour, but set to zero again after this. Then the speed was brought up to the running speeds of nine and eleven kilometers per hour. After the runs there was a cooling period of four minutes, sitting down. Finally, there was a few minutes on the ergo bike with an increasing cadence, and a final rest sitting again. The power setting on the bike was determined by the sex and activity classification of the subject. In this mode of the ergo bike, the resistance varies depending on the cadence, so that there is a constant power by decreased resistance when the cadence rises, and vice versa. For unconditioned individuals, that is, activity class lower than five, the load was 50 watts no matter whether the subject was male or female. For conditioned males with activity class 5 or more, the load was 100 watts, and for females it was 75 watts. Regardless of the power load, the subjects were told to try and keep their cycling cadence at 60 revolutions per minute in the first part and 90 in the second part. This increase in intensity was again meant to result in more work for the cardiovascular system, and an increased heart rate.

Table 1. Laboratory test protocol

Protocol task	Duration (minutes)	Starting time
Sitting at rest	1	0:00
Warming up on the ergo bike, 10 Nm resistance, approximately 75 rpm cadence	3	3:00
Standing still	1	4:00
Walking 3 km/h, 0 % inclination	3	7:00
Walking 3 km/h, 5 % inclination	3	10:00
Walking 3 km/h, 10 % inclination	3	13:00
Walking 5 km/h, 0 % inclination	3	16:00
Walking 5 km/h, 5 % inclination	3	19:00
Walking 5 km/h, 10 % inclination	3	22:00
Running 9 km/h, 0 % inclination	3	25:00
Running 11 km/h, 0 % inclination	3	28:00
Sitting at rest	4	31:00
Cycling with the ergo bike, 60 rpm cadence	3	35:00
Cycling with the ergo bike, 90 rpm cadence	3	38:00
Sitting at rest	4	42:00
Total duration		**46:00**

5.2 Measurement subjects

There were 20 subjects recorded in these heart rate sensor validation measurements. All of the volunteers gave an informed consent for their measurement data to be used in the evaluation, and anonymous information to be used for statistical purposes (Appendix 1). Information asked in the questionnaire included the date of birth, gender, height, weight, wrist circumference, dominant hand, skin color, smoking and health and activity level in general.

The skin color was assessed on the Fitzpatrick scale (Canadian Dermatology Association, 2014). This scale has six types of skin ranging from type I that is a pale white skin and usually has lightly colored hair, to a type VI, which is a highly pigmented dark brown or black skin. Most Finnish people fall under the type III or sometimes type II category. All participants were asked to confirm that they didn't have any illness or disease which would prevent them from doing a quick exercise required by the protocol, but furthermore their overall activity level was assessed. This was done on an activity scale from zero to ten, established by Firstbeat Technologies Oy (2014). They use this scale in the pre-questionnaires handed out to subjects before their own measurements. The activity scale is used only to record the exercise levels of the volunteers and have them comparable to each other. On this scale, a zero means no exercising at all, and a value of eight or higher usually means the individual is involved in some goal-oriented training on a near-daily basis. People who do regular exercise a few times a week are usually somewhere between four and seven on the scale.

Table 2 shows some of the statistics of the measurement subjects in detail. There were ten of both male and female participants in the study. The age span was from 22 to 46 years, with an average of about 28 years. The height and weight of the subjects was from 155 to 190 centimeters and 52-99 kilograms respectively, the averages being around 174 centimeters and 73 kilograms. These measurements result in a normal body-mass index, BMI, with only a few people having a value that signifies slight obesity. None of the volunteers were smokers, and all but one of them, who had skin type I, had the skin type of category II or III. All of the subjects were involved in moderate physical activity at least a few times a week. Some even exercised almost daily, having several hours of total weekly training time.

Table 2. Statistics of the measurement subjects

	Range (min – max)	mean ± standard deviation
Age [years]	24 (22-46)	27.6±5.67
Height [cm]	35 (155-190)	174±10.9
Weight [kg]	47 (52-99)	72.5±12.1
Activity level	4 (4-8)	6.3±1.3
BMI	10.90 (19.66-30.56)	23.77±2.70
Fitzpatrick skin type	I (n=1), II (12), III (7)	

Most of the measurement subjects were of a highly similar demographic; healthy, normal weight Caucasian people around the age of 25, with at least moderate amount of activity and no smoking. A few outliers to this group existed, but they are acceptable and even desirable. It is good to have tested the system with at least some variation in the measurement settings and background. Of course, it is much more important that the recorded subjects were of similar type, so that we can say something about the reproducibility and repeatability of the measurement protocol and the performance of the devices. The subjects also represent the target demographic of the device quite well. The product is designed for active individuals who are not necessarily involved in professional sports, but like to have knowledge of their activity and recovery continuously. However, as the device is meant for global markets, the skin type range of the Caucasian measurement subjects doesn't adequately cover all variations of the skin color the end-users of the product might have.

For a heart rate sensor evaluation, it is beneficial that there is a clear resting heart rate that is lower than the exercise heart rate, which can and should increase into significantly higher readings. Even maximum or nearly the maximal heart rate would be desired during the measurement to ascertain that the wrist monitor is capable of sensing and displaying these high figures. The same goes for the other end of the scale. There were a few participants whose resting heart rate was close to, or for one even under 40 beats per minute. The activity levels of the subjects are also of high importance. In a subject who is in a fit condition, the heart rate levels change more quickly according to the training intensity than with a subject with a poorer fitness level. This applies also to the decreasing of the heart rate, not only to increasing. For a person in bad shape, the heart rate may start rising rapidly right after higher effort, but it can stay high for a long time even after the exercise is finished and the subject is resting. When the person is more fit because of regular exercise, the recovery of the cardiovascular system happens more instantly when the load to the body is ceased.

5.3 Devices

All four of the devices and their correct placement can be seen in Figure 11. The main device under evaluation in this thesis, and used in the protocol measurements, was the PulseOn heart rate wrist monitor. The device was placed on the non-dominant hand of the measurement subject determined in the pre-questionnaire according to the manufacturer's instructions, a few centimeters above the wrist bone, the measuring unit facing the back side of the hand and tight enough not to let any ambient light shine under the unit to the sensors from the sides. The device was sending its raw data continuously to the PulseOn application open on a Samsung Galaxy S4 Active. The data of the Mio Link, which was worn on the other wrist, was not used in this thesis.

For double-checking and backup, two different measurement devices were used to record reference heart rate data. First one was Firstbeat's Bodyguard 2. The sensor isn't particularly designed to be used solely in exercise activities, and so it can show erroneous readings during high speed running, for example. This is especially true if the subject sweats excessively during the exercise and the electrodes start to lose their hold on the skin, or the wire joining the electrodes is not closely attached to the skin with tape and is left to hang loosely in the rhythm of the movement. However, when these things are taken into account and cared for properly, and the electrodes are placed carefully and tightly, the Bodyguard offers an adequate backup reading should the other device fail momentarily. Two electrodes were placed on cleaned areas of the skin, one under the right clavicle and the other on the left side of the chest, approximately at the level of the lowest rib bone as can be seen in Figure 11. Skin tape was used to hold the adjoining wire in place during movement. The main unit stores the RR-interval times to be transferred to a computer in an easily-read format later.

The other, main reference device used was the Polar RS800CX. This set includes the usual soft heart rate belt where the main sensor unit Polar H3 can be attached to. The belt has large sensor pads on both sides of the chest which have to be adequately moistened with water before the recording to ensure a good connection. The proper tightness of the belt is also crucial so that it doesn't begin to slide up or down during the exercise, but still doesn't feel uncomfortable for the wearer. The sensor transfers the heart rate data wirelessly as RR-intervals to the wrist unit. The PC application Polar ProTrainer 5 comes with the package and can be used to store and visualize the heart rate recordings later. The wrist unit can send the data to the computer through an infrared dongle that is attached to an USB port. This data can be exported from the program as Polar's own HRM format that has, on top of many other bits of information, also the RR-intervals of the heartbeat.

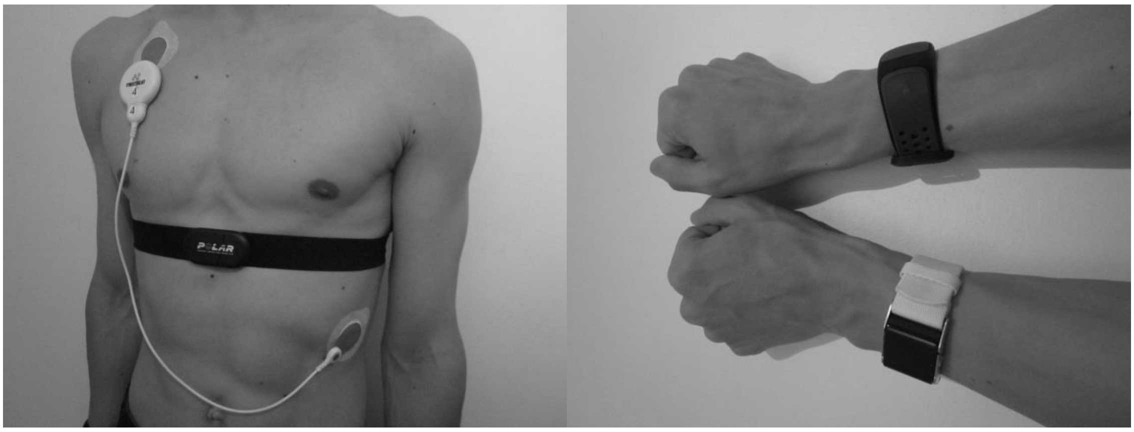


Figure 11. Subject wearing the measurement devices. Polar belt and Firstbeat Bodyguard 2 on the left, Mio Link (above) and PulseOn on the right

6 RESULTS AND DISCUSSION

We have gone through the methods and principles of physics of how to measure heart rate with an optical sensor. Finally, to evaluate the performance of such a sensor we need to assess the quantitative data it provides and produce it into figures of measurement quality.

In this chapter, first the pre-processing actions made on the raw data from the devices are presented. Then we will go through a few reliability values from the data sets, which tell roughly how useful the device would be in practical consumer use. Finally different statistical methods with equations will be used to calculate the accuracy of the sensor inspecting it from various aspects.

6.1 Data pre-processing

From the Bodyguard 2, the data was loaded using Firstbeat's own programs. First, the Firstbeat Uploader was used to getting a raw SDF format file from the device. This file has a header with information about the starting timestamp of the recording, and a field with the RR intervals listed in a column. This information was parsed in another Firstbeat software, the SPORTS. This program also has the functionality to have the data corrected by removing obvious outliers or missing data points. The heart beat information was saved again in CSV format as the RR intervals and their corresponding timestamps in milliseconds, calculated with a cumulative sum.

The data from the Polar belt and the accompanying wrist device was first sent to the Polar ProTrainer 5 software. This data could then be exported as Polar's HRM format. This is a slightly inconvenient form of data for analysis purposes, because it changes its formation depending on the device the data is recorded with, and its settings. For these measurements, however, the settings were kept the same throughout the subjects, so it was easier to parse the raw data from the files. Polar also had the option of storing RR interval data and the information from the HRM-file was parsed in a similar fashion to the Firstbeat SDF-files.

The PulseOn raw data was sent from the device to a mobile phone via Bluetooth. In normal customer use, only the heart rate would be recorded in certain intervals. However, the application and device used were development versions, which allowed for all of the actual raw data to be recorded. This meant the application saved a CSV in the memory of the phone, which had more information than just the heart rate that is the output of the detection algorithm. In the raw format there are records of several time instances during one second. The accompanying timestamps in milliseconds are saved with signal values from the acceleration sensors and intensity values from different optical channels. From this data, the actual heart rate is calculated with a sophisticated algorithm which takes into account the movement detected from the acceleration signals, which might distort the optical signal.

To make the readings from the Polar heart rate belt and the PulseOn device comparable, we first had to manipulate the data. The raw optical signal information from PulseOn was fed to the HR detection algorithm with the acceleration data in MATLAB (MathWorks Inc., 2014). This algorithm was an offline version of the one working inside the device online. It had been tuned from the time of the actual protocol measurements to its current state, which is also used in the device version available to the consumers. The raw data that the optical sensors picked up during the laboratory measurements would be the exact same with the current algorithm, only the results interpreted by the device differ from that which it would have shown earlier.

When the heart rate data from both devices was available, it still had to be synchronized in time and made so that the individual measurement points could be compared. To time-synchronize the data cross correlation was used. Because two data files were created with differing sampling frequencies, they had to be interpolated to the same sampling frequency of one sample every millisecond. Also, because the reference recording was started before and ended after the PulseOn measurement, its size had to be cut from both ends before additional operations. After this, the cross correlation could be conducted. The function in MATLAB gave the correlation values and their corresponding time lags. By finding the maximum value of these correlations the exact time shift of the data of the devices could be determined. Then the synchronizing meant to just change the timestamps of the other recording accordingly.

After the synchronization the original data points were on identical time scales, but still with different sampling rates and data sizes. To make the readings finally comparable, the average heart rate value over a five second time window was taken to represent the device over that time. This averaging was done starting from the first time instant after the warm-up ergo-cycling, four minutes into the PulseOn data and the last window was taken from the end of the PulseOn data. These last windows had to be of similar size for both Polar and PulseOn in case the latter was shorter than five seconds. Now the individual data points could be compared respectively, each representing a corresponding section of the signals.

6.2 Reliability

To calculate the reliability of our measurements and therefore get one measure of their performance, we will use a few different kinds of criterions. As mentioned in the earlier theory section, the reliability of this type of a heart rate recording can be assessed by looking at the ratio of measurement points which are close to the real value in a certain degree to those points which are not. The definition of when a measurement is close enough to the true value can be agreed on, and the value range can be made either absolute or relative. A constant range in the units of heart beats can be more forgiving or tight depending where we are in the spectrum, but it is also easier to calculate. A relative range of a certain percentile adapts to the heart rate zone the measurement is at, giving more space for error when the heart rate increases and vice versa.

In this chapter we will use both a constant range, and a relative range. The constant ranges used are from five beats below the true value to five beats above it, and another range is same with ten beats as the limits. These reliabilities can be expressed in equation form as

$$reliability_{5/10bpm} = \frac{N_{<5/10}}{N_{all}} * 100\ \%, \tag{18}$$

where $N_{<5/10}$ is the number of data point pairs where the error between the PulseOn device and the reference is less than 5 or 10 beats per minute, and N_{all} is the total number of data pairs. A difference of five beats per minute is not much in this kind of a measurement, and getting these reliability readings to over 90 percent would be excellent for the evaluated device. This also requires extreme accuracy from the reference device, as the threshold can be passed by having errors of just a few beats in the opposite directions in both devices. Ten beat difference is still quite strict to be applied throughout the measurement, but errors larger than this tend to be highly critical for the purpose of accurate heart rate measurement, and they shouldn't happen regularly during stable conditions. In Figure 12 is an example of one measurement subject's data, where curves for both devices can be seen, and also points where the error is more than 10 bpm are highlighted. Similar figures for each subject can be seen in Appendix 2.

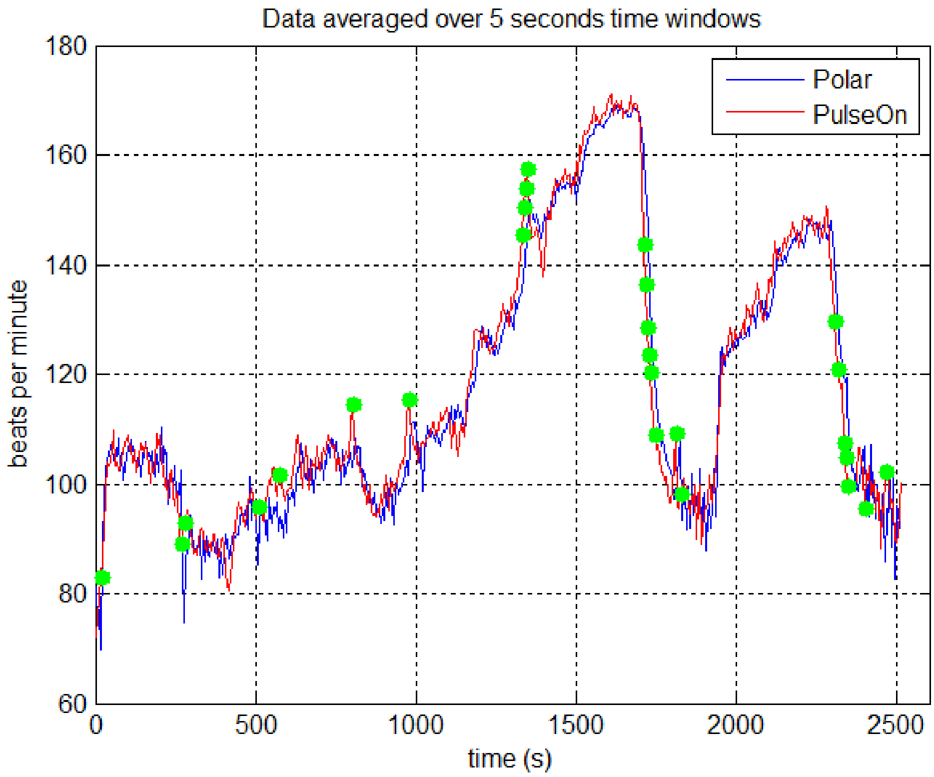


Figure 12. Example of interpolated data. Sections where error is greater than 10 bpm highlighted in green.

The relative ranges used are somewhat similar to the constant ones, but probably somewhat looser in general. These reliability ranges are five and ten percent of the reference reading, up or down. The form of calculation used is similar to equation (18), but in this case the numerator is the number of data pairs having a relative error of less than 5 or 10 percent compared to the reference value. For most people in these protocol measurements, the heart rate values during anything but sitting were at least close to one hundred, and during exercise much higher, of course. This means the ranges in percentage are slightly broader than in absolute heart beats during the majority of the measurements. Also, the parts of exercising with higher heart rates tend to be the more tedious section of the recordings for the sensors since there is usually increased movement involved, which means more interference. So, it is naturally logical to have a narrower range of allowed error in the lower heart rate zones and a wider range during the increased heart rate. Even though this might mean error ranges smaller than five beats during sitting, for example, the relative percentile error ranges are still generally more forgiving than constant ranges, and it is also more logical to use them.

6.3 Measurement errors

In the previous chapter, we went through measures of reliability, which take the difference of the measured and real value, and deduct a simple true or false answer to that data pair. The number of these true or false answers was then calculated as a percentage of all the pairs. In this chapter, we will take a similar kind of an approach, but instead of a binary value of yes or no, we get a real number value from the difference of the measurement. When we aggregate these values, we will get statistical figures that give us quantitative data about the recording more precisely. We will use some of the indicators of data discussed in chapter three for calculation. These different calculation methods describe the data with differing emphasis. We will also use a few other methods that quantify the data, which are more commonly used in data modeling and in making forecasts or simulations of these models.

The principle of linear regression can be used in our heart rate performance evaluation. We are assuming that the real heart rate is accurately measured by our reference device. Then we will try to model this heart rate with another measurement, the optical sensor. This model of the heart rate should, of course, be equal to the true value and change linearly whenever the true value changes. So, we can say that our optical measurement of the heart rate should be linearly correlating with the reading of the reference device. We can evaluate how correct this assumption is with regression analysis. We can see the correlation of all the data points of every subject in Figure 13 in terms of linear regression. The correlation coefficient of 0.96 is usually squared and then expresses the coefficient of determination r^2.

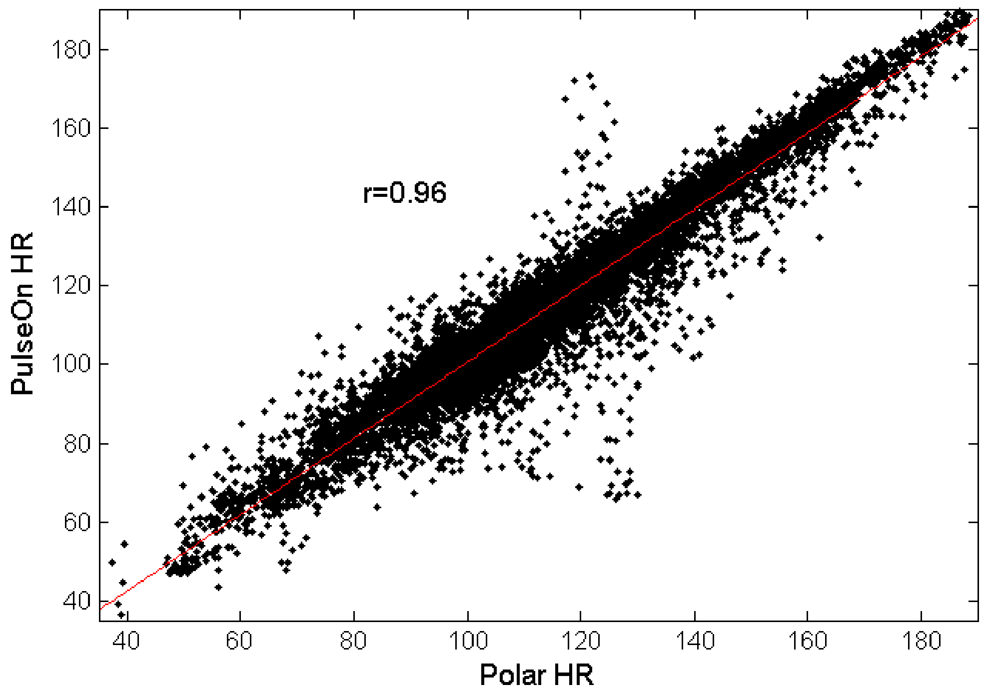


Figure 13. Linear regression fit to the entire data set.

In Figure 14 we can see the difference plot, or the Bland-Altman plot (Bland & Altman, 1995) of the whole data set. The plot shows the difference in each of the data pairs against the average of those two points.

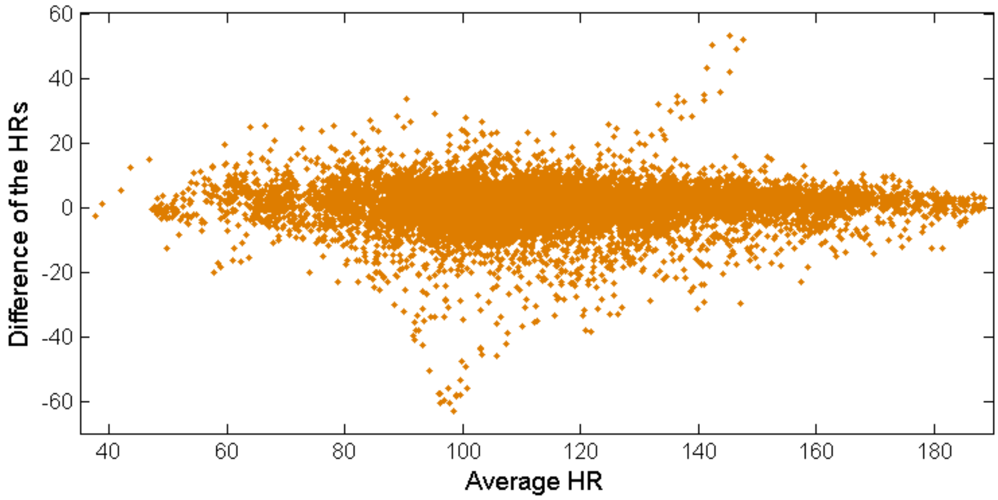


Figure 14. The Bland-Altman plot of the entire data (Bland & Altman, 1995)

To visualize the distribution of our data, basic histograms are also a simple yet powerful method to utilize. It can be used to distinguish the differences in frequency distributions between certain error ranges more clearly, and show if there are any underlying statistical distributions of the error that could be identified. For example, if the data is skewed in one direction it can be seen quite clearly from a histogram with evenly distribute bins, and also the amount of outliers in data points is more obvious. Also, histograms can just visualize the classification of data. For example, in our case we can differentiate how much of the data was recorded during the lower, resting heart rates of under 60 beats per minute compared to some higher heart rate zones during exercising in the 140 to 160 beats per minute area. This distribution tells us of the significance of other measures, too. An excellent performance figure might be less important if we are only working in the lower, easier-to-detect heart rates.

In the next Figures 15 and 16 we can see one example of heart rate and error distributions in histograms. The first figure shows how the heart rate readings are spread out from lower values to higher ones, the majority being near the 100s. In this case over half of all the readings are also more than 100, which means that the relative error range limits expressed in percentages would be more than those given in the same absolute values. For example, five percent error would be more than a five beats per minute error. From the second figure we can see that a clear majority of the errors are within the narrowest range of plus-minus five percent.

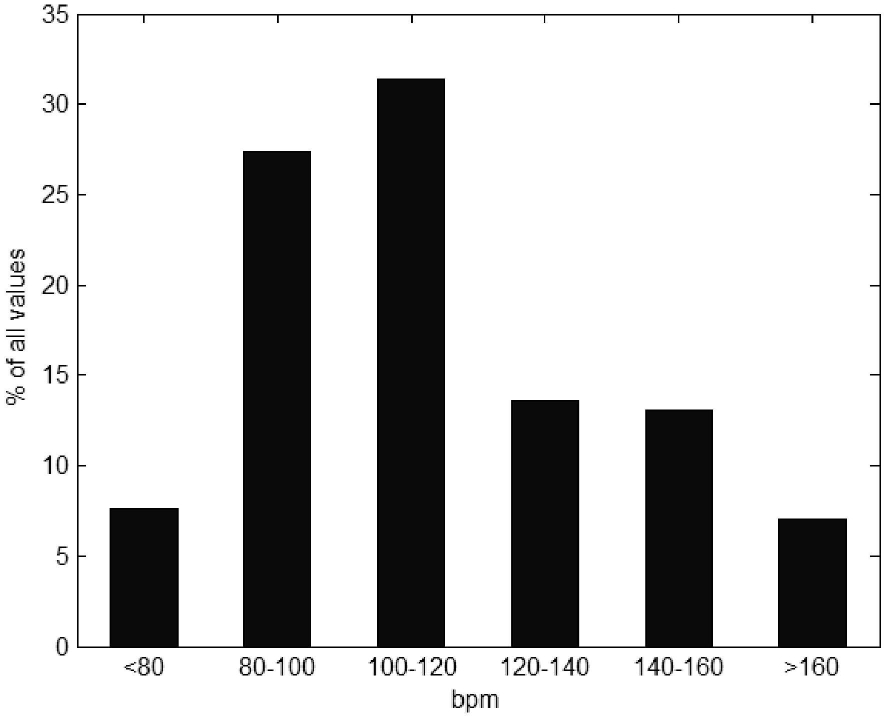


Figure 15. The distribution of a subject's measurement readings

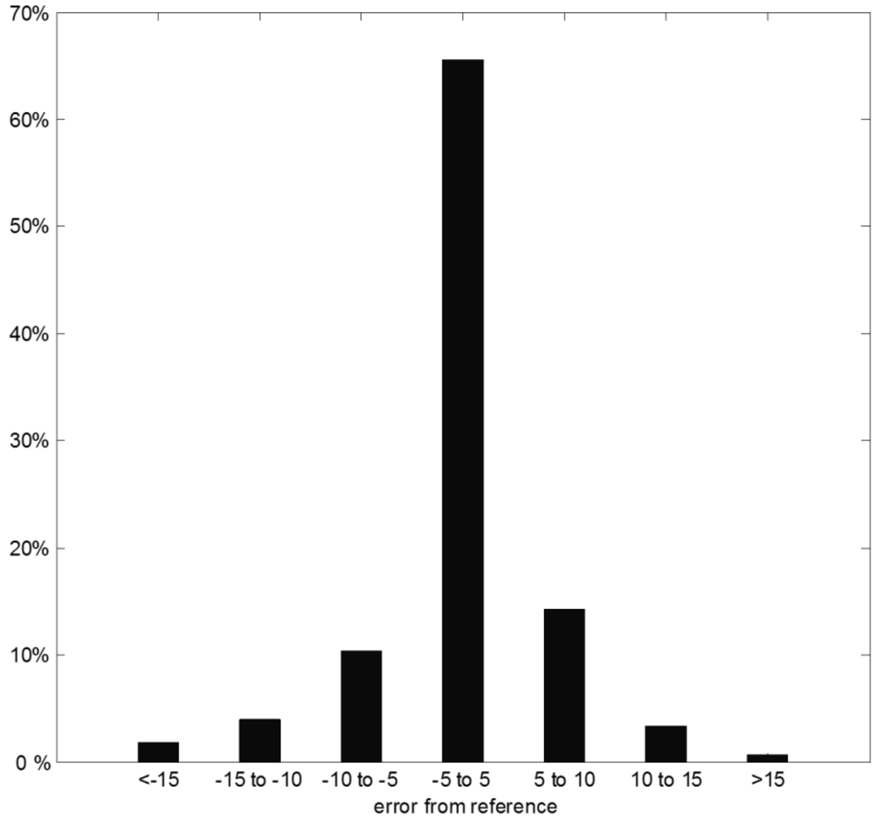


Figure 16. The distribution of errors in bpm

In Table 3 below we can see different results for all the measurement subjects recorded for the study in this thesis. Their ages and skin types are shown in columns two and three. Values are given for reliability of the errors being smaller than 10 % of the reference reading, and also the mean absolute error, the normalized root-mean-square deviation and standard error of estimate as expressed by equations (7), (13) and (17) respectively. Also, the coefficient of determination is given in the last column.

Table 3. Measurement results for all the subjects

Subject ID	Age	Skin type	reliability <10% [%]	MAE [bpm]	NRMSD [%]	SEE [bpm]	R^2
1	25	III	95.83	3.73	5.11	5.08	0.956
2	26	II	97.42	3.27	4.30	4.31	0.958
3	26	III	88.10	4.78	7.44	6.91	0.894
4	27	III	81.75	4.54	7.12	6.53	0.942
5	28	II	90.67	5.42	6.97	8.26	0.898
6	24	III	89.29	4.76	5.45	6.04	0.959
7	39	II	90.28	4.34	5.61	5.60	0.923
8	22	II	94.84	4.35	5.51	5.72	0.948
9	46	II	82.94	4.29	6.31	6.13	0.923
10	27	III	98.41	3.53	4.88	4.50	0.949
11	26	I	89.29	5.76	13.58	10.92	0.648
12	29	II	97.22	3.88	5.00	5.03	0.955
13	23	II	85.91	5.53	8.81	7.62	0.888
14	23	II	84.13	5.81	10.80	8.10	0.794
15	26	III	99.40	3.21	4.36	4.28	0.959
16	26	II	84.33	6.68	13.02	10.29	0.799
17	26	III	83.73	6.37	10.69	10.19	0.819
18	25	II	93.25	4.56	5.36	6.17	0.954
19	22	II	90.87	5.61	10.11	9.25	0.810
20	26	II	90.28	4.78	8.14	6.30	0.870
Average			**90.40**	**4.76**	**4.73**	**7.14**	**0.924**

As we can see, the average values are around 90 percent for reliability, a little less than five bpm or percent for the mean absolute error and normalized root-mean-square deviation, a little over seven bpm for the standard error of estimate and about 92 % for the coefficient of determination. The range of reliability is quite wide; from around 81 % up to a 99.4 %. For R^2, the range would be slightly narrower if we don't include subject number 11's result of 0.648, but the highest figures are less than those of reliability. The values of the different errors more than double in some cases from subject to subject, again subject 11 showing up to 13.58 % for the NRMSD when compared to number 15's 4.36 %, for example.

6.4 Discussion and comparison

A few results from other publications are now given for comparison with our values. Terbizan et al. (2002) assessed the validity of seven commercial heart rate belts in their study. These belts use the traditional measurement of electric potential on the skin surface as does our reference device, so they can't be equally compared with our PPG device, but these performance figures give some idea of the ballpark we are talking about. Terbizan et al. had 14 male subjects, who they measured in four different conditions for a duration of 10 seconds; once during rest, and during three speeds on a treadmill. Their criterion for a valid measurement reading was that the correlation should be over 90 %, and the standard error of estimate (SEE) less than 5 bpm. None of the devices gave a valid reading according to these restrictions when it came to the fastest speed on the treadmill, which was about 9.6 km/h. Four of the devices filled the criterions on the slower speeds and rest, and one of the devices also was valid during the first two speeds, but failed during the rest period. Two of the measurement devices failed to achieve the limits completely in any situation. The two Polar devices they evaluated were among the top performers, though the other one gave the worst results during the fastest running.

As mentioned earlier in chapter 4, Parak and Korhonen (2014b) evaluated the performance of Mio Alpha and Scosche Rhythm against an ECG reference with a similar measurement protocol. Their results showed that the devices had 10 % reliabilities of 87.49 % and 86.26 %, and mean absolute errors of 4.43 bpm and 6.82 bpm, respectively. Comparing these with the results in this thesis we can see that the Mio Alpha had a slightly smaller error but also a smaller reliability and the Scosche device performed worse in both aspects in comparison with the PulseOn device.

In his white paper, Eschbach (n.d.) assesses the validation and reliability of the Valencell PerformTek® optical heart rate technology. This sensor was integrated into an earbud device and its results were compared with a 12-lead ECG. Doctor Meir Magal from the American College of Sports Medicine conducted the data collection from the subjects. They had 41 subjects with an almost equal distribution of men and women, and 36 of these subjects returned later for a second measurement to evaluate the test-retest reliability of the device. The measurement's duration was 15 and a half minutes, and it involved first sitting in a chair, then standing and walking and slowly increasing speed to a maximum of 6 miles per hour, and then decreasing the speed in the same fashion. These speeds were similar to the ones used in the measurements for this thesis, although having a slightly different profile, and for nine subjects the speed was decreased even further.

Eschbach presents a few metrics for their data, namely correlation, the standard error of the estimate, confidence intervals and five beats per minute reliability range ratios. He also notes that the test-retest reliability of 36 follow-up measurements was close to that of the reference ECG's value. The correlation coefficient between the device and the reference was found to be 0.99, with a standard error of estimate of 4.43 beats per minute. The 95 % confidence interval limits were at -9.55 and 8.18 bpm, with a bias of -0.68 bpm. The amount of readings that were inside five units of the ECG's values was 87 %.

These results are clearly better than those obtained for the data collected for this thesis. The 5 bpm reliability is almost as good as our 10 % reliability, which translates to more than five beats during the majority of the measurements. Eschbach's data shouldn't be directly compared with ours, though. The optical heart rate device inserted into the ear is obviously a totally different configuration as opposed to our wrist device. The sensor is probably more protected from outside light sources because it sits deep in the ear, and also the blood vessels inside the ear cavity walls are much more clearly expressed. Furthermore, wrist is more prone to movement artefacts than the ear.

Another reason for the difference, besides the obvious dissimilarities in the measuring technology, can be in the measurement device to which the readings are compared. The reference ECG of Eschbach's measurements is the golden standard of heart rate measurement and is surely much more accurate than our commercially available heart rate belt. As discussed in the theory section, the total error we get as a result from our calculations is comprised of the error between the measurement device and the reference and also the error of the reference device itself. We have chosen the Polar belt to be the reading to turn to when comparing our values, but we must keep in mind all the time that this reference doesn't tell the exact truth, even the ECG cannot perfectly.

Lu et al. (2009) have also studied a PPG device attached to the earlobe, comparing it to an ECG. Even though they had a highly limited measurement setting and a short duration of the recording, their results showed significant correlation between the two devices. The median correlation was 0.91, ranging from as low as 0.14 up to 0.99.

The measurements only lasted a few minutes, and all the time the subjects were sitting and asked to keep as still as possible. Even in this setup the results varied immensely, which tells clearly about the difficult nature of the photoplethysmography measurement. The signals received by the optical sensors are so delicate and prone to even the slightest errors that coming up with a device that could accurately measure the heart rate in each and every situation encountered is truly a great feat. The PulseOn device and the software working inside of it during the laboratory measurements were still in the development state, although the raw signals recorded were gotten with a similar instrumentation as the current device. Also, the algorithm used to get the heart rate readings from this data was the latest one, and already shows significant improvement from earlier tests. Still, the work for developing the device and its heart rate detection algorithm is ongoing and will never be perfect. The ideal device would be able to record the heart rate from any individual in every condition, which is an impossible task considering that each measurement subject, human being, is a unique black box with many unknown parameters that should be known completely to get the perfect reading every time.

7 SUMMARY

In this thesis, we have gone through some basic principles of photoplethysmography, and how it can be applied to consumer products that can be easily worn on the wrist to give a heart rate reading that is adequately accurate to the needs of an everyday athlete. We started from the traditional methods of measuring the heart rate from electric potential, and how these principles have been used in the more widely used heart rate belts of companies like Polar. However, these belts have their shortcomings when it comes to comfort or ease of use although giving the most reliable readings of commercially available fitness devices.

Photoplethysmography is a technology that has been used in clinical settings for years already in blood oxygen saturation measurements. The PPG signal changes according to the constituents and volume of blood through which the measuring light is shining. The information of this waveform can be used to assess many aspects about the physiological system, but one useful application is to detect the heart rate. This can be achieved because the blood pulsates in the vessels in the same rhythm as the heart pumps it out to the cardiovascular system. The frequency of this pulsation can quite effortlessly be detect with simple instruments. However, even though something is simple, it doesn’t necessarily mean that it is easy, too. This has been sadly true with many attempts to try and commercialize the technology, as companies try to give the common people the tools for their need of self-quantification.

The Finnish company PulseOn Oy has made their take on the optical heart rate measurement industry, and the performance of the device was the main aspect under evaluation in this thesis. Controlled laboratory measurements were conducted with differing activity tasks on 20 volunteers and the resulting data from these activities was compared to a Polar heart rate belt acting as the reference device. The resulting errors between the two devices were analyzed in MATLAB to arrive at a few quantified figures of performance. These results showed that the PulseOn device had a good accuracy in being in the same heart rate zones as the reference, and the calculated errors were decent enough to have the device show correct readings most of the time during normal use. In these assessments, one has to remember that the reference device isn’t perfect, either, and also has some degree of error in its own measurement.

Optical heart rate measurement with PPG is highly susceptible to noises from movement artefacts and ambient lights. What seems like an easy setup with a light emitter and a photodetector, is in fact far from simple when accurate heart rate figures are desired, especially with a target of the measurement as complicated as the human physiological system. PulseOn has done years’ worth of research into bringing heart rate monitoring available for everyone, and their device, although constantly under improvement heart rate detection-wise, is already a tempting alternative for the people of the quantified self –movement to lose their heart rate belts and wear the sensor on their wrist.

REFERENCES

Adidas International Trading BV. 2014. adidas miCoach: The Interactive Personal Coaching and Training System, website. Available (accessed on 4.11.2014): http://micoach.adidas.com/smartrun/

Allen, J. 2007. Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement 28, 3, pp. R1-R39.

Aumala, O. 2002. Mittaustekniikan perusteet. 11. edition. Otatieto, Helsinki, 223 p.

Bland, J.M., Altman, D.G. 1995. Comparing methods of measurement: why plotting difference against standard method is misleading. The Lancet, 346.8982, pp. 1085-1087.

Canadian Dermatology Association. 2014. Know Your Skin Type, website. Available (accessed on 4.11.2014): http://www.dermatology.ca/skin-hair-nails/skin/photoaging/know-your-skin-type/

Dynastream Innovations Inc. 2014. THIS IS ANT – What is ANT+, website. Available (accessed on 12.11.2014): http://www.thisisant.com/consumer/ant-101/what-is-ant

Eccles Health Sciences Library. 2014. ECG Learning Center, website. Available (accessed on 4.11.2014): http://ecg.utah.edu/lesson/1

Eschbach, C. n.d. Validation and Reliability of PerformTek® Earbud Heart Rate Sensor Utilizing 12 Lead ECG. Valencell Inc. Available (accessed on 19.8.2014): http://www.valencell.com/white-papers

Fallow, B., Tarumi, T. & Tanaka, H. 2013. Influence of skin type and wavelength on light wave reflectance. Journal of clinical monitoring and computing 27, 3, pp. 313-317.

Firstbeat Technologies Oy. 2014. Analyzed by Firstbeat, website. Available (accessed on 4.11.2014): http://www.firstbeat.com/consumers/firstbeat-intelligence-in-heart-rate-monitors

Foo, J.Y.A. & Wilson, S.J. 2006. A computational system to optimise noise rejection in photoplethysmography signals during motion or poor perfusion states. Medical and Biological Engineering and Computing 44, 1-2, pp. 140-145.

Hayes, M.J. & Smith, P.R. 1998. Artifact reduction in photoplethysmography. Applied Optics 37, 31, pp. 7437-7446.

Hillman Lab. 2012. Skin Imaging, website. Available (accessed on 4.11.2014): http://orion.bme.columbia.edu/~hillman/Skin_Imaging.html

Huang, F., Yuan, P., Lin, K., Chang, H. & Tsai, C. 2011. Analysis of reflectance photoplethysmograph sensors. world academy of science, engineering and technology 59, pp. 1266-1269.

Lane, D.M. 2014. Online Statistics Education: A Multimedia Course of Study, website. Public Domain. Available (accessed on 4.11.2014): http://onlinestatbook.com/

LearntheHeart.com. 2014. QRS Complex, website. Available (accessed on 4.11.2014): http://www.learntheheart.com/ecg-review/ecg-interpretation-tutorial/qrs-complex/

Lu, G., Yang, F., Taylor, J.A. & Stein, J.F. 2009. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. Journal of Medical Engineering and Technology 33, 8, pp. 634-641.

Maeda, Y., Sekine, M., Tamura, T., Moriya, A., Suzuki, T. & Kameyama, K. 2008. Comparison of reflected green light and infrared photoplethysmography. Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE. pp. 2270-2272.

Spl4. 2006. File: PVC detectionUsing PGG.png. GNU Free Documentation Licences / Creative Commons, website. Available (accessed on 4.11.2014): http://en.wikipedia.org/wiki/File:PVC_detectionUsing_PGG.png

Mathworks Inc. 2014. MATLAB – The Language of Technical Computing, website. Available (accessed on 4.11.2014): http://www.mathworks.se/products/matlab/

McGraw-Hill Global Education Holdings. 2014. Home – AccessScience from McGraw-Hill Education, website. Available (accessed on 4.11.2014): http://www.accessscience.com/

Mio Global. 2014. Strapless Heart Rate Monitors, website. Available (accessed on 4.11.2014): http://www.mioglobal.com/

Morris, A.S., Langari, R. 2012. Measurement Uncertainty. Measurement and Instrumentation, Butterworth-Heinemann, Boston, 2012, pp. 39-102.

Quantified Self Labs. 2014. Quantified Self | Self Knowledge Through Numbers, website. Available (accessed on 3.11.2014): http://quantifiedself.com/

Parak, J., Korhonen, I. 2014. Accuracy of Firstbeat Bodyguard 2 beat-to-beat heart rate monitor, website. Available (accessed on 4.11.2014): http://www.firstbeat.com/userData/firstbeat/tiedostolataukset/white_paper_bodyguard2_final.pdf

Parak, J., Korhonen, I. 2014. Evaluation of wearable consumer heart rate monitors based on photopletysmography. 36th Annual International IEEE EMBS Conference 2014. 4 p.

Polar Electro. 2014. Who we are, website. Available (accessed on 3.11.2014): http:// http://www.polar.com/uk-en/about_polar/who_we_are

PulseOn Oy. 2014. PULSEON WRIST DEVICE & APPLICATION, website. Available (accessed on 3.11.2014): http://www.pulseon.com/pulseon-wrist-device-application

Rama. 2005. File:Saturometre_2.jpg. CeCILL, website. Available (accessed on 4.11.2014): http://commons.wikimedia.org/wiki/File:Saturometre_2.jpg

Samsung. 2014. Wearables / Smartwatches, website. Available (accessed on 3.11.2014): http://www.samsung.com/uk/consumer/mobile-devices/wearables/

Stein, S. 2014. Samsung Gear Fit review – CNET, website. Available (accessed on 4.11.2014): http://www.cnet.com/products/samsung-gear-fit/

Suunto. 2014. About Suunto, website. Available (accessed on 3.11.2014): http://www.suunto.com/en-GB/About-Suunto/

Tamura, T., Maeda, Y., Sekine, M. & Yoshida, M. 2014. Wearable Photoplethysmographic Sensors—Past and Present. Electronics 3, 2, pp. 282-302.

Terbizan, D.J., Dolezal, B.A., Albano, C. 2002. Validity of Seven Commercially Available Heart Rate Monitors. Measurement in Physical Education & Exercise Science, vol. 6, issue 4, pp. 243-247.

TomTom International BV. 2013. Multi-Sport Cardio, website. Available (accessed on 4.11.2014): http://www.tomtom.com/en_gb/products/your-sports/tomtom-multi-sport-cardio-gps-watch/white/

Vehkaoja, A., Salo, A., Peltokangas, M., Verho, J., Salpavaara, T., Lekkala, J. 2014. Unconstrained Night-Time Heart Rate Monitoring with Capacitive Electrodes. XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013 41, IFMBE Proceedings. Springer International Publishing, pp. 1511-1514.

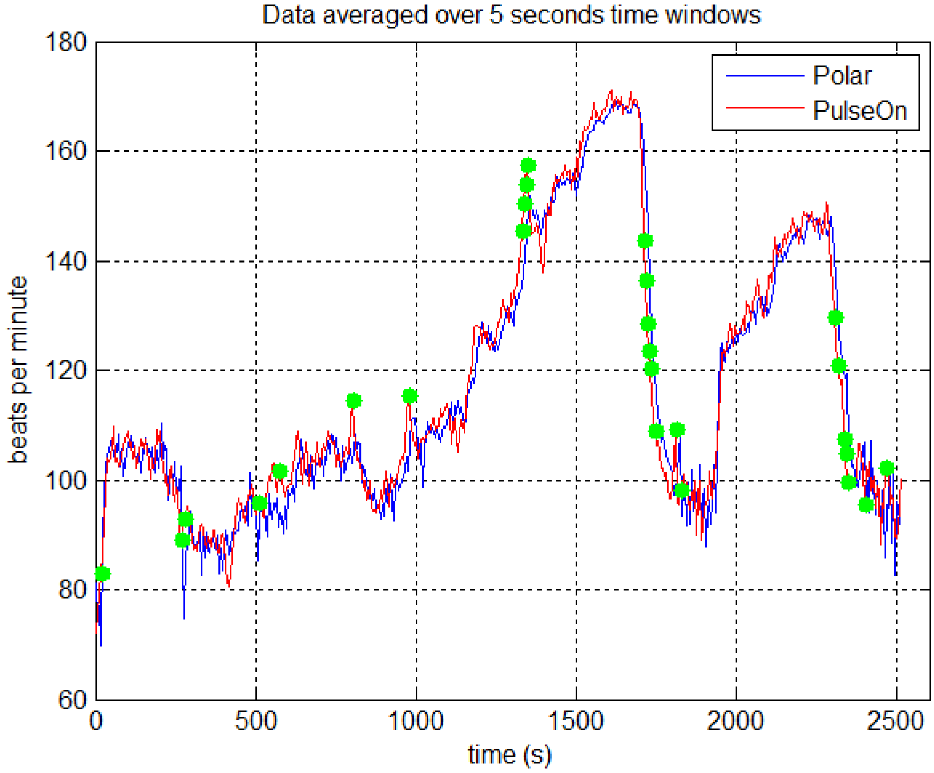
Weijia Cui, Ostrander, L.E. & Lee, B.Y. 1990. In vivo reflectance of blood and tissue as a function of light wavelength. Biomedical Engineering, IEEE Transactions on 37, 6, pp. 632-639.

Appendix 1: Subject information questionnaire

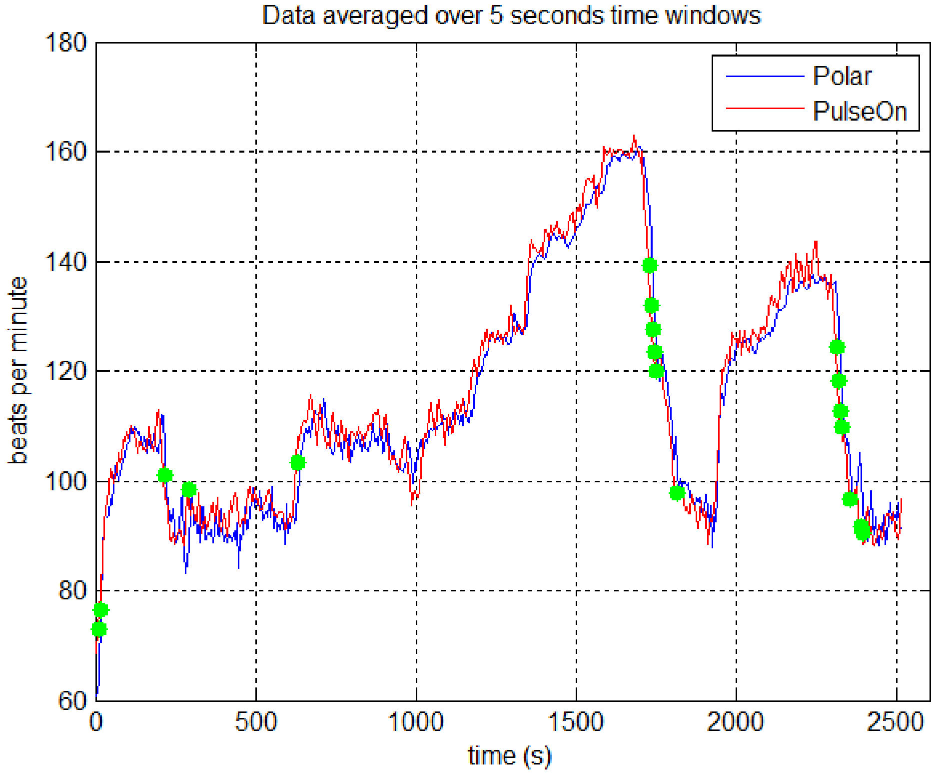
Subject information list - PulseOn testing			
Measurement basics:			
	Subject ID number:		[Sxxx format]
	Measurement type ID:		[Maaa format]
	Date of recording:		[dd.mm.yyyy]
	Start time:		[hh:mm] according marker
	End time:		[hh:mm] according marker
Subject Basic Identification			
	Date of birth:		[dd.mm.yyyy]
	Gender:		[male / female]
Physical Characteristics			
	Weight:		[kg]
	Height:		[cm]
	Wrist circumference:		[cm]
	Dominant hand:		[right / left / both]
Health Status:			
	Do you have any illness or disease which would prevent from physical exercise of ~30min ?		[yes / no]
	Activity class:		[0 - 10 according FirstBeat]
	Smoking:		[according table below]
Skin color:			
	Fitzpatrick scale:		[I - VI]
Smoking classification:			
	no-smoker (0 cigarettes)		
	ex-smoker		
	sometimes (a few cigaretes per month)		
	occasionally (less than 1 pack per week)		
	smoker (more than 1 pack per week)		

Appendix 2: Heart rate figures

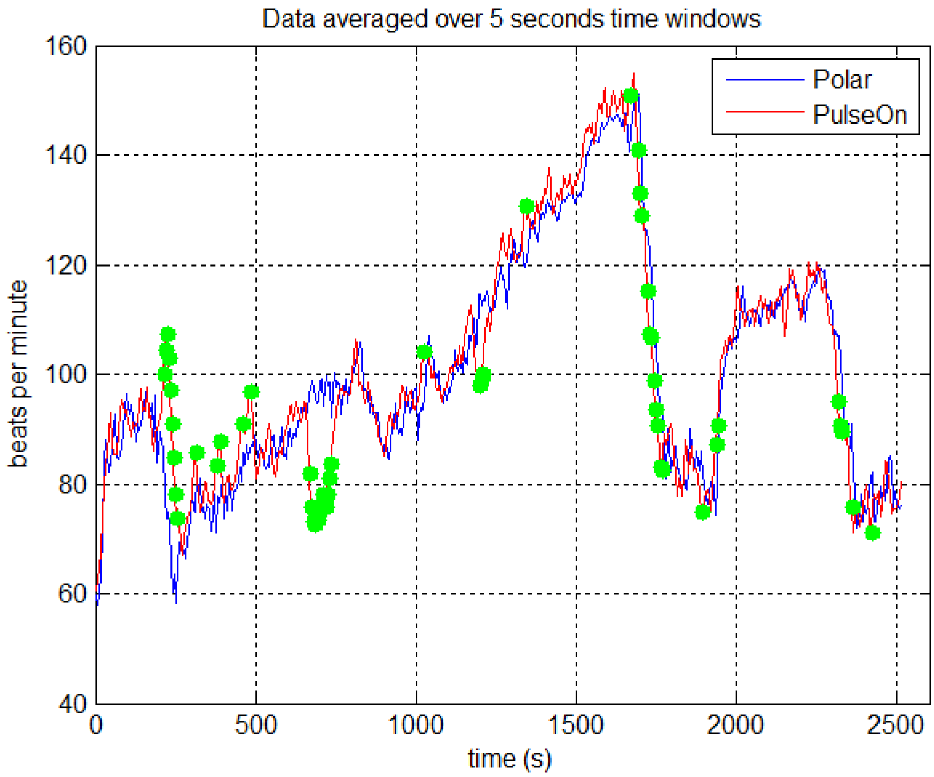
Subject 1



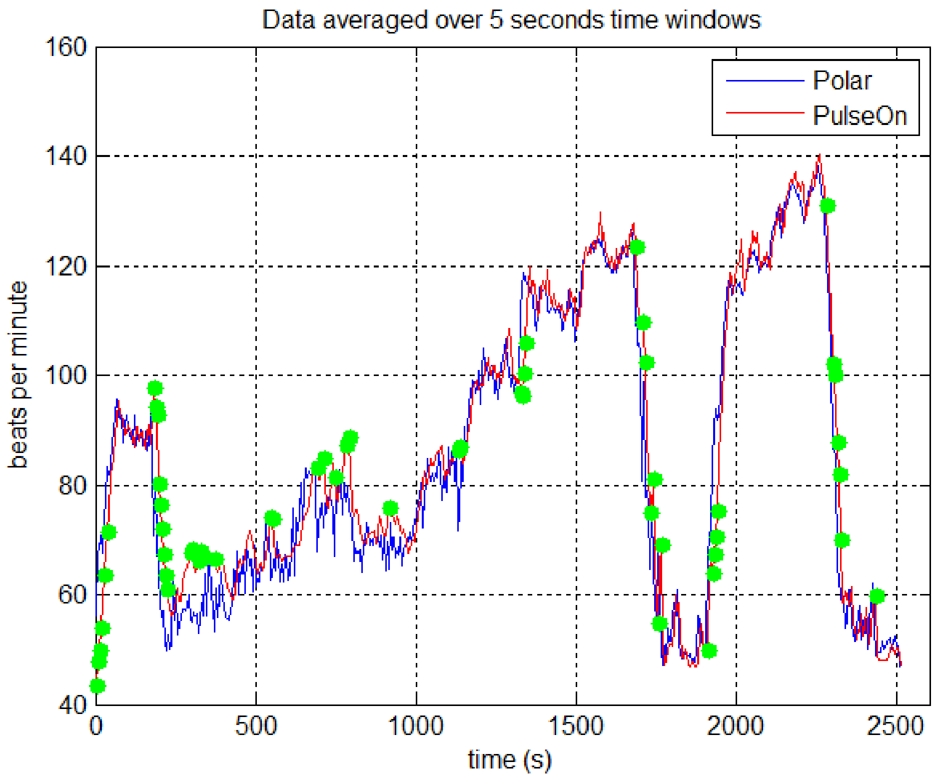
Subject 2



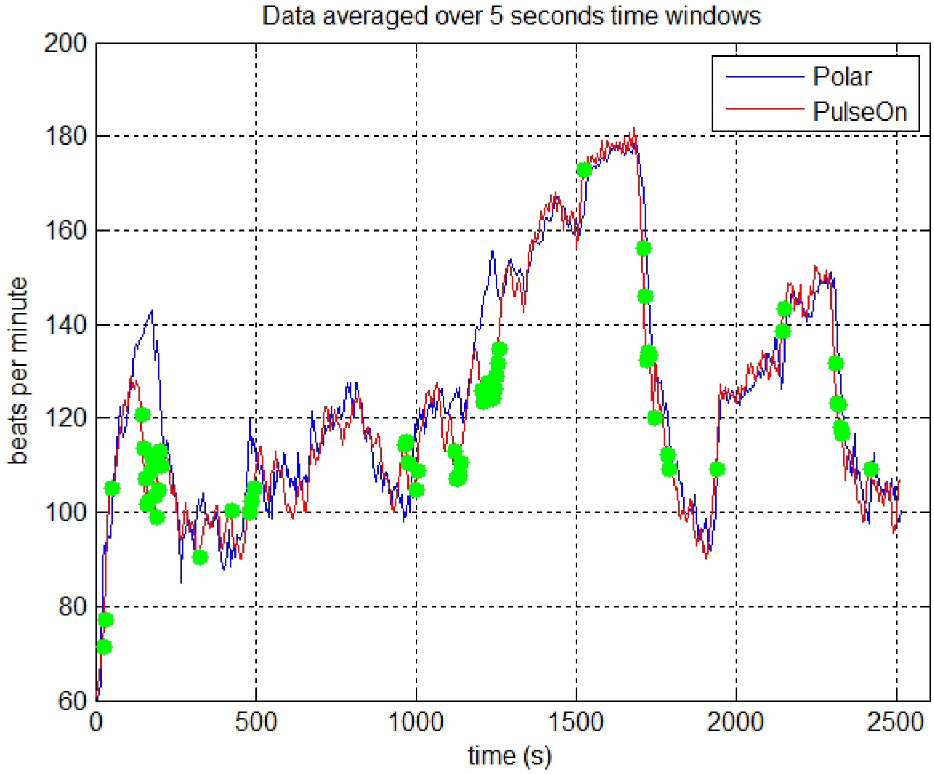
Subject 3



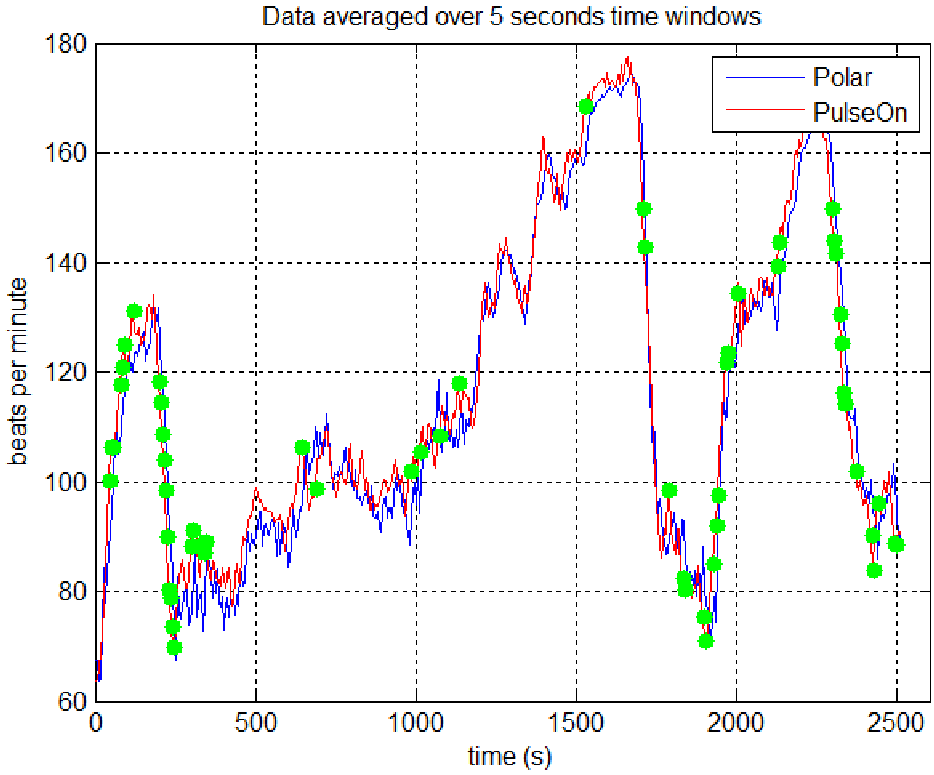
Subject 4



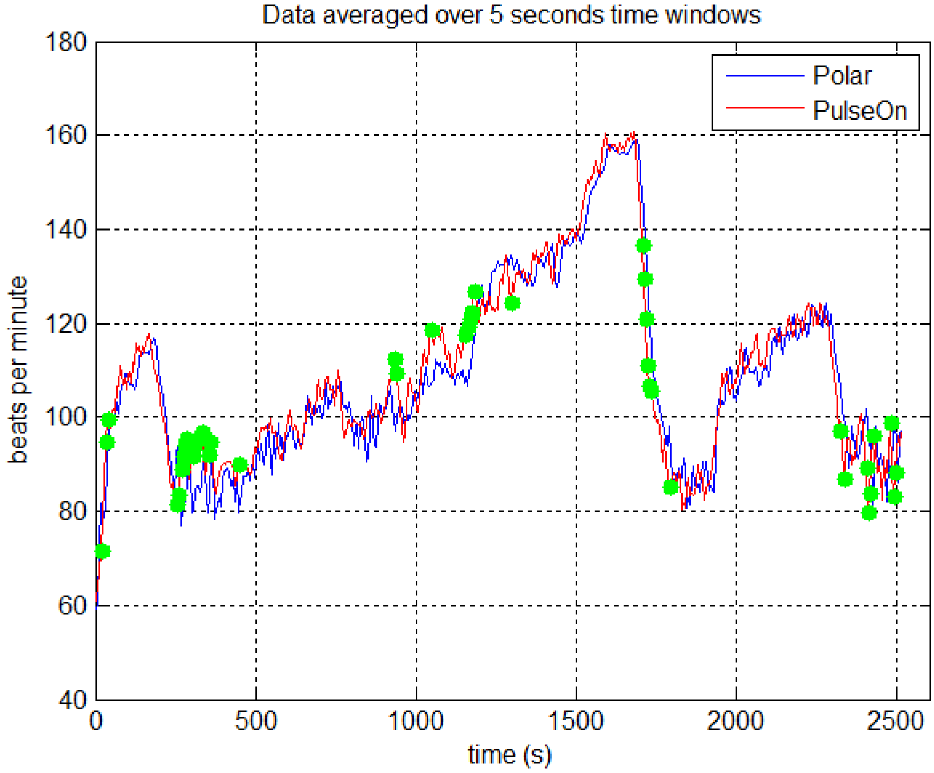
Subject 5



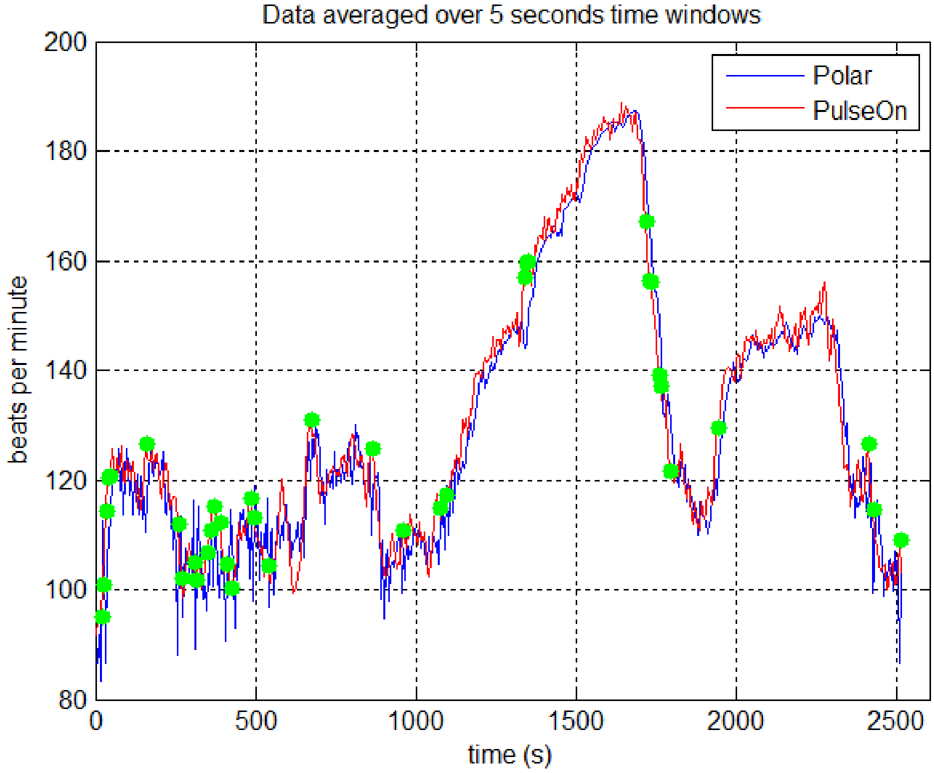
Subject 6



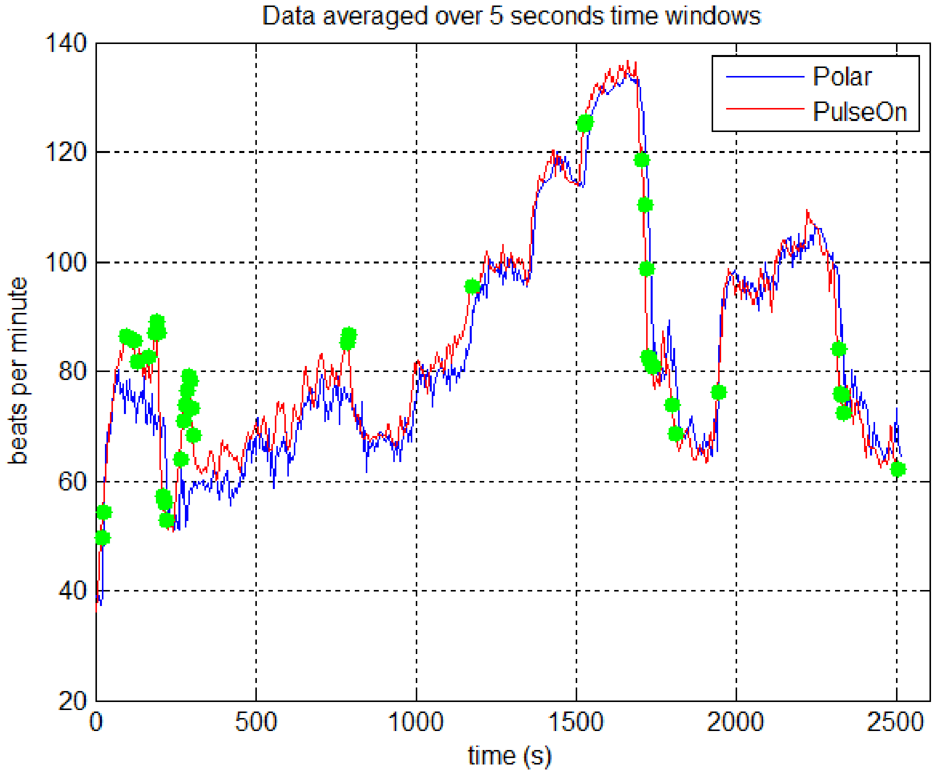
Subject 7



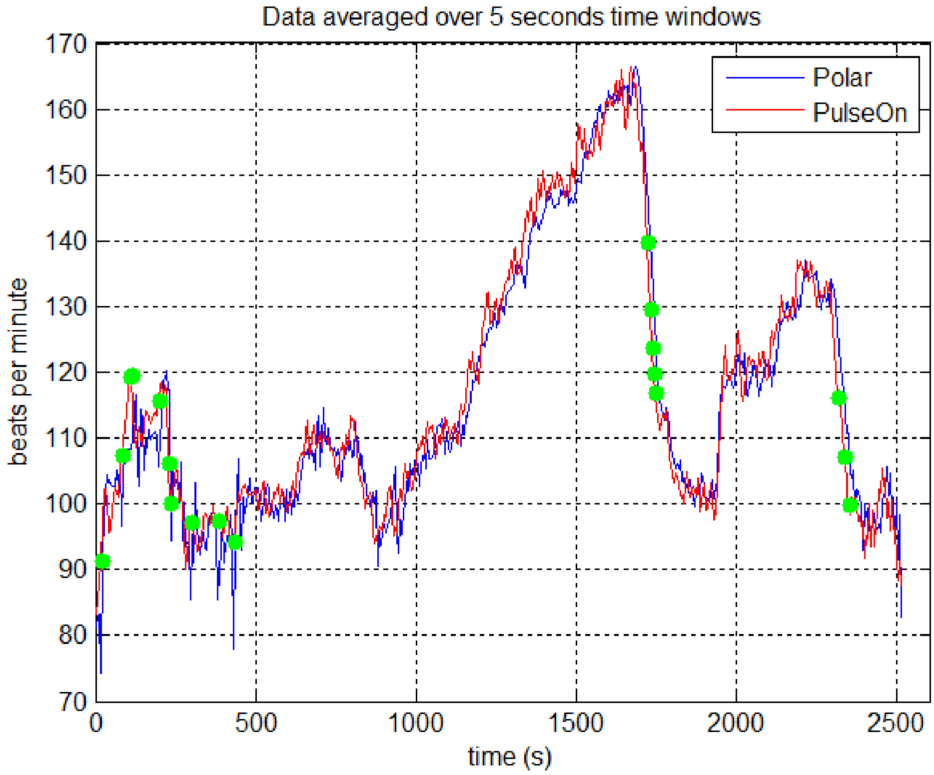
Subject 8



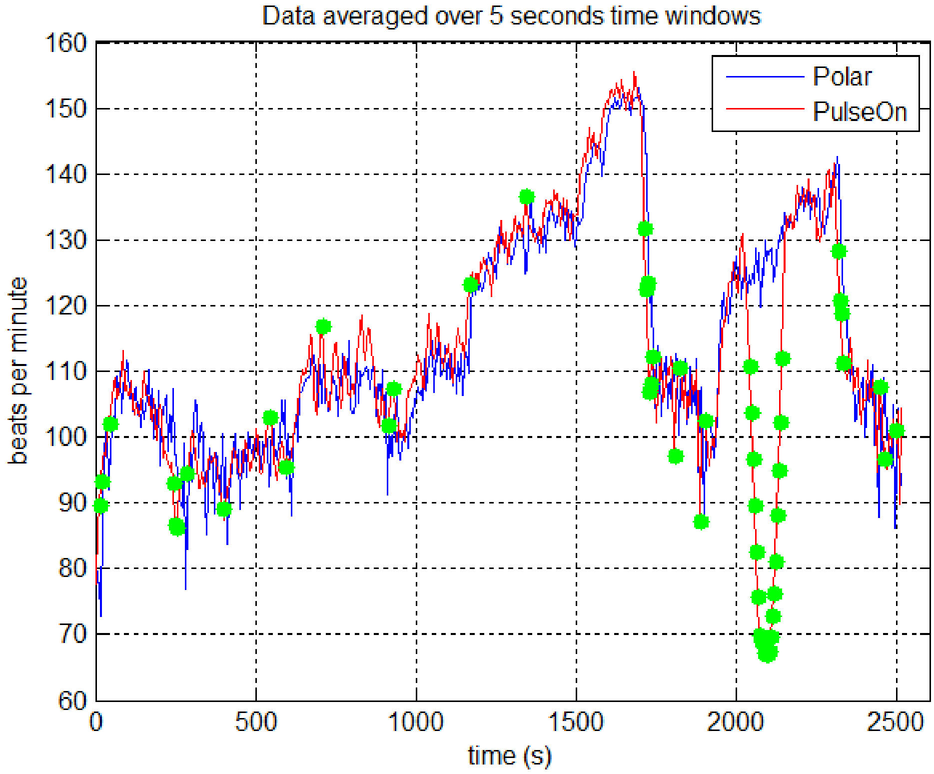
Subject 9



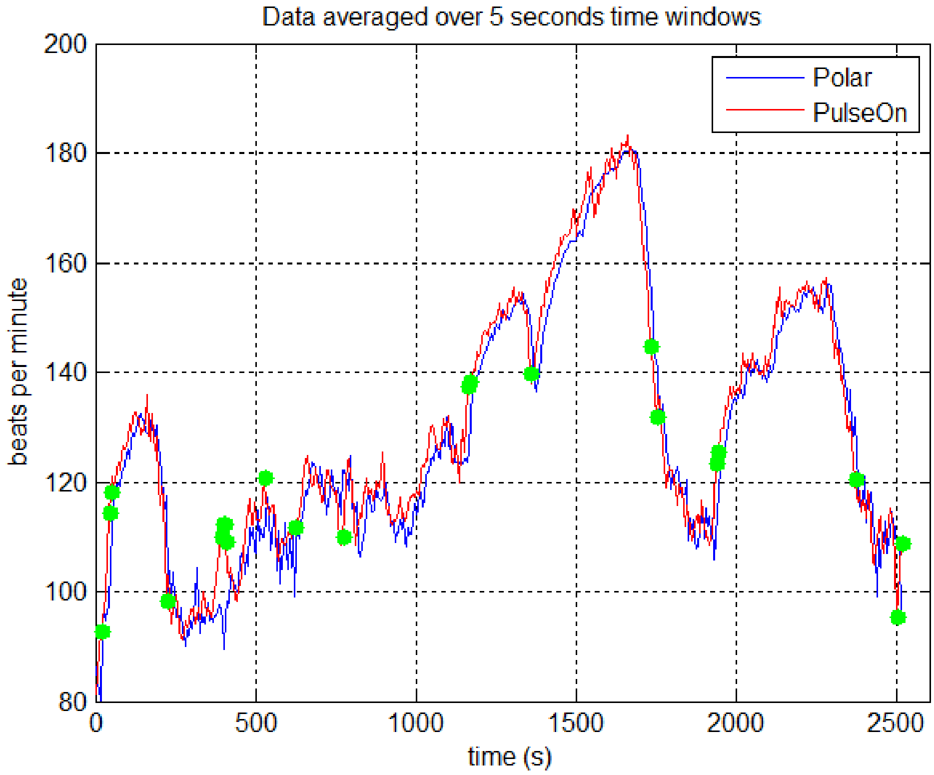
Subject 10



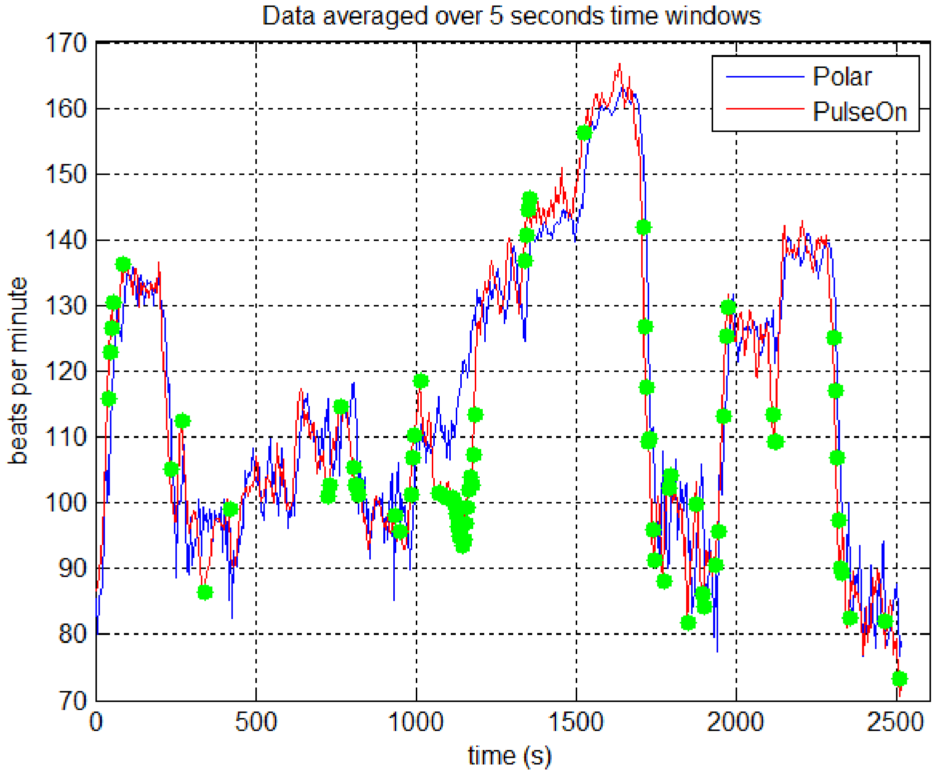
Subject 11



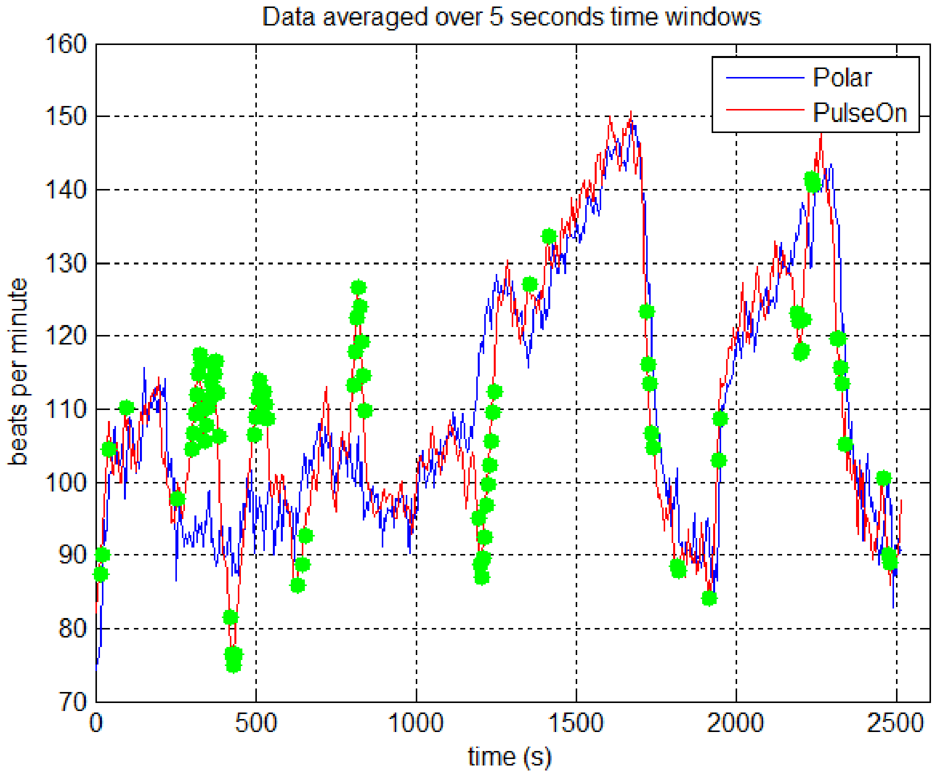
Subject 12



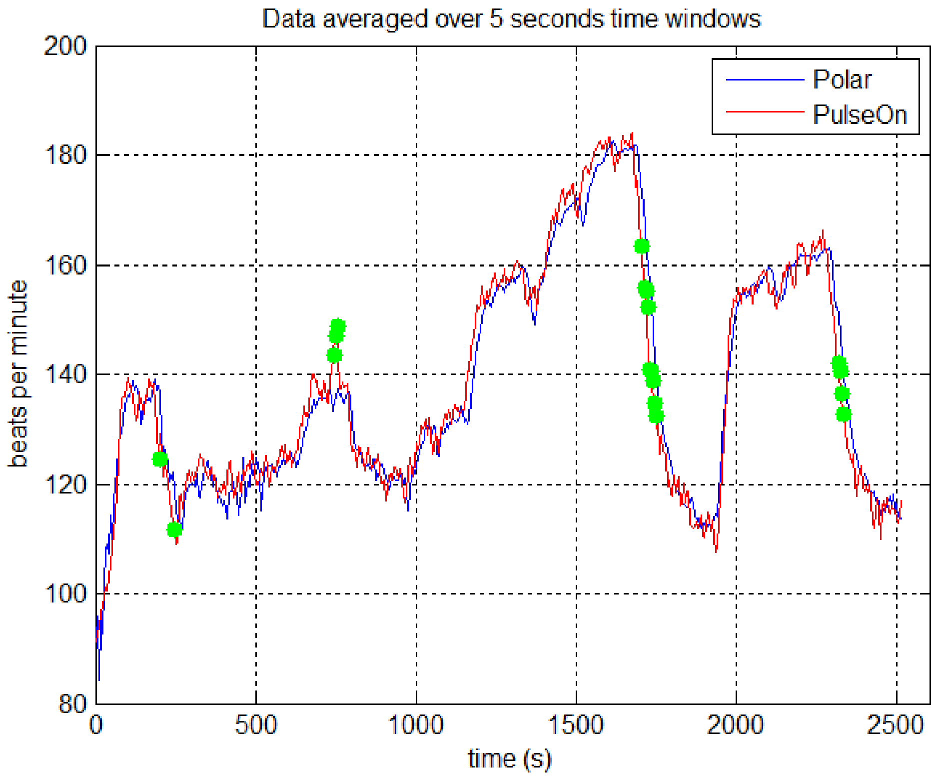
Subject 13



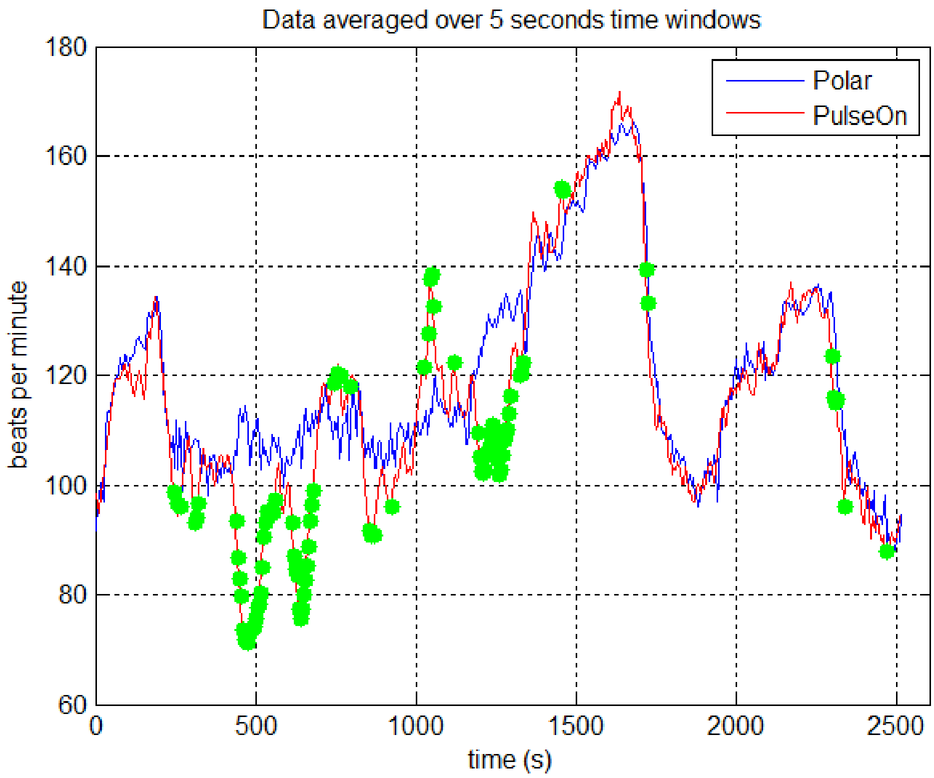
Subject 14



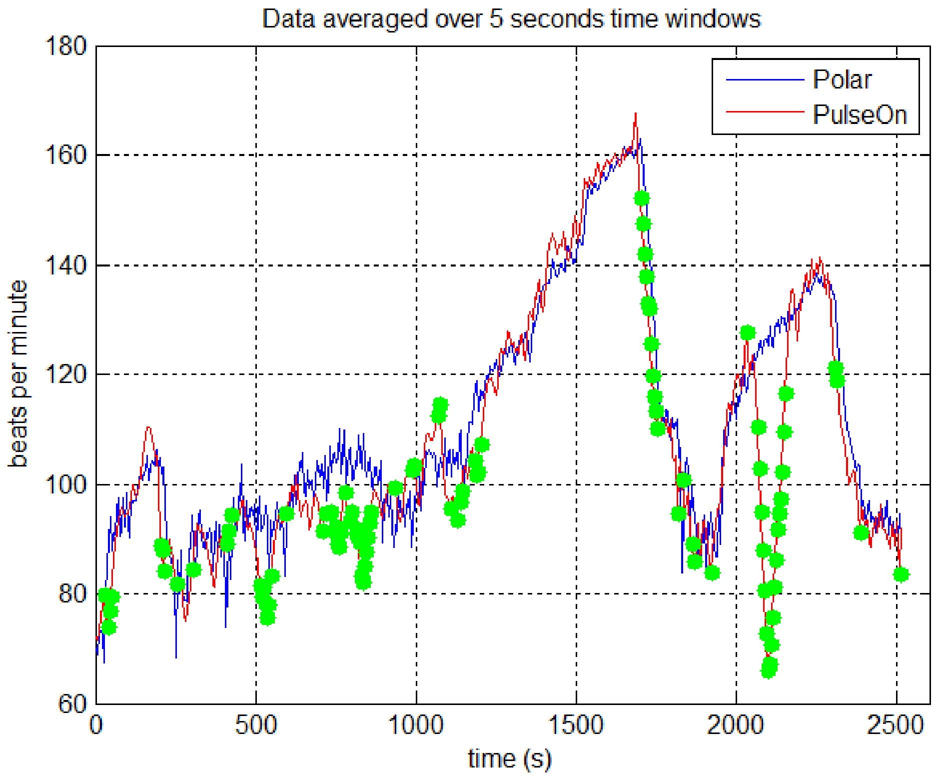
Subject 15



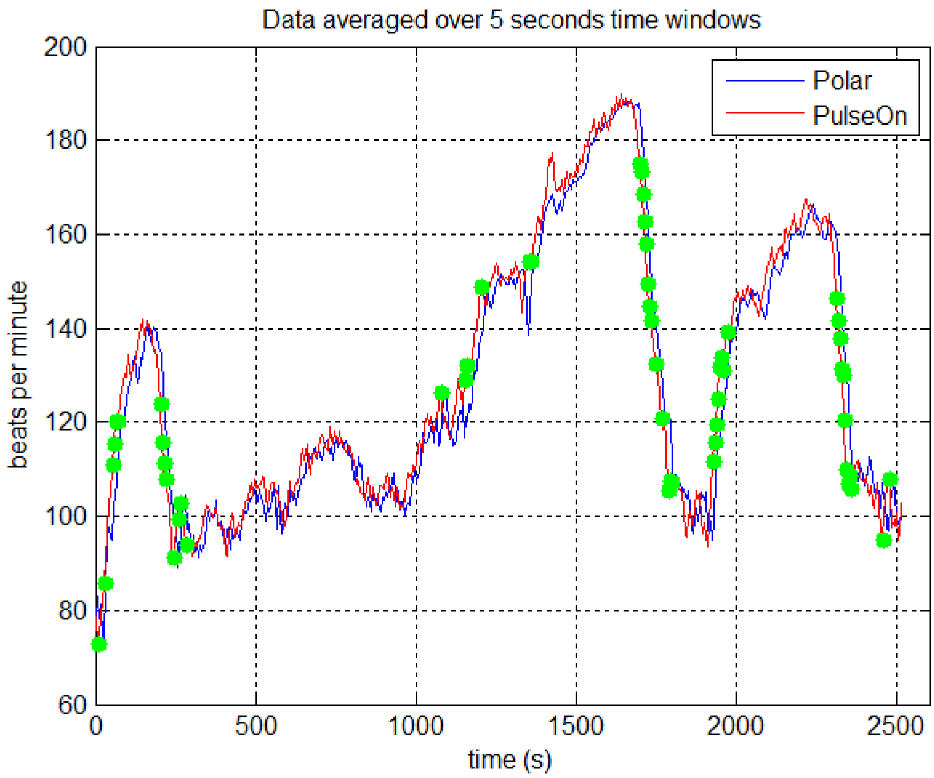
Subject 16



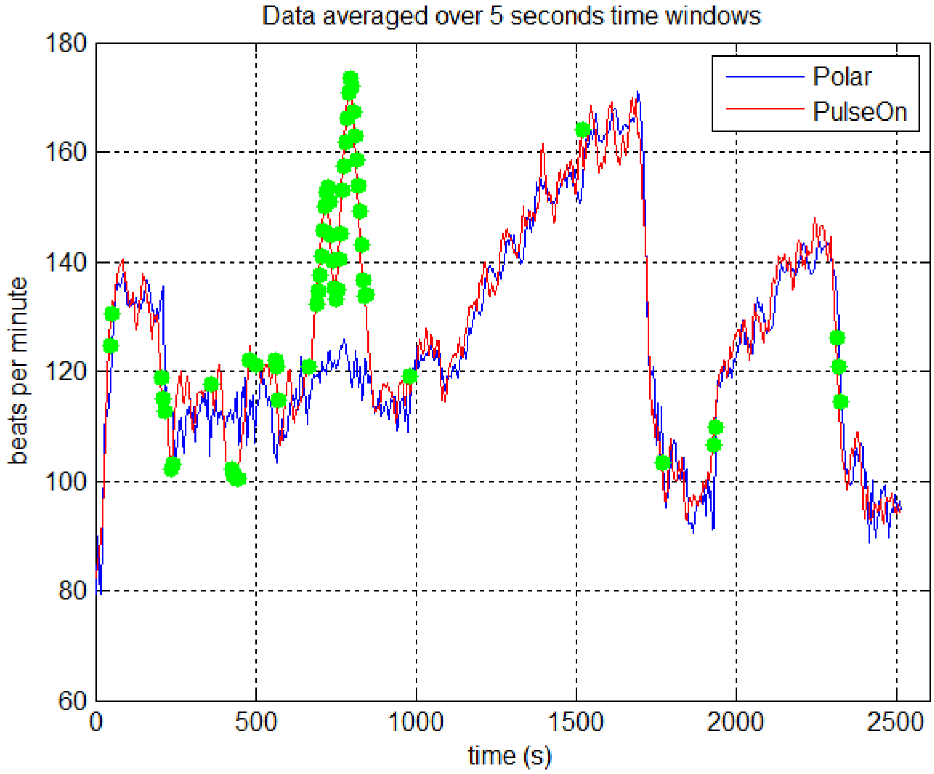
Subject 17



Subject 18



Subject 19



Subject 20