NTNU – Trondheim
Norwegian University of
Science and Technology

Ambient-light Photoplethysmography

- How can I tell your pulse from looking at
your face?

Åsmund Rustand

Sammendrag

I pasientovervåkning, både på sykehus og i ambulerende behandling, er det dyrt utstyr som er normen, spesialutviklet og produsert for å utføre bestemte oppgaver. Såvel i disse sammenhengene som i hjemmemarkedet vil det allikevel alltid finnes en etterspørsel etter billige og allsidige løsninger, der hvor disse kan gi fornuftige og nyttige opplysninger.

Selv om pulsoksimetri og andre typer biologisk anvendt spektroskopi er nokså velkjent teknologi, og på tross av at studier er gjort på feltet, kan det for noen være et overraskende faktum at man kan hente ut fysiologiske parametre som hjerte- og pustefrekvens kun ved å filme lysstyrkevariasjoner i noens hud ved hjelp av konvensjonelt forbrukerutstyr for videoopptak: kameraet kan "se" disse variasjonene, men vi kan ikke selv uten videre oppfatte dem.

Denne avhandlingen er et forsøk på å utforske visse aspekter av temaet som kalles *kontaktfri, omgivelsesbelyst fotopletysmografi*, hvor vi håper på å avdekke fordelaktige egenskaper teknikken måtte ha. Vi har et klart fokus på tekniske muligheter, men medisinsk relevans blir også diskutert og bekreftet.

Konklusjoner fra tidligere arbeider eksemplifiseres—muligheten vi har med denne metoden til å beregne nøyaktige estimater av hjertefrekvens, fra hudoverflaten både i ansiktet og i håndflaten, med en forsøksperson som beveger seg eller sitter stille; tilstedeværelsen av signalfluktuasjoner beslektet med respirasjon. Dertil, at pulseringen er sterk nok til å muliggjøre underinndeling i mindre segmenter og kartlegging av pulsamplitudens distribusjon dem imellom.

Det lykkes oss å innhente annen informasjon: signalets faseinformasjon er robust nok til at ikke bare hjertefrekvensens gjennomsnitt over lengre tid, men også korttids variabilitet kan ses med mer enn et minimum av detaljoppløsning. Fra et opptak hvor både ansikt og hånd er synlig estimerer vi faseforsinkelsen, dvs. forskjellen mellom blodtrykkbølgens gangtider til panne og til håndflate, og sannsynliggjør derved ytterligere påstanden om faserobusthet.

Konseptet *uavhengighetsanalyse* (Independent Component Analysis) har blitt foreslått som en måte å forbedre metodens resultater på – en fremgangsmåte vi undersøker og finner begrenset støtte for. Ved andre, enklere tilnærmelser til signalbehandlingen avdekkes positive egenskaper når det gjelder å oppnå et klart signal.

Problem Statement

The medical care system is going through radical changes caused to a large degree by technological innovation. At the same time as increasingly expensive diagnostic and therapeutic methods are being developed, utilization of new mass-produced electronics makes the screening processes for many serious diseases easier and cheaper.

In the present project, research is intended performed that may result in new diagnostic tools for the evaluation of abnormalities in the cardiovascular system. Video recording of reflected light in the RGB channels and subsequent signal processing represent extremely inexpensive and, when developed, relatively simple screening tools. These methods will be unlikely replacement candidates for present diagnostic modalities, but may well become tools for the screening and identification of serious conditions.

Investigate the use of low-cost equipment (e.g. web camera) in non-contact ambient-light reflectance photoplethysmography. Explore previous work. Questions to answer include how to extract the cleanest possible signal, and what information apart from heart rate can be acquired from it. Study signal processing algorithms for estimating heart rate and other relevant parameters and, where possible, compare the results with standard techniques.

Abstract

This thesis is an attempt to explore certain aspects of the subject called *non-contact, ambient-light photoplethysmography*, where we hope to reveal beneficial properties the technique may have. We have a clear focus on technical possibility but the medical relevance of it is also discussed and confirmed.

Conclusions from previous work are exemplified – the ability we have with this method to compute an exact heart rate estimate, from the skin surface both on our face and in the palm of our hand, with and without the subject moving during recording; the presence of signal fluctuations akin to those of breathing. Moreover, that the pulsations are strong enough to enable subdivision into smaller segments and a mapping of how the pulse amplitude is distributed among them.

Other information is obtained: the phase information contained in the signal is robust enough that not only the average heart rate over some time but also its short time variability can be seen with more than a minimum of detail resolution. From a recording where both face and hand is in view, we further substantiate the claim of phase robustness by estimating the phase delay, i.e. the difference in blood pressure wave travel time between forehead and palm.

The concept of *Independent Component Analysis* has been suggested as a way of improving the output of the method—a procedure we investigate and find limited support for. Other, more simplistic signal processing approaches are found to have positive traits in achieving overall signal clarity.

Contents

Chapter 1

Introduction

1.1   Motivation

In patient monitoring, clinical as well as ambulatory, expensive devices developed and built for performing exact tasks are the norm. Both in such situations and in home care, however, there will always exist the demand for low-cost, multi-purpose solutions, whereever we can expect these to provide us with sensible and useful data.

Despite pulse oximetry and other kinds of biologically applied spectroscopy being fairly well known technologies, and even though studies in the field have been done, some may regard it a surprising fact that physiological parameters like heart and respiratory rates can be extracted using only conventional, consumer-grade video recording equipment to film the brightness variations in someone's skin: the camera can "see" these variations but we cannot ourselves sense them.

One study ([1], see ch. 2.4) suggests the use of Independent Component Analysis (ICA) in extracting photoplethysmogram (PPG) heart rate information and backs this with experimental results. Specifically, ICA is in that study applied on a signal extracted from a video segment where (after use of automatic face detection on a recording where the subject is moving) all of the subject's face is visible. We want to investigate this further, by applying ICA also in other situations. We know that the wanted PPG signal can be "seen" most clearly in the skin on the forehead, cheaks etc., meaning the whole-face signal will be a noisy starting point. Will other, less noisy starting points also benefit from the application of ICA? This is a question we here seek answered. Can other, less sophisticated approaches be used to our advantage? We suggest the rather obvious in not using the results from face detection directly, but only the part containing the forehead. By mapping the spatial distribution of pulse amplitude on the skin, we will also see if we can pre-confirm the pulsatility of the signal we wish to obtain.

Another study ([2], see 2.2) gives a more general assessment of what we can expect from the method, an example of which is the presence of respiration infor-mation. Likely there will exist other studies that we weren't aware of and did not

include in our reading (see bibliography p. 51).

Some further interesting questions to be raised are: What is the minimum recording duration needed to estimate "instantaneous" heart rate? Can we record simultaneously the skin on two locations (e.g. hand, face) and find such information as the phase delay between them? Can we apply the technique to not only skin but also internal organs? These questions are attempted answered in this report.

1.2 Report Structure

The report starts by laying out background information on a few physically and medically important issues in chapter **2**: the behavior of light absorption in hemoglobin as the basis for the photoplethysmogram in **2.1**, the diagnostical relevance of pulse transit time in **2.5** and of heart rate variability in **2.6**. In addition, two recent studies are revisited and their results described in some detail in **2.2** and **2.4**.

General aspects and specific involved methods of analysis are dealt with in chapter **3**, some of them explained in detail. This includes simple procedures for phase delay estimation between two blood-pressure-like signals in **3.3** and mapping of the distribution of the pulse amplitude on skin surface in **3.4**, as well as two ways of estimating instantaneous heart rate from a video "snapshot" of only a few seconds duration in **3.5**.

A short preliminary exploration is done in chapter **4** in an attempt to give clues to what information we can expect to reveal and to affirm some conclusions from previous studies on the subject. Chapter **5** gives only a brief overview of the experiments that are to be executed, as more detail is given alongside the presentation of the results in chapter **6**:

6.2 and **6.1** give the results of heart rate estimations respectively with and without movement. The application of JADE independent component analysis (see **2.3**) is invoked in several different cases to assess the benefits thereof. The pulsatility mapping procedure is tried out in **6.3** on videos of the face and the hand, separately and both combined in one video. This latter combination is investigated further in **6.4**. A small test case is presented in **6.5** in which we try to confirm the applicability of the pulsatility mapping method, as well as shed some light on what minimum duration of video is needed to achieve a stable heart rate estimation. In **6.6** a short laparoscopy recording of a cancerous liver is attempted analyzed.

The preliminary explorations of photoplethysmogram contents are discussed in **7.1**. Following this, the rest of chapter **7** discusses the other results sections in the same order that they appear in chapter 6. Conclusions are made in chapter **8** by answering two questions from the problem description—**8.1**: How do we extract the best possible signal? and **8.2**: What information can we acquire from this signal?—as best we can from presented results and subsequent discussion.

Chapter 2

Background

2.1 Photoplethysmography

Stemming from the Greek plethysmos *to increase*, *plethysmography* means "[...] finding variations in the size of a part owing to variations in the amount of blood passing through or contained in the part." [3].

Using conventional plethysmography one measures pulsatile tissue volume directly, e.g. with a strain gauge that measures changes in the circumference of an extremity. This will measure the sum total of volume changes in any and all blood vessels. Of these, arterial pulsations are the most significant. Capillaries are largely non-compliant and will register with only minor pulsations. Venous oscillations may be evident but are, depending on measurement technique, often avoided by applying external pressure [4]. The plethysmogram is used as an indirect measure for arterial blood pressure (ABP) [5].

*Photo*plethysmography is related to, but not equivalent to, traditional plethysmography. *Transmission mode* PPG measures how light is obstructed and absorbed through tissue (e.g. finger tip, earlobe) by use of an LED on one side and a photodetector on the other. When the LED and the photodetector are placed alongside eachother on a tissue surface to measure the amount of light that is reflected, we have *reflectance mode* PPG.

Several factors play a role in the formation of the AC[1] component of the PPG signal, the most important of which are given in the schematics in fig. 2.1. Grey boxes signify externalities that differ according to measurement methodology: amount of light, physical properties of the sensor and applied external pressure; while yellow are simple physiologic parameters. ABP is the main source of PPG signal variations, as suggested by their morphological similarities (compare fig. 2.2 and fig. 2.4(a)), and it is indeed reported that the dependence of the PPG on the ABP is similar to that of the true plethysmogram on the ABP [4,6].

In *non-contact PPG*, changes in ambient light and automatic brightness ad-

---

[1]Where AC is used specifically in reference to pulsation generated by individual heartbeats, as opposed to their baseline.

justment in camera or software have a significant influence on the baseline[1] of the signal, but also physiologic parameters like changes in capillary density[7] and in venous volume fluctuations [4].

An application of non-contact reflectance mode PPG for combat or field triage situations is investigated in [8]. A recent innovation in contact reflectance PPG are applications using popularly available smartphone technology, able to record a high quality PPG signal when you place your finger covering both the photo-light LED and the camera lens [9]. Experimental results have been presented and discussed regarding the use of PPG and pulse oximetry in laparoscopic[2] surgery [10].

Transmission mode PPG is routinely used in clinical settings (pulse oximetry). **Non-contact reflectance mode PPG using ambient light** is a field of research on the incline, offering promising insights into the future of non-invasive circulatory monitoring [1,2].
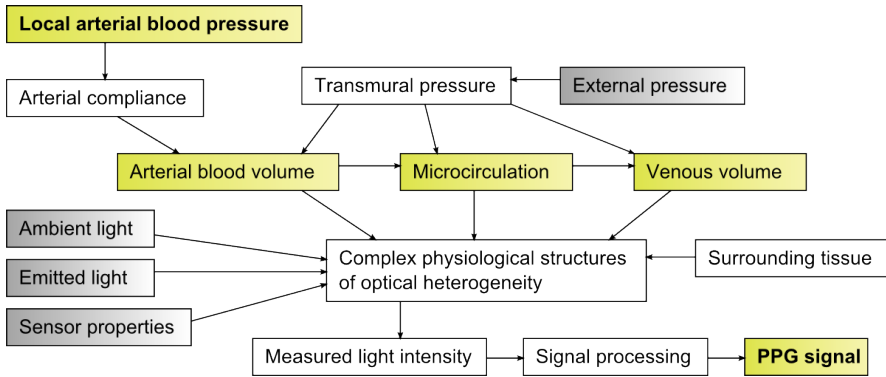

Figure 2.1: General processes and interactions that underlie the photoplethysmogram (PPG) [4]
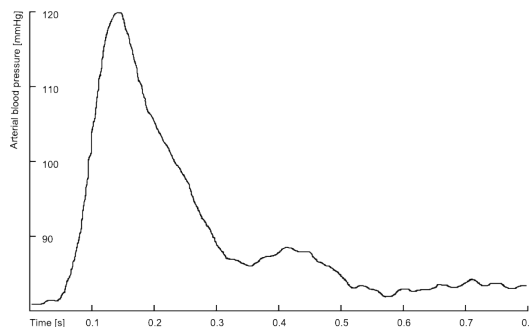

Figure 2.2: Radial ABP waveform cycle

---

[2] "Laparoscopy. Examination of the interior of the abdomen using a laparoscope, which is a type of *endoscope* [..] (see *minimally invasive surgery*)." [3].

2.1.1 The Hemoglobin Absorption Spectrum

The complex hemoglobin molecule makes up most of the content of the blood and is the carrier of oxygen in the blood stream. It is precisely therefore also the physiologic basis for the PPG: incident light on the skin travels into the underlying layers where a significant part is absorbed in blood and surrounding tissue, and the remaining part reflected back to the surface.

Fig. 2.3 is taken from [11] and modified to include an approximate color spectrum. It shows the absorptivity spectrum functions for (1) oxyhemoglobin (HbO_2) and (2) deoxygenated hemoglobin (Hb) in the wavelength span 450–700 nm. We observe: on red wavelengths absorption is low – higher on blue and yet higher on green, including peaks in both Hb and HbO_2. The Hb absorptivity is generally lower than that of HbO_2 on blue and green wavelengths.
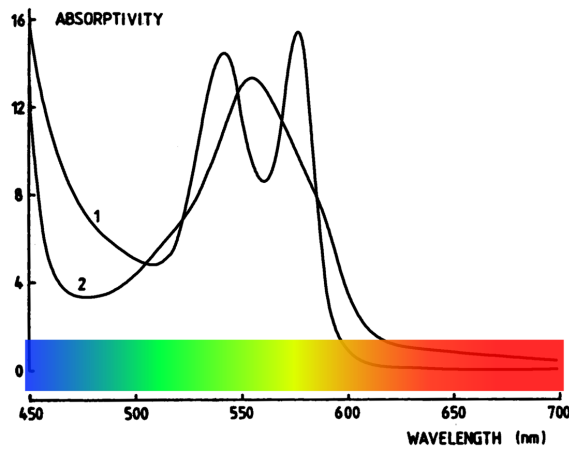

Figure 2.3: HbO_2 (1) and Hb (2) absorption spectra $[L \cdot mmol^{-1} \cdot cm^{-1}]$

The values from [11] are obtained by *transmission* of light through *in vitro* blood. The interpretation (as pertaining to PPG) of these values is supported by [12], where the *reflectance* of *in vivo* tissue is theoretically assessed and the findings corroborated by experiment.

In a PPG recording using an RGB colorspace we would expect seeing heart rate pulsations in the green channel, where the absorption is the highest, and to a lesser extent in the blue channel. More light is absorbed when the blood is more dense, i.e. "on the beat", meaning the heart rate pulsations will be **inversely related to the ABP**. Signal variations due to changes in the closely interrelated concentrations of Hb and HbO_2 are also expected to arise in both the green and the blue channel where their absorptivities are differing significantly. The red channel will have higher pixel values because of more light being reflected, but less variations owing to physiological parameters.

2.1.2   Pulse Oximetry

Optical oximetry was developed during the last world war and was able to provide a continuous estimate of blood oxygen saturation (SaO_2) from a light-weight device placed on the earlobe. The same principle is applied here as in laboratory oximetry, in which different wavelengths of light is shone through a small container of blood. Using knowledge of the hemoglobin absorption spectrum, one measures HbO_2 and Hb concentrations with one distinct wavelength for each, and calculates

$$SaO_2 = \frac{HbO_2}{HbO_2 + Hb}.$$

Pulse oximetry (contact transmission mode PPG) is an extension on this, overcoming difficulties regarding other light absorbers than hemoglobin by using flashes of red and infrared light to obtain the pulsatile signal purely related to arterial blood. The possibility of heart rate measurements is a byproduct of this. Today's finger pulse oximeters use this method, and also compensate for the unwanted influence of external light. They are regarded as standard in clinical monitoring care but are also cheaply available for home use [13].

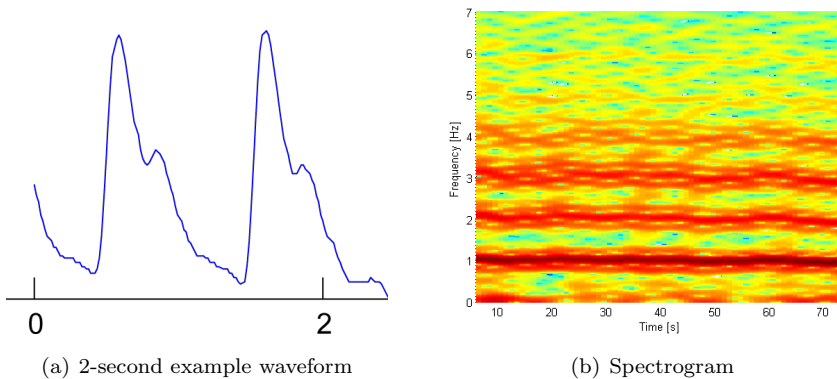(a) 2-second example waveform   (b) Spectrogram

Figure 2.4: Output from finger pulse oximeter

In fig. 2.4 is displayed example output from the finger pulse oximeter used for heart rate control purposes in this report (see B.2), (a) 2 seconds of pulse waveform when a heart rate of about 60 bpm was measured, (b) spectrogram from about 1 minute of the same measurement. SaO_2 measuremeants were not used.

An application with some similarities has recently been made popular through smartphone programs that are able to record a high quality (contact reflectance mode) PPG signal when a finger is placed covering both the photo-light LED and the camera lens [9]. The lack of control over lighting means only the pulse *waveform* is available in this way, no absolute values and *no oximetry*. In this it is similar to the non-contact ambient-light PPG investigated in this report.

2.2 Study by Verkruysse, Svaasand, Nelson

A survey of non-contact ambient-light PPG is done in the 2008 article titled "Remote plethysmographic imaging using ambient light" [2], using a consumer level digital photo camera in video mode. It is shown that PPG signals can be measured on the human face with normal ambient light as the source, from a distance of more than one meter; revealing up to four harmonics of the fundamental heart rate (HR), in addition to the respiration rate (RR), and modulation products between these.

Videos of different subjects were recorded with durations of 30 seconds to a few minutes, where they where sitting or laying down to reduce movement, with daylight as the only source of illumination. Frame rates of 15 or 30 fps and resolutions of 640×480 or 320×240 were used. Regions of interest (ROI's) were selected manually, and the pixel values in these averaged for each of the R, G and B channels and for each frame to give one raw signal for each channel. The assumptions that are made about the PPG signal in ch. 2.1.1 are confirmed: the green channel shows clearer signs of both HR and RR than the blue channel, while the red is mostly lacking in any interesting content apart from sometimes showing traces of the RR.

A case is made for the use of pulsatility (pulse amplitude) mapping and phase mapping on a grid of 40×50 cells. Each of these cells is considered an ROI, and maps are constructed displaying their phases and amplitudes at the fundamental HR frequency (from the green channel). Phase mapping is applied to a case of so called "portwine stains"[3] right after laser surgery, showing clear contrast to the surrounding areas.

Another experiment is conducted in which the subject is instructed to hyperventilate. This has a clear influence on the heart rate, but also the signal content that is seen at around 0.15 Hz is found to increase – a clear indication that these are variations that are caused by respiration (breathing causes varying oxygen concentration) as the authors are indeed arguing.

Further arguments are made that the signals acquired are true PPG signals: (1)–HR is not visible in areas affected by involuntary movement but not containing any bare skin (e.g. hair, clothing) and (2)–the green channel contains by far the strongest HR signal, even though its actual pixel values are much lower than the red's. (2) is strong evidence that the signals are filtered by variations in blood volume (due to the absorption bands for Hb and HbO_2 for yellow and green light).

2.3 Independent Component Analysis (ICA)

Sensors employed in measurements of any quantity will commonly output a signal which is a mixture of several independent physical processes. In independent component analysis (ICA), a special case of the wider concept of blind source separation (BSS), a representation is sought where this mixture is a linear transformation of the original data (for reasons of computational and conceptual simplicity), i.e.

$$x = \mathbf{M} \cdot s$$

---

[3] "A purple-red birthmark that is level with the skin's surface. It is a permanent type of haemangioma." [3]

or with a dimensionality of 3 and vectors as functions of time,

$$[s_1(t), s_2(t), s_3(t)] = \mathbf{M^{-1}} \cdot [x_1(t), x_2(t), x_3(t)] \qquad (2.1)$$

where x is the vector of knowns and s the underlying physical signals. $\mathbf{M}$ is found in ICA as the matrix maximizing some measure of the statistical independence of the components of s [14]. Often used measures are based on the *nongaussianity* of the variables, a property that (by the central limit theorem) implies independence[4]. The well-known FastICA algorithm is based on *negentropy*, a measure of nongaussianity based on the information-theoretic quantity of (differential) entropy [15].

ICA is very well exemplified by the application to speech data [16] (see illustration in fig. 2.5). If several people are in the same room, and what they are saying is recorded by several microphones placed there, each recording will be a different mixture of the speech signals. If the number of people is the same as the number of microphones, ICA can recover separately each of the source signals, on the basis of the statistical independence that follows from their physical independence.

There are thoughts of applying some form of constrained ICA (cICA) to the subject matter of this report, but this is not investigated further here [17].

2.3.1   JADE

The Matlab implementation of Joint Approximate Diagonalization of Eigenmatrices (JADE) used in this report is available online [18], an implementation by the original author of the algorithm [19, ch. 4.2]. JADE approximates the statistical independence of the sources with their *kurtosis* (fourth moment) as the measure of nongaussianity [1, 19].
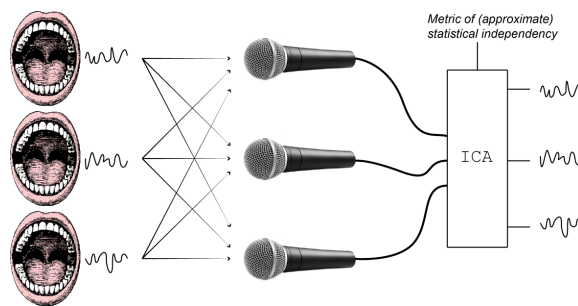

Figure 2.5: ICA source separation illustrated

---

[4]If we can find a set s of variables each less gaussian than the ones in x, they will likely be less codependent. This because the sum of any two signals will tend to be more gaussian than each of them.

2.4   Study by Poh, McDuff, Picard

This 2010 MIT article, titled "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation" [1], explores the possibility of automatically measuring heart rate using a normal web camera and face recognition. Videos of 15 fps were recorded in 24-bit RGB color space at a resolution of 640×480 pixels with an Apple iSight built-in camera at a distance of approximately 50 cm. The experiments were conducted indoors and with a varying amount of sunlight as the only source of illumination, on 12 subjects of varying age and of both genders. Results were compared to and validated by measurements from a blood volume pulse (BVP) sensor. Subjects were told to be moving around gently as if interacting normally with the computer.

The analysis was performed on videos of 1-2 minutes length. Firstly, face recognition (OpenCV Viola and Jones) [20] was used on each frame to identify the region of interest (ROI). Secondly, the three raw traces were extracted by averaging the pixel value within the ROI, for each frame, for each of the R, G and B channels. Finally the R, G and B raw traces were fed through the JADE algorithm, resulting in three new, statistically more independent signals. From these, the second channel was the one most often found to be containing the most pulsatile information, and it was generally chosen as the basis for estimating values for the heart rate. A 30 second epoch was used for heart rate estimations.
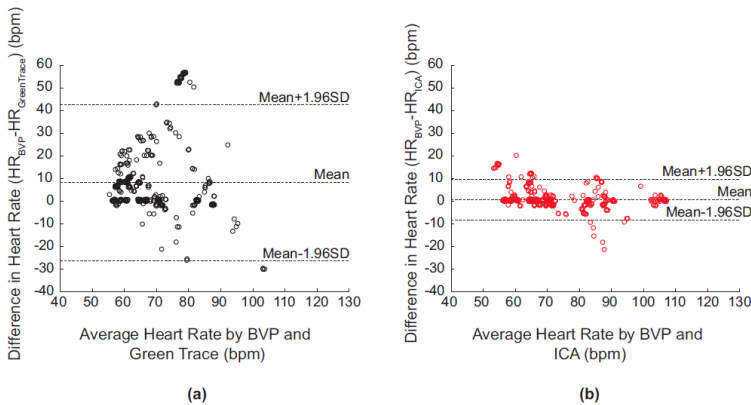

Figure 2.6: Results from Poh, McDuff, Picard study

Fig. 2.6 is borrowed from the article and gives an overview of the agreement between the BVP sensor and heart rate estimations from the raw green trace (a) and the ICA method (b). The plots are *Bland-Altman* plots, which plot the average vs. the difference of the estimated value and the control value[5], for 372 measurement points from the 12 participants. The overall root mean square error (RMSE) was 19.36 bpm for (a) and 4.63 bpm for (b).

---

[5]In a Bland-Altman plot, a narrow vertical spread indicates good accuracy, while a wide horizontal spread indicates an appropriate testing range.

2.5   Pulse Transit Time

Pulse transit time (PTT) is the time it takes the pulse pressure wave to travel from the heart to a given measurement point on the arterial tree. The pressure wave originates at the left ventricle of the heart where blood is ejected through the aortic valve into the aorta, and the pulse wave velocity (PWV) is much greater than the forward movement of the blood itself [21]. A decrease in the compliance (an increase in stiffness) of the arteries leads to an increase in the PWV. The compliance is a function of several physiological parameters, also partly of the blood pressure – indeed in some conditions, PWV and blood pressure are highly correlated [22].

Computing the PTT requires one proximal and one distal measurement. Proximal options include valve sounds, PPG and even direct arterial blood pressure measurements. The ECG, often being readily available in a monitoring situation, is also used [4]. PTT measured by ECG R-peak, though simple to detect and tolerant of motion artifact, is known not to consistently coincide with the mechanics of the valve movements, and is reported to be an unreliable marker of beat-to-beat blood pressure [23].

The difference in PTT (PTTD) between two distal points can be measured directly by PPG on both points (e.g. toe and finger), which is done in [24] in addition to controlling the values by use of ECG. The relation of PTTD to diastolic blood pressure is investigated but in this case found not to be statistically significant. PTTD is however not ruled out as an assessment of arterial compliance.

2.6   Heart Rate Variability

Short- and long-term variabilities in someone's heart rate as measured from the ECG R-peak reflect properties of the autonomic nervous system. Heart rate variability (HRV) can therefore be used as a monitoring tool in clinical conditions with altered autonomic nervous system function. In postinfarction [25, 26] and diabetic patients, low heart rate variability is associated with an increased risk for sudden cardiac death. Heart rate variability analysis is easily applicable in adult medicine, but physiologic influences such as age must be considered [27].

The significance of HRV has also been examined in sports medicine, both on athletes [29] and on sedentary subjects [30]. A discovered correlation between resting HRV and maximal oxygen uptake (VO_2-max) is applied in some consumer heart rate monitoring devices as an indicator of physical fitness [28]. Although ECG is the preferred method for measuring the HRV, PPG is also applicable.

Chapter 3

Method

All spectrograms were done on normalized but unfiltered signals using a window width of 12 seconds and an 11 second overlap, and shown logarithmically on arbitrary scale.

Pulse frequency estimation was done with the same window length and overlap, by FFT and peak detection, limited to an accuracy of ± 0.1 bpm. A change in detected heart rate of more than 10% in either direction from one data point to the next was considered to be an error, and the value of the second highest peak was used instead. Peak detection was done between 45 and 200 bpm where nothing else is said. See A.2. *Respiration frequency* was estimated using the same method, only with peak detection between 0.055 and 0.75 Hz.

All video material was recorded using a Microsoft Lifecam, a $\sim$500 NOK camera of relatively high quality, at 30 frames per second and 640$\times$360 pixels resolution (unless otherwise specified), further cropped as detailed in each case. See B.3. The subject was the author of this report unless otherwise stated. All experiments carried out were done on several video recordings, the ones reported in ch. 6 being representative examples. One must however be careful not to interpret any result as conclusive evidence.

3.1 Segmenting

In footage without movement, the segmenting was done manually, i.e. the exact area of pixels to be used was found and cut out either in video editing software or directly in Matlab. In chapter 6.1 this is done, one segment containing only the forehead and another containing the whole face.

In footage with movement, face recognition was applied as in [1] (see ch. 2.4). In 6.2.1, the signal is extracted from the face rectangle on each frame. In 6.2.2, it is extracted from a smaller segment of the rectangle which is made sure to contain only the bare skin on the forehead. In 6.2.3, due to the "choppy"[1] nature of the

---

[1]the coordinates returned vary significantly with only small differences between two video frames

face detector, the coordinates are filtered before signal extraction.

Using "pulsatility mapping" (see 3.4 and 6.3) a segment was divided into sub-segments using a grid of squares.

3.2   Signal Extraction

Signal extraction was done, by for each video frame averaging the pixel values of the green channel within the chosen segment. Wherever ICA/JADE was used, this was done also for the red and the blue channel, using all three as input signals for the algorithm. Of the three output signals was chosen the one which contained the most pulse information. For signals extracted from the green channel a capital G is used, for ICA/JADE signals a capital J followed by the number of the channel used, i.e.

$$[J1(t), J2(t), J3(t)] = \mathbf{M}^{-1} [R(t), G(t), B(t)],$$

where M is the resulting mixing matrix.

Where pulsatility mapping was used, signals where extracted individually for each subsegment in a grid overlayed the original larger segment.

3.3   Phase Delay Estimation

An estimation of the phase delay between two pulse signals was done (similarly to the pulse frequency estimation) as follows (see A.4):

1. The two signal $G_a(t)$ and the face signal $G_b(t)$ were both split into N windows of 12 seconds and 11 seconds overlap. Hamming windowing was applied.

2. For each window
$$\{W_a(t; i)\}_{i=1}^N, \qquad \{W_b(t; i)\}_{i=1}^N$$
the index of the peak in the absolute value of its FFT corresponding to the pulse frequency was found, yielding
$$\{I_a(i)\}_{i=1}^N, \qquad \{I_b(i)\}_{i=1}^N.$$

3. The absolute phases at these points were found as
$$\Phi_{aa}(i) = \angle FFT\{W_a(t; I_a(i))\}, \qquad \Phi_{bb}(i) = \angle FFT\{W_b(t; I_b(i))\}.$$
The phases
$$\Phi_{ab}(i) = \angle FFT\{W_a(t; I_b(i))\}, \qquad \Phi_{ba}(i) = \angle FFT\{W_b(t; I_a(i))\}$$
were also found.

4. Two estimates of the phase delay were found,
$$PD_1(i) = \Phi_{aa}(i) - \Phi_{ba}(i), \qquad PD_2(i) = \Phi_{ba}(i) - \Phi_{bb}(i),$$
this to account for the fact that the peaks of the two signals were more often than not offset by a few points, i.e. $I_a(i) \neq I_b(i)$ but $I_a(i) \approx I_b(i)$.

5. Wherever a discontinuity/likely erroneous value was found in the one esti-
   mate, the value of the other was used. Wherever both had discontinuities,
   the previous value was used. Otherwise the average of the two was taken.

6. Result divided by $Fs = 30$ to give values in seconds.

It is pointed out that this is a method of estimating average phase delay, and the
phase delay at the base frequency only. In the case of blood pressure-like signals,
an estimation of peak-point distance may be a more accurate way to a meaningful
value, although it may prove problematic on noisy signals. See further discussion
in ch. 7.5.1.

3.4   Pulsatility Mapping

Matlab scripts were written which create a map of the relative values of the pul-
satilities – within each square in a grid segmenting the video frame. The procedure
used was as follows (see also A.3):

1. Create grid dividing video frame into N squares. Square size is chosen by ex-
   periment in each case to give a balance between resolution and signal strength.

2. For each frame and for each square the average pixel value in the green
   channel of the video is calculated to give the set of signals $\{g_i(t)\}_{i=1}^{N}$.

3. For all $i = 1, ..., N$, $G_i(f) = FFT\{g_i(t)\}$ is calculated. The indices $f_{Gmax}(i)$
   and values $G_{max}(i)$ of the maximum in the interval $(45, 200)$ [bpm] are ex-
   tracted.

4. The pulse rate is assumed to be the most frequently occurring value i.e. the
   mode of $f_{Gmax}(i)$. The squares corresponding to the i's for which $f_{Gmax}(i) \neq$
   $mode\{f_{Gmax}(i)\}$ are excluded.

5. The colormap $[red, yellow, green, blue]$ is assigned to $G_{max}(i)$ using
   $[3, 2, 1, 0.5] \cdot std\{G_{max}(i)\}$ as lower thresholds, accordingly. Maps are shown
   overlayed an averaged video frame indicating the slight movements of the
   subject.

A new PPG signal is constructed by averaging the content of a combination of the
red, yellow, green and blue sets. What sets to use was determined by experiment
to see what combinations gave the most pleasing result in each case. It is pointed
out that the thresholds for the different pulsatility levels are based on what levels
exist in the given video, and do not represent an absolute scale.

Repeating the procedure for $\{r_i(t)\}_{i=1}^{N}$ (R channel) was tested as a means of re-
moving false positives caused by movement/mechanical pulsations, but was deemed
unnecessary as only very few were found.

3.5   Short-Time Heart Rate Estimation

(i) Heart rate estimation using forehead segment   A simple Matlab script was written executing the following procedure. A subject is assumed to be situated with face directed towards the camera in normal computer interaction distance, i.e. about 40 cm.

1. Construct webcam video object with resolution 640×360 pixels, color space 24-bit RGB and frame rate 10 fps.

2. Capture single frame and perform face recognition [20]. As ROI choose top, middle 96×32 pixels from the rectangle returned.

3. Record L seconds of this ROI. Average contents of green channel for each frame.

4. Calculate PSD and detect its peak in the interval [45, 200] bpm.

(ii) Heart rate estimation using pulsatility mapping   A second Matlab script was written executing the following procedure. Subject is situated as before.

1. Construct webcam video object with resolution 640×360 pixels, color space 24-bit RGB and frame rate 10 fps.

2. Capture single frame and perform face recognition [20]. The height of the returned rectangle is increased by a factor of 1.5 to make sure all of the subject's face is contained within it. It is also made sure to be divisible by d in both directions.

3. Record L seconds of this ROI. Create grid and within each $d×d$ square average the contents of the green channel.

4. Calculate PSD for each square and detect its peak in the interval [45, 200] bpm.

5. Detected pulse is taken to be the most frequently occurring peak index (with a few exceptions, see A.5).

Chapter 4

Preliminary

4.1 Exploring Contents of the ambient-light PPG

4.1.1 Forehead

From a video of about two minutes' length of the subject keeping his face still
were cut two segments, (i) a segment on the forehead of 93×54 pixels and (ii) a
segment of the same size on the subject's t-shirt. These segments were averaged
over each frame, both resulting in separate signals for the red, green and blue
channels. Fig. 4.1 shows (in their respective colors) the red, green and blue signals
(here shown above each other without regard to amplitude/scale) and their FFT
contents between 0.1 and 1.1 Hz for segment (i). The green and blue signals both
have peaks at around 0.15 and 1 Hz, but the green peaks are seen to be significantly
higher. Fig. 4.2 shows the same information for segment (ii), all three signals with
a peak at around 0.2 Hz. In fig. 4.3 an example view of the relative R, G, B
absolute pixel values of segment (i) are seen (segment (ii) has values that do not
differ significantly between the channels).
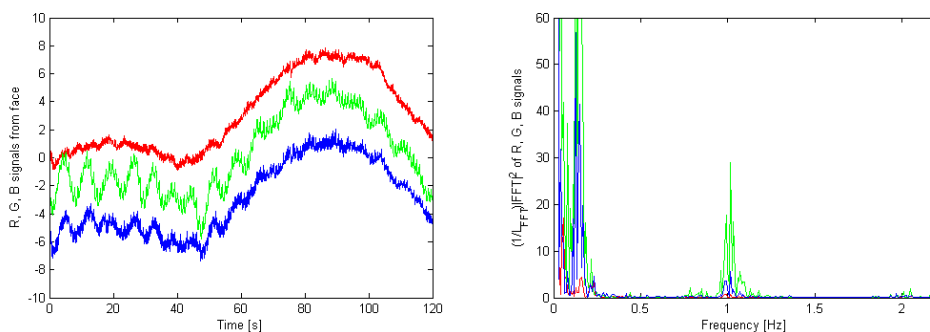
Figure 4.1: Red, green and blue channels, segment from forehead (i)

15

A view of the green signal from fig. 4.1 is given in fig. 4.4 where it is split into
four frequency bands delimited at 0.055, 0.75 and 3.75 Hz. (Bands hypothesized
to be influenced by *a) external, b) respiration, c) blood pressure/heart rate, d)
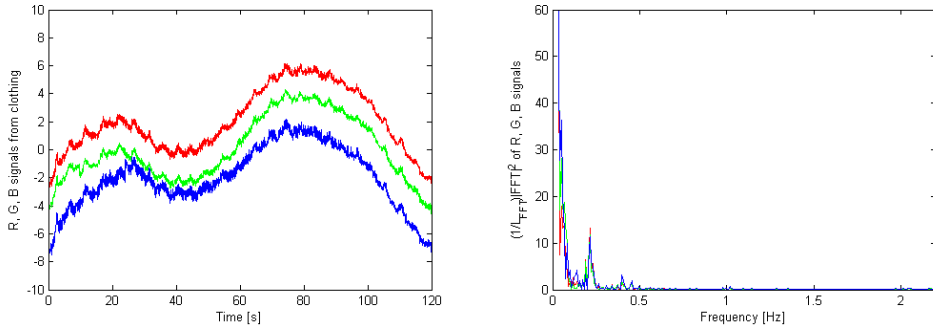HF/noise.*) The spectrogram for this same green signal can be seen in fig. 6.2(a).

Figure 4.2: Red, green and blue channels, segment from subject's clothing (ii)
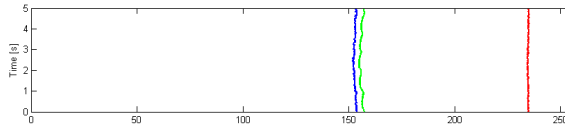

Figure 4.3: Amplitudes of R, G and B channels, segment (i), to scale. Channel bit
depth of 8 gives pixel values in the interval [0, 255].
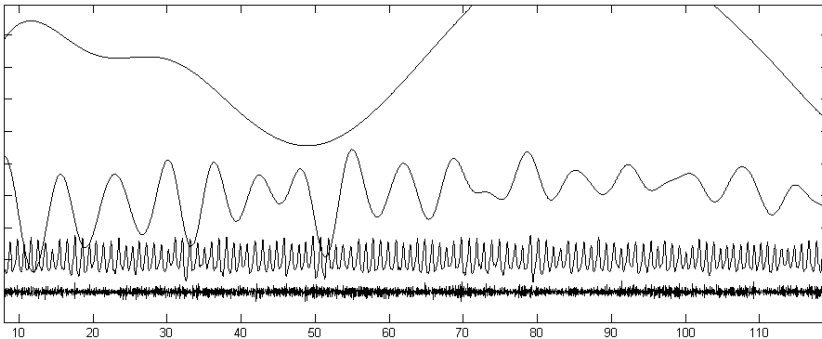
Figure 4.4: Forehead PPG signal in frequency bands (from top to bottom): a, b,
c, d.

4.1.2 Palm

From 136×196 pixels closeup video footage of the subject's palm were extracted RGB signals. The green channel's spectrogram is shown in fig. 4.5 and its division into frequency bands in fig. 4.6.
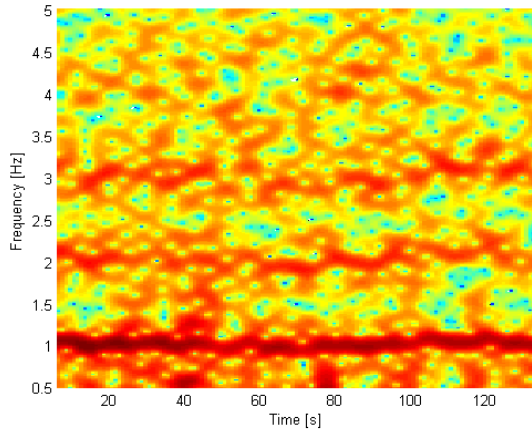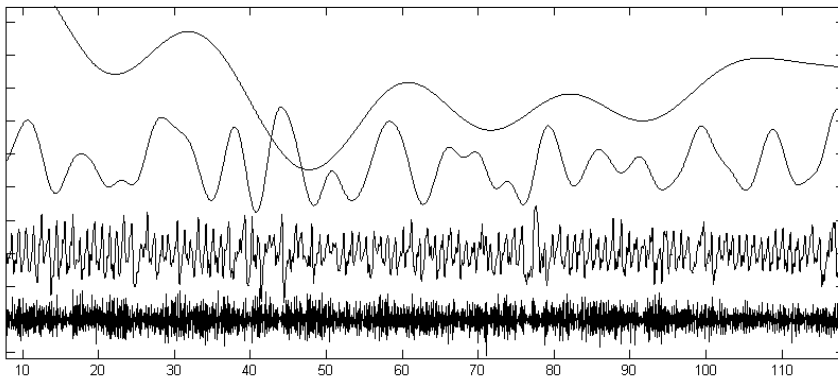

Figure 4.5: Spectrogram from palm

Figure 4.6: Palm PPG signal in frequency bands

4.2   ICA-JADE Test Case

We synthesize a signal consisting of an ABP-like AC component with varying heart rate of 1.1-1.3 Hz, baseline modulated by a respiration-like part at 0.11-0.13 Hz. See spectrogram in fig. 4.7 and waveform $s_1(t)$ in fig. 4.8. Together with the two low pass noise signals $s_2(t)$ and $s_3(t)$ it was linearly mixed using a mixing matrix

$$\mathbf{M} = \begin{bmatrix} 0.004 & -7 & 3 \\ 0.04 & 2 & 4 \\ 0.01 & 6 & 6 \end{bmatrix}$$

to give the signals $x_1(t)$, $x_2(t)$ and $x_3(t)$ according to eq. 2.1 on page 8. We see in fig. 4.9 that the synthesized signal has been essentialy drowned in noise because of the lower values in the left column of M.
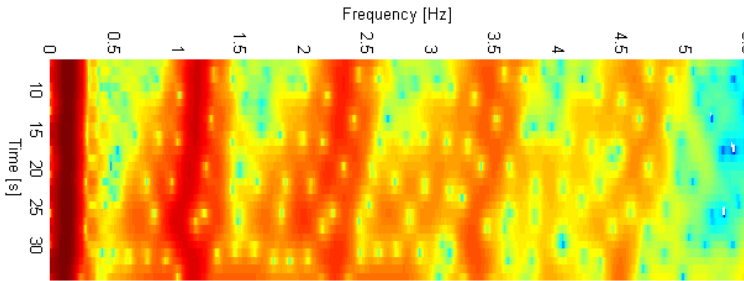

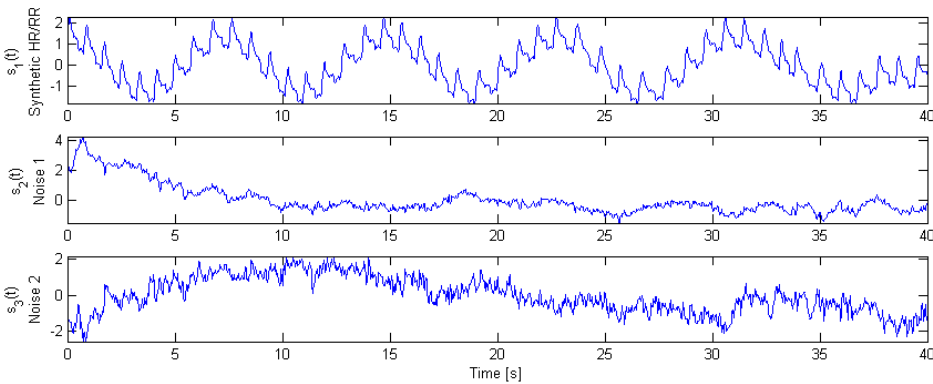Figure 4.7: Spectrogram of synthesized heart/respiration signal

Figure 4.8: Original signals

The signals $\varsigma_1(t)$, $\varsigma_2(t)$ and $\varsigma_3(t)$ (fig. 4.10) were recovered using JADE according to

$$[\varsigma_2(t), \varsigma_1(t), \varsigma_3(t)] = \mathbf{M}_\varsigma^{-1} \cdot [x_1(t), x_2(t), x_3(t)]$$

with

$$\mathbf{M}_{\varsigma} = \begin{bmatrix} 7.6323 & 1.5394 & 0.0802 \\ -0.1296 & 0.0230 & 4.3537 \\ -2.9090 & -0.5603 & 7.6536 \end{bmatrix}.$$
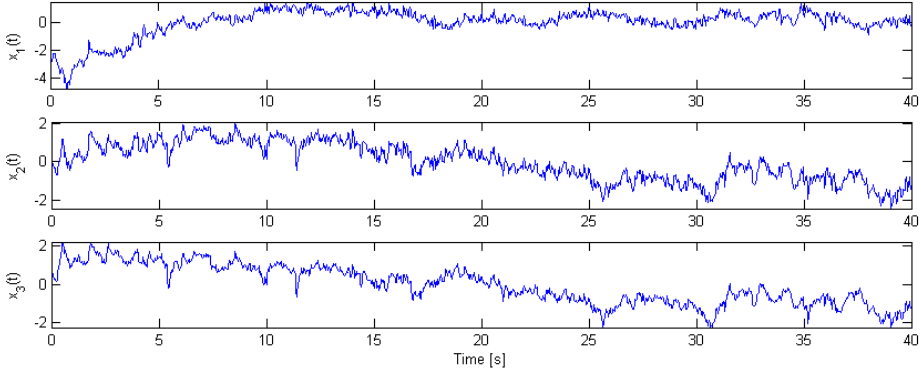
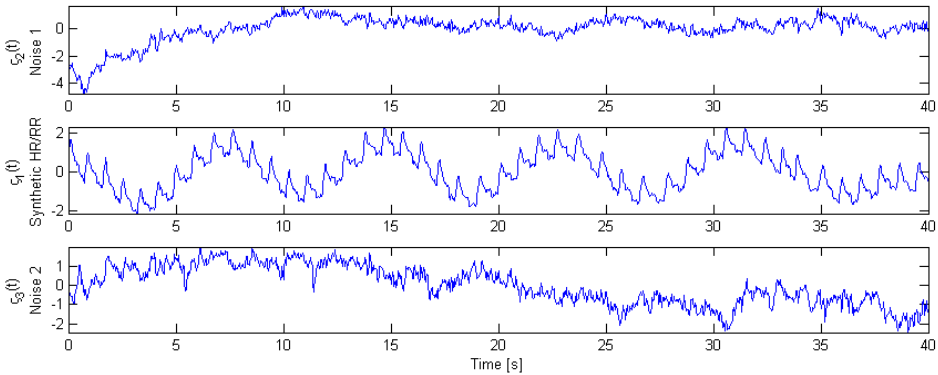Figure 4.9: Originals signals mixed – input to JADE.

Figure 4.10: Output from JADE.

At the scale shown here, the recovered signals are very much alike the original ones. Although the numbers in the mixing matrix are not exactly the same (and the column order has been changed), we recognize that the middle column of $\mathbf{M}_{\varsigma}$, corresponding to the pulse signal, contains lower values than the others, as in $\mathbf{M}$.

Chapter 5

Experiments

1. From video footage where the subject is not moving, a pulsatile signal is extracted in different ways:

    (a) from a segment containing the entire face of the subject

    (b) from a segment containing only the bare skin on the forehead of the subject

    (c) by applying JADE source separation to (a)

    The results are compared visually by spectrograms and by their ability to track the heart rate as compared to a finger pulse oximeter, in **ch. 6.1**, and discussed in **ch. 7.2**.

2. From video footage where the subject is moving, a pulsatile signal is extracted in different ways:

    (a) from a segment containing the entire face of the subject, as found by face detection

    (b) from a segment containing only the bare skin on the forehead of the subject, i.e. the top, middle segment cut out from (a)

    (c) by applying JADE source separation to (a)

    (d) by applying JADE source separation to (b)

    (e) (a)-(d) repeated with filtered face coordinates

    The results are compared visually by spectrograms and by their ability to track the heart rate, in **ch. 6.2**, and discussed in **ch. 7.3**.

3. In **ch. 6.3 and 6.5** a method of pulsatility mapping is tried out as a means of (i) mapping the distribution of the pulsatile signal, (ii) semi-automatically extracting the pulsatile signal and (iii) estimating heart rate. This is discussed in **ch. 7.4**.

4. In **ch. 6.4**, experiments are done on a recording where parts of both the hand and the face of the subject are visible, efforts are made to extract such information as heart rates, respiratory rates and phase delay. Discussed in **ch. 7.5**.

5. In **ch. 6.5**, efforts are made to find out what duration of video is needed for a stable heart rate estimation, using the methods of (i) forehead segment and (ii) pulsatility mapping. Method (ii) is then subject to a somewhat larger test on 6 subjects to confirm the stability. Discussed in **ch. 7.6**.

6. In **ch. 6.6**, a brief analysis is done of a short laparoscopy video recording of a liver. The JADE algorithm is applied on four separate ROI's, one of which contains a cancerous growth. Discussed in **ch. 7.7**.

Chapter 6

Results

6.1 Heart Rate Estimation from Footage without Movement

6.1.1 Segment on Forehead

About 120 seconds of footage was recorded where the subject was keeping his face still. A segment of $93{\times}54$ pixels on the forehead was cut out and $G_{forehead}(t)$ extracted. The spectrogram of $G_{forehead}(t)$ is shown in fig. 6.2(a). A clip of a few seconds of the waveform is shown in fig. 6.1.
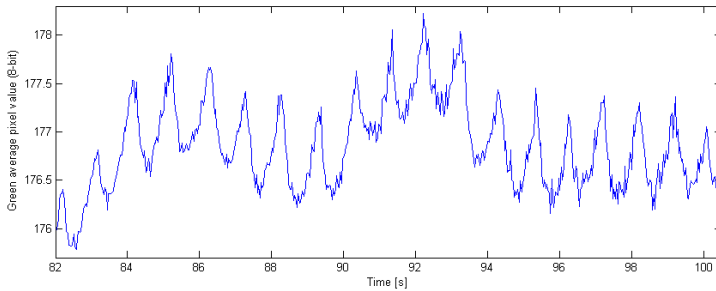

Figure 6.1: Cut from $G_{forehead}(t)$

6.1.2 Segment containing whole Face

A new segment of $158{\times}212$ pixels was cut from the same video containing the whole face. $G_{face}(t)$ was extracted. Its spectrogram is shown in fig. 6.2(b), and a clip of the waveform in fig. 6.3.

$R_{face}(t)$ and $B_{face}(t)$ were also extracted and together the three were used as input for a JADE ICA source separation (see 3.2 and 2.3.1). Of the output

(a) $G_{forehead}(t)$ (b) $G_{face}(t)$
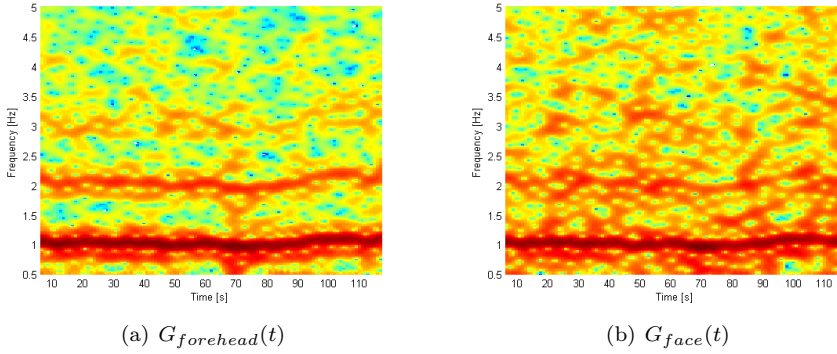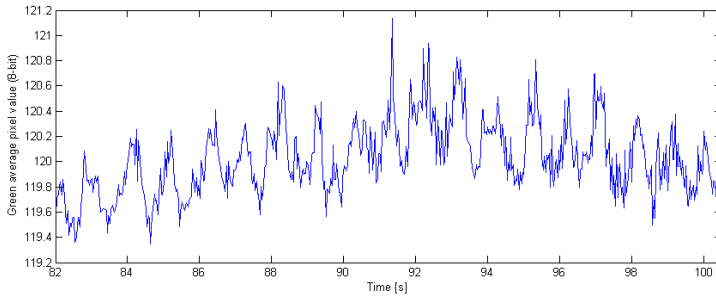
Figure 6.2: Spectrograms

Figure 6.3: Cut from $G_{face}(t)$

signals $J1_{face}(t)$ was the one containing the most pulsatile information, and its spectrogram is shown in fig. 6.4. The mixing matrix found was

$$\mathbf{M}_{face} = \begin{bmatrix} 0.9147 & 0.8269 & 5.6251 \\ 0.0891 & 0.9296 & 3.9579 \\ 0.0950 & 0.4633 & 4.0161 \end{bmatrix}$$

Heart rate estimations were performed and are shown in fig. 6.5 for the pulse oximeter, $G_{forehead}(t)$, $G_{face}(t)$ and $J1_{face}(t)$. Error is quantified in tab. 6.1, where values are listed for average heart rate (mean), maximum absolute error ($e_{max,abs}$) and the standard deviation of the error (σ_e), when results from pulse oximeter are taken as correct.
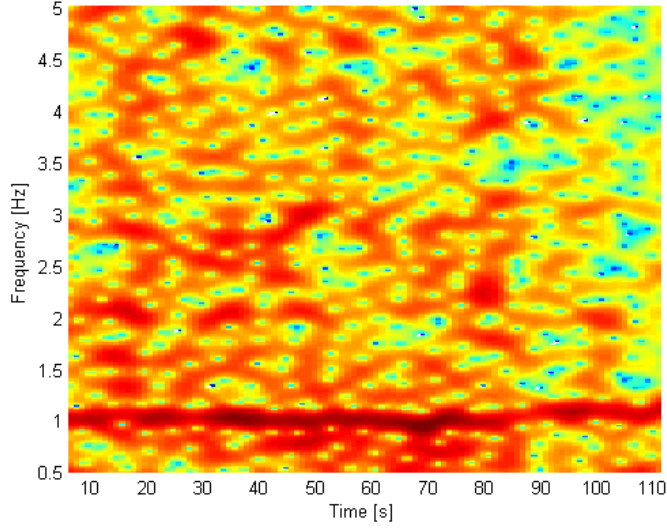
Figure 6.4: Spectrogram of $J1_{face}(t)$

Table 6.1: Error in estimated pulse frequency, video without movement

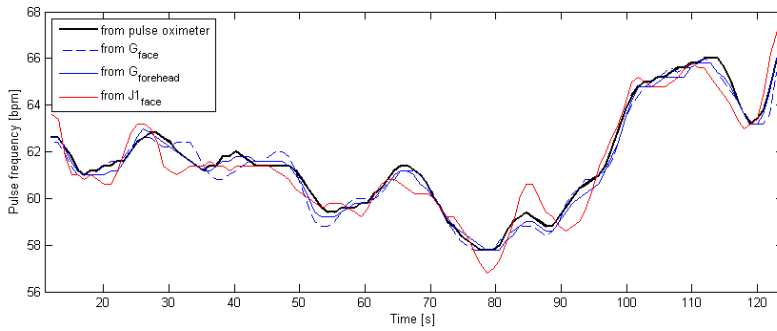| Signal	Mean [bpm]	$e_{max,abs}$ [bpm]	σ_e [bpm]
Pulse oximeter	61.6	0	0
$G_{forehead}(t)$	61.4	0.6	0.20
$G_{face}(t)$	61.4	1.2	0.36
$J1_{face}(t)$	61.4	1.4	0.61

Figure 6.5: Estimated pulse frequency compared for the different signals

6.2   Heart Rate Estimation from Footage with Movement

6.2.1   Segment containing whole Face

A video of 100 seconds was recorded with the subject moving around as if he were interacting normally with the computer, i.e. no abrupt movement and face always directed towards the camera. Face detection was carried out as described in 2.4. Traces of resulting face position are given in fig. 6.6, the blue trace (X-position) here quantifying the horizontal movement. For comparison a simultaneous recording was done with the finger pulse oximeter.
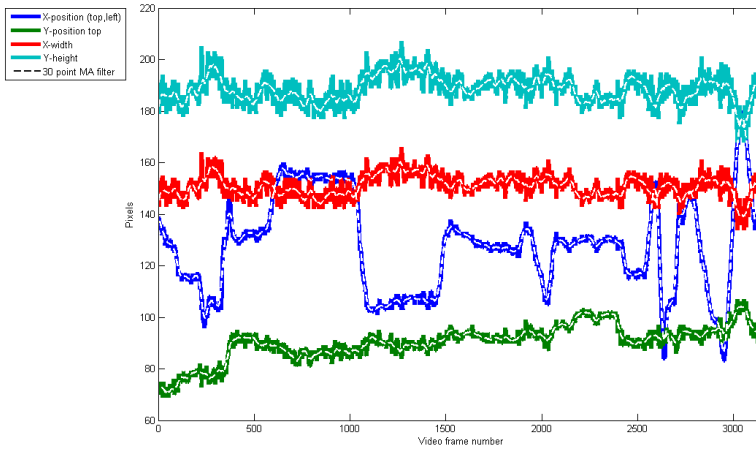

Figure 6.6: Result of OpenCV face recognition on 360×360 pixels video

The raw trace from the green channel, $G_{recface}$, was extracted as before, now from the segment returned from the face recognition procedure. Its spectrogram is shown in fig. 6.7(a). With its corresponding red and blue channels it was then used as input to the JADE algorithm, the result signal containing the most pulsatile information was $J2_{recface}$, whose spectrogram is shown in fig. 6.12(a). The mixing matrix was

$$\mathbf{M}_{recface} = \begin{bmatrix} 4.2975 & 1.0157 & 0.0333 \\ 1.6173 & 1.2284 & -0.0513 \\ 1.9027 & 0.7555 & 0.7371 \end{bmatrix}.$$

The heart rate was estimated from the two signals and are compared to the heart rate estimated from the finger pulse oximeter in fig. 6.8. Minimum allowed frequency for the pulse was raised to 52 bpm for $G_{recface}$ when low frequency disturbances were seen. Error data is compared in tab. 6.2
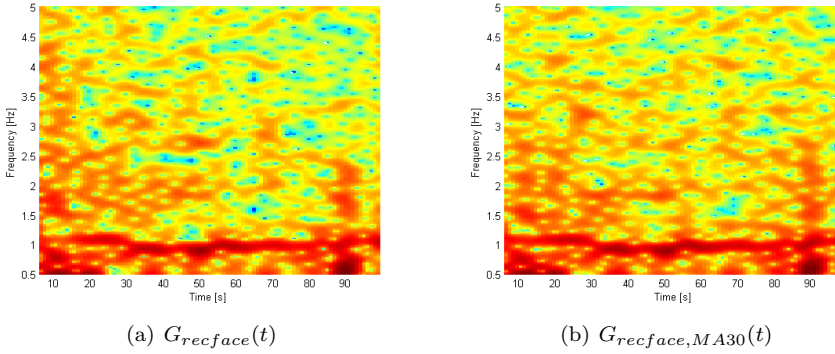
(a) $G_{recface}(t)$

(b) $G_{recface,MA30}(t)$

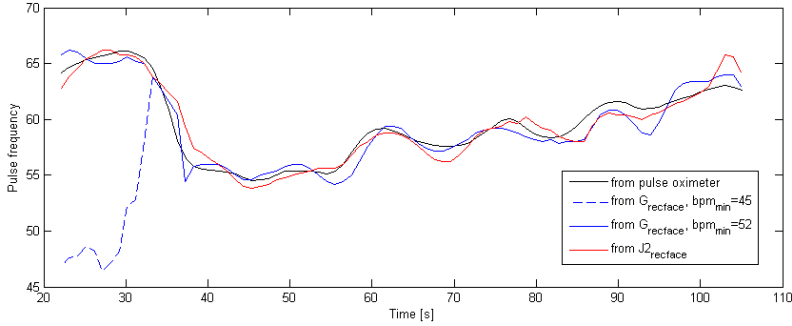Figure 6.7: Spectrograms

Figure 6.8: Estimated pulse from face detection

6.2.2 Segment on Forehead

Based on the rectangles returned from the face recognition procedure, a segment of 64×32 pixels was extracted from the top middle part of the face, i.e. a segment containing a part of the forehead, moving with the face throughout the video. The signals $G_{recforehead}$ and $J2_{recforehead}$ were extracted in the same manner as before, directly from the green channel and by JADE separation, respectively. The spectrogram from $G_{recforehead}$ is shown in fig. 6.9(a) and from $J2_{recforehead}$ in fig. (excluded, see 7.3).

Heart rate was estimated also from these signals, result shown in fig. 6.10 in comparison with pulse oximeter. Here, too, minimum pulse frequency was raised to 52 bpm, in this case for the $G_{recforehead}$ signal.
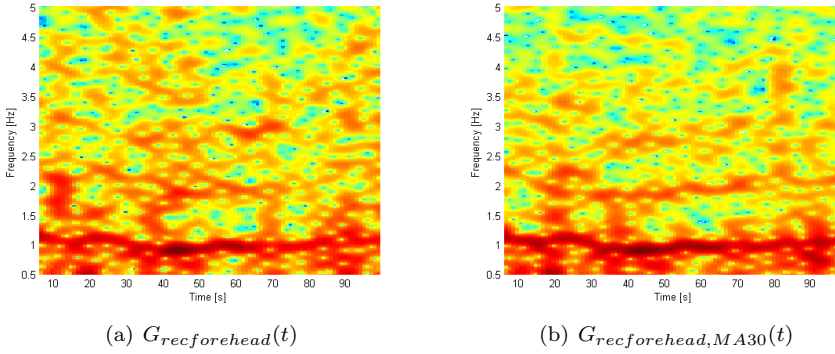
(a) $G_{recforehead}(t)$ (b) $G_{recforehead,MA30}(t)$
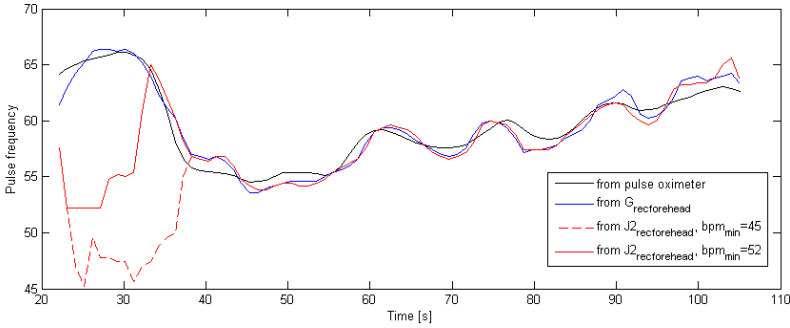
Figure 6.9: Spectrograms

Figure 6.10: Estimated pulse from face detection, segment on forehead

6.2.3   Smoothing out Face Coordinates

To smooth out the movement from frame to frame of the face detection rectangle, its coordinates were filtered in with a 30-frame (1 second) moving average (MA) filter. New coordinates are shown in fig. 6.12(a). Signal extraction was repeated, giving the signals $G_{recface,MA30}$ (fig. 6.7(b)), $G_{recforehead,MA30}$ (fig. 6.9(b)) and $J2_{recface,MA30}$ (fig. 6.12(b)). BPM values for these signals are shown in fig. 6.11. Error data is compared in tab. 6.2, where values are listed for average heart rate (mean), maximum absolute error ($e_{max,abs}$) and the standard deviation of the error (σ_e).
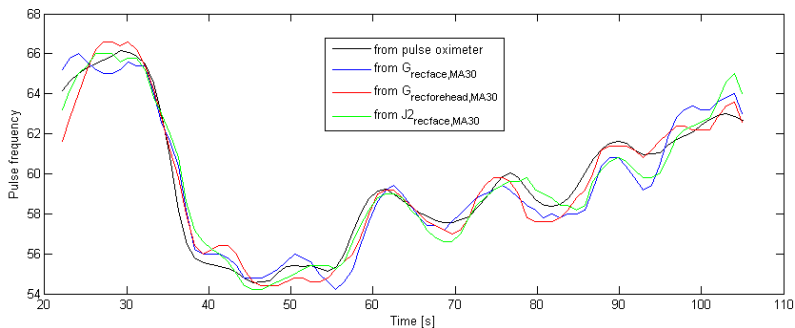
Figure 6.11: Estimated pulse when face detection coordinates are filtered

Table 6.2: Error in computed pulse frequency, video with movement

Signal	Mean [bpm]	$e_{max,abs}$ [bpm]	σ_e [bpm]
Pulse oximeter	59.6	0	0
$G_{recface}(t)$	59.4	2.4	0.95
$J2_{recface}(t)$	59.5	3.5	1.0
$G_{recforehead}(t)$	59.6	2.8	0.95
$G_{recface,MA30}(t)$	59.5	2.1	0.68
$J2_{recface,MA30}(t)$	59.4	3.0	0.80
$G_{recforehead,MA30}(t)$	59.5	2.1	0.71
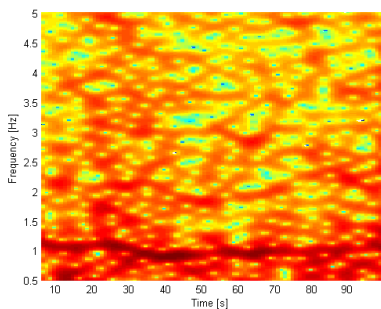
(a) $J2_{recface}(t)$
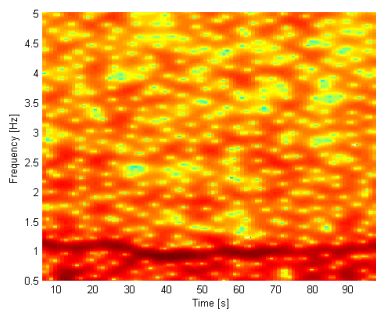
(b) $J2_{recface,MA30}(t)$

Figure 6.12: Spectrograms

6.3   Pulsatility Mapping

6.3.1   Inside of Hand

The procedure described in 3.4 was carried out on a two minute video of the subject's palm using a grid of 8×8 pixel squares. The resulting pulsatility map is shown in fig. 6.13, and a new signal was constructed by averaging the content of the red squares. The spectrogram of this signal is shown in fig. 6.14(a) compared to the spectrogram of a signal that was extracted manually from a rectangular segment of 136×96 pixels in fig. 6.14(b).

Figure 6.13: Hand pulsatility map

(a) Signal made from summing pulsatile areas

(b) Signal from rectangular segment

Figure 6.14: Spectrograms hand

6.3.2 Face

The procedure was done similarly for a two minute face video (same video as in ch. 6.1.1), using 4×4 pixels grid size, result shown in fig. 6.15. Spectrograms are compared in the same manner as above in fig. 6.18 (using the forehead segment of 93×54 pixels from fig. 6.2(a)). The signal in 6.16(a) was taken this time from the red, green and yellow squares.
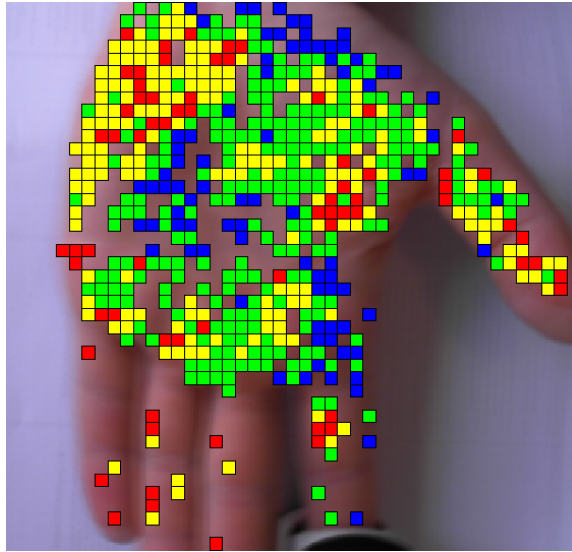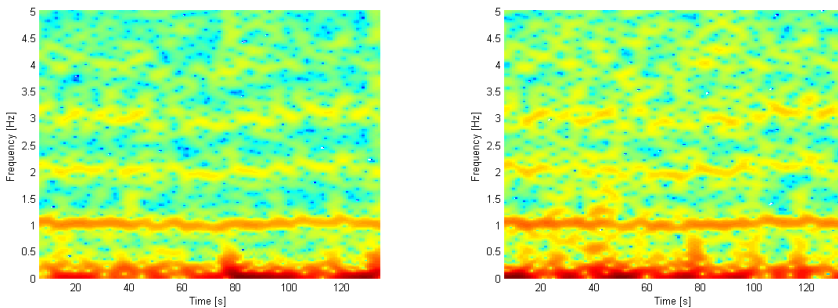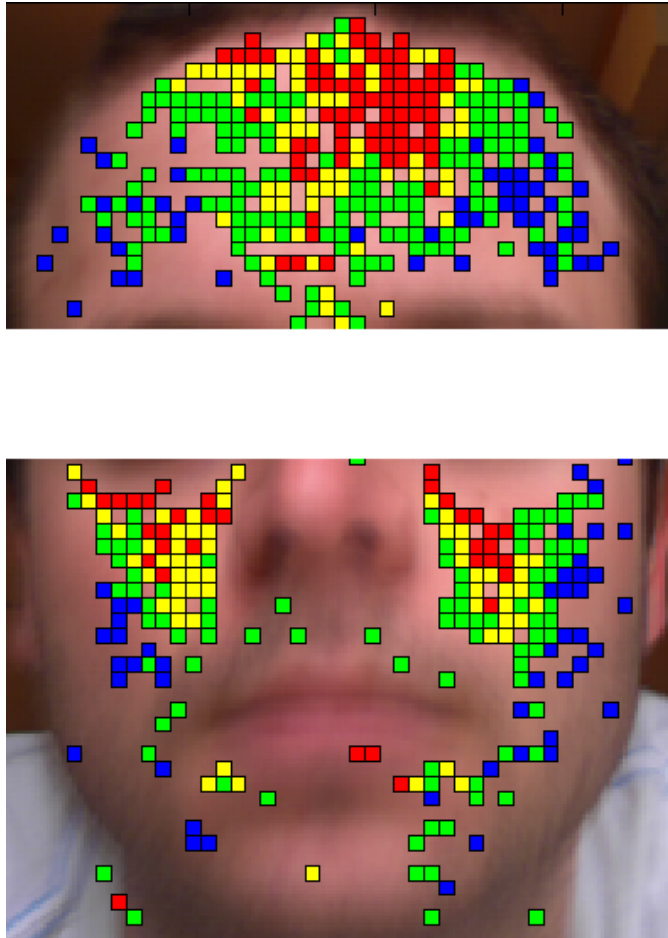
Figure 6.15: Face pulsatility map

(a) Signal made from summing pulsatile areas   (b) Signal from rectangular segment
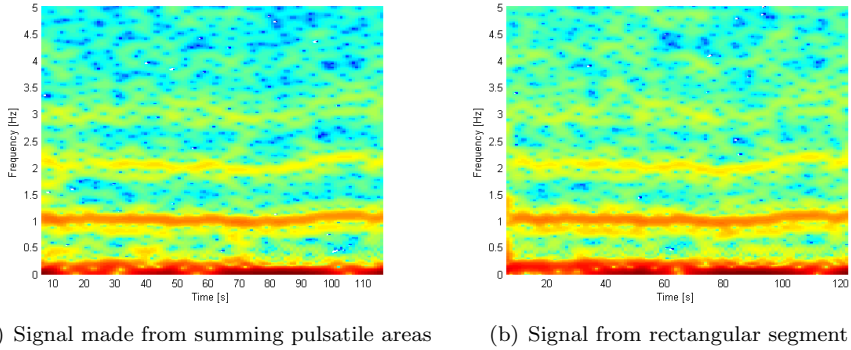
Figure 6.16: Spectrograms face

6.3.3   Face and Hand combined

A video was recorded in which parts of both face and hand were visible, and pulsatility mapping done on an 8×8 pixel grid, result shown in fig. 6.17. It is again pointed out that the color scale used is not absolute – we observe lower levels on the inside of the hand than on the face, not evident from figs. 6.13 and 6.15.
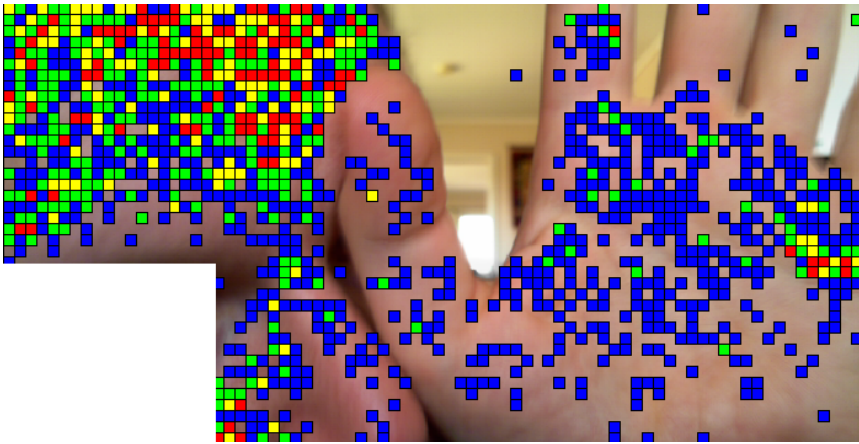
Figure 6.17: Pulsatility map for both face and hand in one frame

6.4 Face/Hand Comparisons

Extending upon the previous section 6.3.3, the video frame was manually separated into one hand segment and one face segment, and signals were averaged over the pulsatile squares within each segment. Spectrograms of these signals are shown alongside spectrograms of signals averaged over manually chosen rectangle segments.
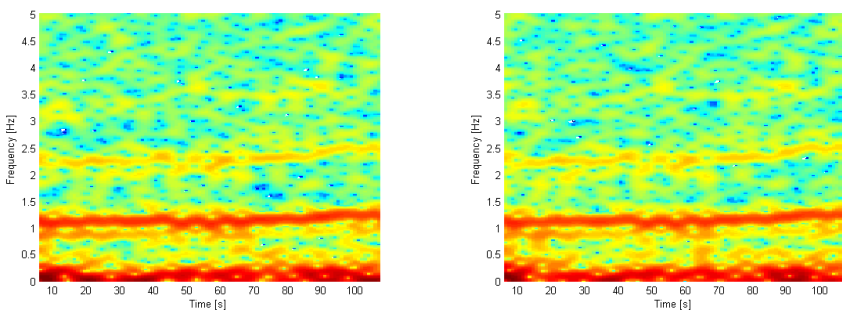
(a) Signal made from summing pulsatile areas   (b) Signal from rectangular segment

Figure 6.18: Spectrograms face
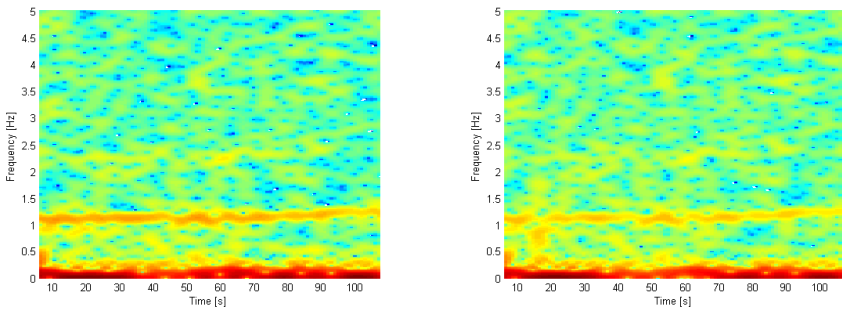
(a) Signal made from summing pulsatile areas   (b) Signal from rectangular segment

Figure 6.19: Spectrograms hand

6.4.1   Phase Delay Estimation
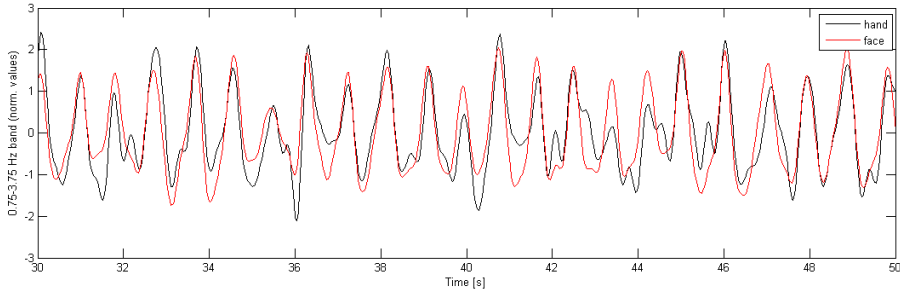
Figure 6.20: Comparing heart rate pulsations

In fig. 6.20 is shown a small clip of the signals from figs. 6.19(a) and 6.18(a) after being bandpass filtered to reveal HR contents and their harmonics in the frequency band [0.75, 3.75] Hz. The HR frequencies were estimated for both signals, shown in fig. 6.22(a). The phase delay between them was estimated as described in 3.3 and is shown in fig. 6.22(c). This phase delay is equivalent to the PTTD for the hand and the face, see Background (ch. 2.5) and Discussion (ch. 7.5.1).

6.4.2   Respiration Rate Estimation
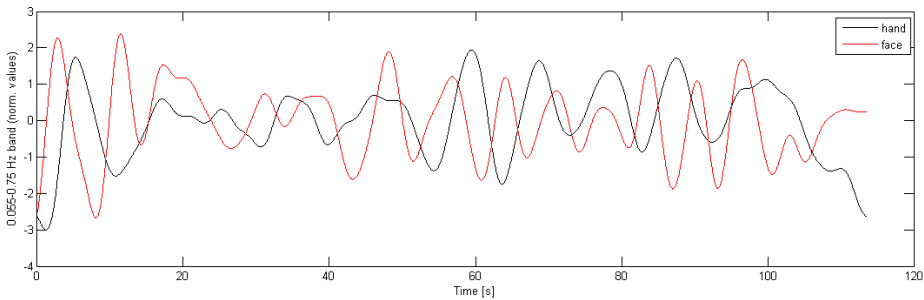
Figure 6.21: Contents of the face and hand signals in the frequency band [0.055, 0.75]

Fig. 6.21 shows the contents of the two signals in the band [0.055, 0.75] Hz, assumed to be conditioned by respiratory variations. Values for their frequencies were estimated and are shown in fig. 6.22(c).
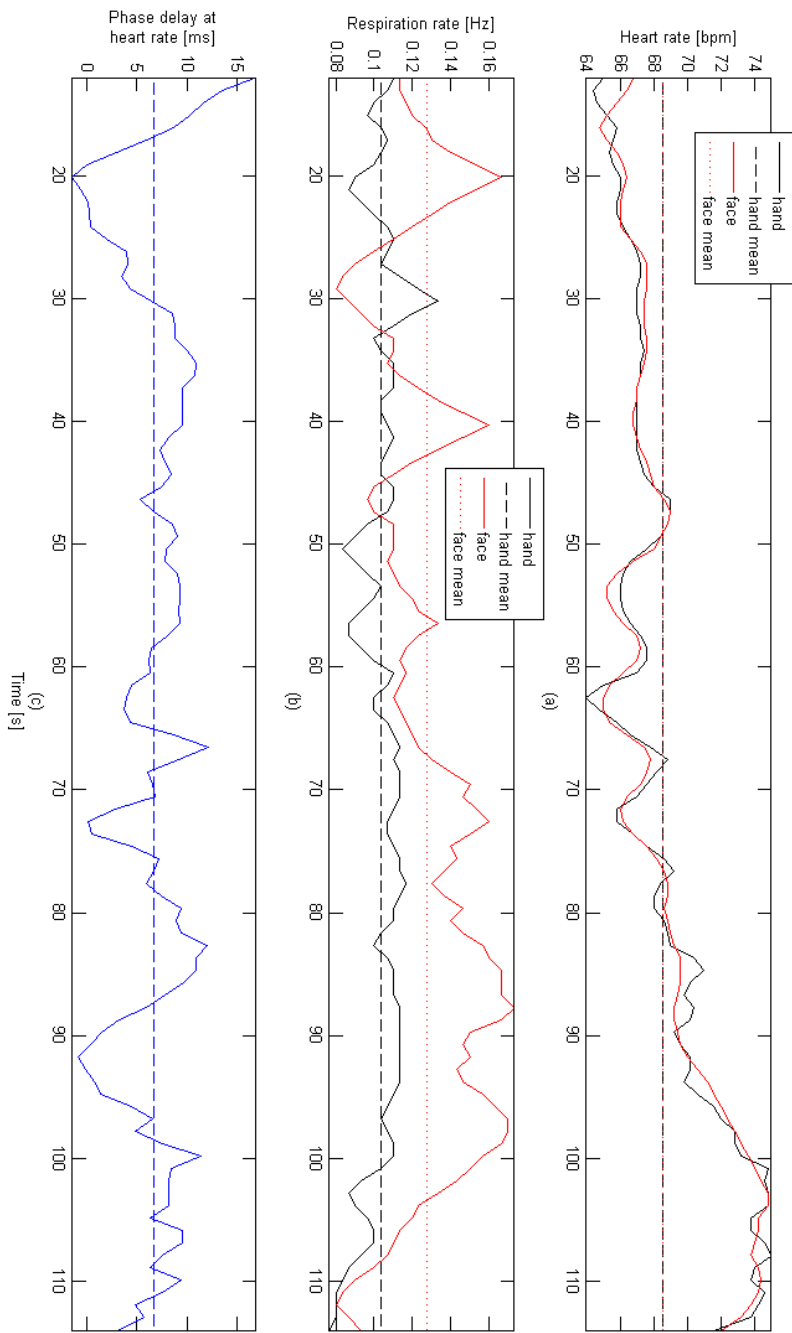
Figure 6.22: (a) estimated heart rates, face and hand, (b) estimated respiration rates, face and hand, (c) estimated phase delay, hand on average 6.6 ms after face

6.5   Short-Time Heart Rate Estimation Testing

(i) Testing   For the application example described in 3.5(i), a single-subject test was carried out for a suggestion of the necessary length L of the recording. For each run the value found was noted as well as the instantaneous value from the finger pulse oximeter. The application was run 24 times at each value of L, and values were calculated for RMSE e_{rms}, bias e_{avg} and maximum absolute error $e_{max,abs}$. Results are listed in tab. 6.3.

Table 6.3: Error results, short-time heart rate estimations (i)

| L [s]	e_{rms} [bpm]	e_{avg} [bpm]	$e_{max,abs}$ [bpm]
4	6.6	-3.8	23
5	6.1	-2.5	15
6	4.2	-2.3	10
7	2.2	-1.1	4.0

(ii) Testing 1   For the application example described in 3.5(ii), a single-subject test was carried out for a suggestion of the necessary length L of the recording. For each run the value found was noted as well as the instantaneous value from the finger pulse oximeter. The application was run 24 times at each value of L, and values were calculated for RMSE e_{rms}, bias e_{avg} and maximum absolute error $e_{max,abs}$. Results are listed in tab. 6.3. The value $d = 10$ was used.

Table 6.4: Error results, short-time heart rate estimations (ii)

L [s]	e_{rms} [bpm]	e_{avg} [bpm]	$e_{max,abs}$ [bpm]
4	7.9	-3.5	26
5	14	1.1	67
6	4.6	0.042	18
7	2.3	-0.92	4.0

(ii) Testing 2   It was chosen $L = 7$ and a test of 24 runs of the script was performed on 6 subjects, a total of 144 data points. For each run the value found was noted (p_{webcam}) as well as the instantaneous value from the finger pulse oximeter (p_{poxi}). The results are shown in a Bland-Altman plot in fig. 6.23. The overall RMSE was 4.95 and the bias -0.77 bpm. Examples of face pulsatility maps from the six subjects are found in B.1.
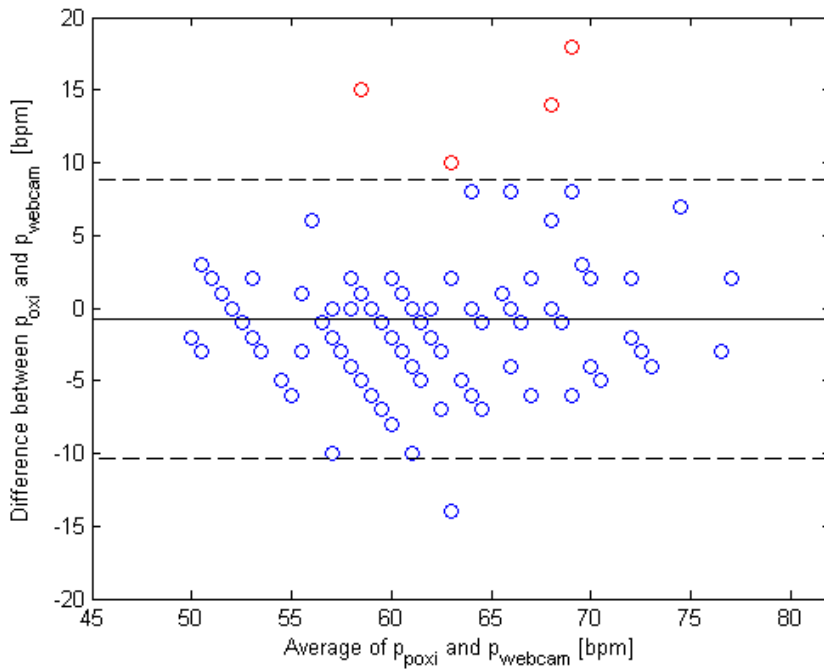
Figure 6.23: Average vs. difference of estimated and control value, short-time heart rate estimation testing. By convention of Bland-Altman plots, $1.96 \cdot SD$ and $-1.96 \cdot SD$ limits are indicated. Data points outside these are in red. Note that a large amount of data points coincide with each other due to integer rounding.

6.6   Laparoscopy Video

On a short clip (7 seconds, the longest available with reasonably low movement) from a liver laparoscopy video were chosen 4 separate ROI's, two on the wall of the abdominal cavity (1, 2) and two on the liver (3, 4) whereof one (4) containing a cancerous growth (see fig. 6.24). The R, G and B raw traces were extracted for each ROI and fed through the JADE algorithm. Normalized RGB traces and the most pulsatile of the independent components are shown in the following figures with their respective FFT's.
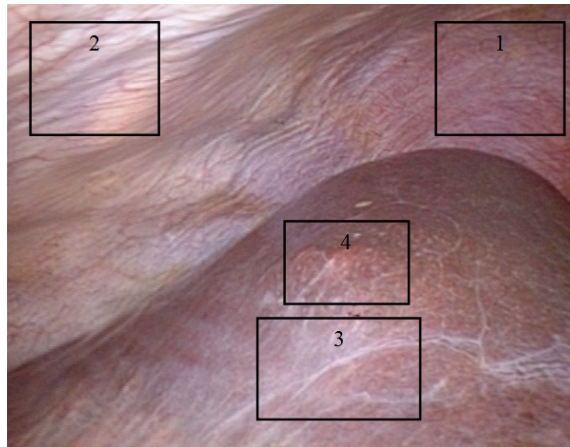
Figure 6.24: Video frame with the 4 ROI's
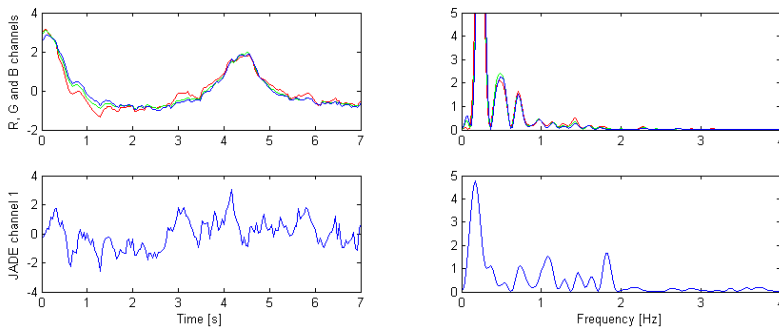
Figure 6.25: Laparoscopy ROI 1
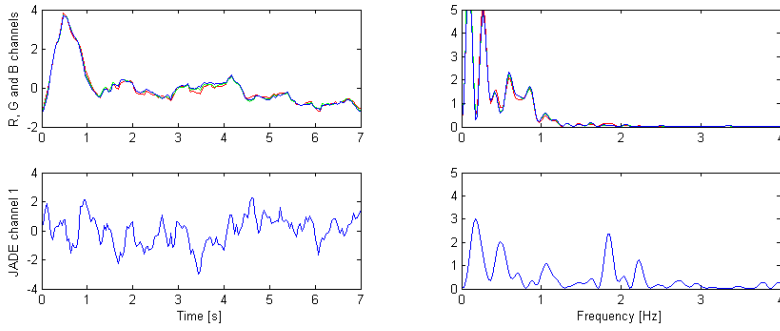
Figure 6.26: Laparoscopy ROI 2
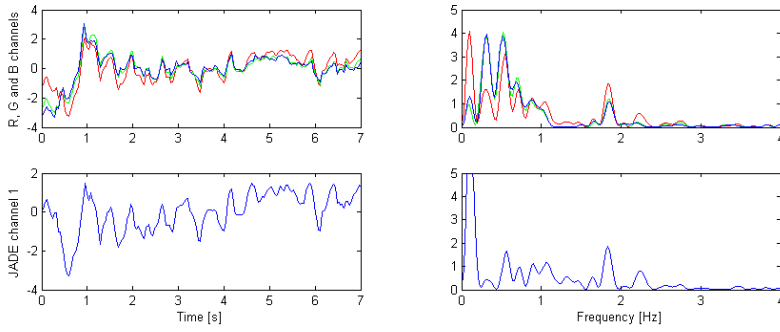
Figure 6.27: Laparoscopy ROI 3
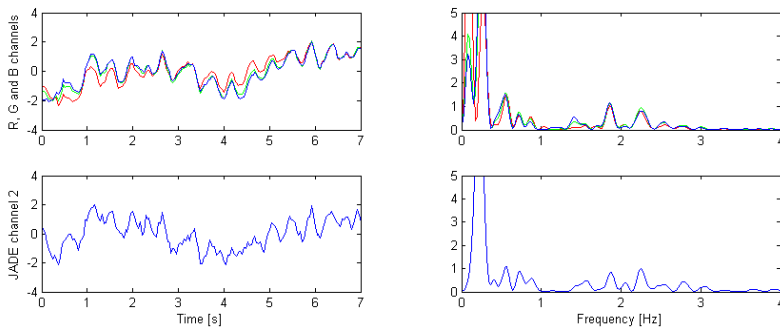
Figure 6.28: Laparoscopy ROI 4

Chapter 7

Discussion

The procedures in this report are focussed on not requiring expensive equipment, e.g. no especially high quality video recording is used. Wherever something is stated to be infeasible, it may not mean that it is impossible to do so, it means, however, that the wanted results may not be readily available at the technologically low threshold wanted. Better lighting and higher quality optics and image sensors will likely yield better results, but when an expressed goal is that a normal webcam should suffice, a procedure is disproven by one of them failing to do so. On the other hand, that something is possible to do with the camera at hand, does not mean that it is possible with any camera. In this regard, any exploration on the topic of this report will be incomplete, but will likely give a very good indication of what is possible with an average web camera.

7.1 The Contents of the Webcam PPG

Discussed here are findings from preliminary explorations in 4. What is evident at first glance from fig. 4.1(b) is the protruding FFT peak at around 1 Hz in the green, to a lesser extent also visible in the blue channel, which it is reasonable to assume is the base harmonic of the heart rate frequency.

Further, it is observed comparing figs. 4.1(a) and 4.2(a) that there are common baseline fluctuations in what is revealed to be the frequency band below 0.055 Hz. The fact that these are found both on bare skin and on clothing makes us conclude that they are related to movement and external lighting but possibly also partly to in-camera adjustments we were not able to control manually.

The peaks in the green and blue channels in fig. 4.1(b) at 0.15 Hz, that are also clearly visible as variations with a period of about 7 seconds in fig. 4.1(a), are assumed to be indicative of respiration rate, based on argumentation and explorations in [2] (see 2.2). That the content found at both 0.15 and 1 Hz is indeed true PPG content stemming from physiological parameters is further backed by the relative pixel values of the R, G and B channels, an argument put forth by [2]. We see in fig. 4.3 that green and blue values are significantly lower than red.

A peak is seen in all three channels from segment (ii) (fig. 4.2(b)) at 0.2 Hz. This is distinct from respiration rate at 0.15 Hz and is likely to come from movement.

The hypothesized frequency band characterization *a) external, b) respiration, c) blood pressure/heart rate, d) HF/noise* is, all in all, confirmed. It is also seen in fig. 4.4 in this case to be very well delimited at the frequencies 0.055, 0.75 and 3.75 Hz.

The same information that we find in the forehead segment also occurs in the palm segment. The base harmonic of the heart rate is clearly seen in fig. 4.5, and the first and second are also clear enough to be made out. Compare fig. 6.2(a), however, to see that the noise level between the harmonics is much higher. The variations due to respiration can be seen in fig. 4.6 although they are more ambiguous than the ones in 4.4.

7.2   Without Movement

From the preliminary exploration in 4 and from previous work, it is expected that a good heart rate estimation should be readily available from a face video without movement. This is confirmed in 6.1.1 and 6.1.2 where a segment containing the whole face (G_{face}) is compared to a segment containing bare skin on the forehead only ($G_{forehead}$). A radical difference between these is expected, and confirmed by figs. 6.1 and 6.3, the latter having a significantly higher SNR. In fig. 6.2 it is seen that going from G_{face} to $G_{forehead}$ we remove noise both above and inbetween the visible harmonics of the heart rate. The second harmonic, partly visible in $G_{forehead}$ is completely camouflaged by noise in G_{face}.

Estimations of the heart rate done from $G_{forehead}$ and G_{face} concur closely with those done by finger pulse oximeter, as the numbers in tab. 6.1 show: $G_{forehead}$ has a small but not insignificant edge[1], nowhere differing from the control value by more than 0.6 bpm.

The use of JADE source separation here is deemed unsuccessful or at the very least unnecessary. Although $J1_{face}$ in fig. 6.4 shows a strong base frequency component – we may even see traces of a first harmonic – the noise level is evidently worse than that in G_{face}. We see a decline in the quality of the heart rate estimations with increases in both maximum error and error variance. The mixing matrix $\mathbf{M}_{face}$ was included in case it could reveal any clue to which of the JADE output signals was the most beneficial choice, as we saw in ch. 4.2, but no such clue was revealed.

We remark that this is not the case for which JADE is used in [1], so different results may be expected when movement and face detection are involved, meaning no conclusions are yet to be drawn. If forehead segmenting can be done in an automatized way, it may provide a method of improving the signal, admittedly less sophisticated, but algorithmically a lot less complex and maybe somewhat more intuitively understandable.

---

[1]The heart rate estimation from the face may not show overall convergence to the one from the finger, even when they are done by the same method – see fig. 6.22(a) where both rates are estimated by use of webcam.

7.3 With Movement and Face Detection

When we include face detection, and the procedure is repeated on recordings where the subject is moving, we see a distinct decline in signal quality. This is expected due to the physical movement of the subject, but also, as we will see, due to the movement of the face tracker itself – even when two video frames are nearly identical, their detected face coordinates will differ, leading to a "choppy" behavior. A smoothing-out of the coordinates is suggested as a way of improvement, and seemingly (see fig. 6.6) a simple one-second MA filter does a good job of this. In tab. 6.2 we see an improvement on the three signals that are based on face detection when the coordinates are filtered ("MA30"), both maximum errors and error variances are down on all three of them. We compare spectrograms: fig. 6.7(a) to fig. 6.7(b) and fig. 6.9(a) to 6.9(b). In both cases and especially in the last, the coordinate-filtered version is perceived to be less noisy: base frequency is more clearly protruding and traces of harmonics are more prominent.

In fig. 6.8 we see the first point in these experiments at which the application of JADE begets beneficial results: for the first 35 seconds we are not able to trace the heart rate from $G_{recface}(t)$ (when the low threshold is 45 bpm), because of low frequency disturbances higher in amplitude than the base heart frequency. This is not the case for $J2_{recface}(t)$ (disturbances present in fig. 6.7(a) but not in fig. 6.12(a)). We remark that this is the case for which JADE was applied successfully in [1], and choose to take our results as indicative of the reason for this success.

JADE is not helpful in the case of a forehead-only segment, we see in fig. 6.10 that a tracking of the heart rate is not successful for $J2_{recforehead}(t)$, not even when the low threshold is raised to 52 bpm. Its spectrogram is not included, as it was not particularly enlightening. As we see from comparing the figs. 6.12(a) and 6.12(b), the coordinate filtering led only to a noise enhancement for $J2_{recface}(t)$, due to this, filtering was not attempted for $J2_{recforehead}(t)$.

7.4 Pulsatility Mapping

The method described in 3.4 and applied in 6.3, 6.4 and 3.5/6.5 may be viewed as a method for either **(i)** mapping the distribution of the amplitude of the pulse signal in the skin that is exposed to the camera, **(ii)** semi-automatically extracting the physiologic signal with both heart and respiration rates including harmonics and modulation products or **(iii)** robustly estimating the bpm value of the heart rate by "majority vote". Each of these discussed separately here:

(i) This is *medically* interesting if it does actually map an underlying anatomical structure – which it is reasonable to assume that it does, to some degree. The resulting distribution map in fig. 6.15 reflects an expected [2] higher pulsatility in the forehead and upper cheak areas. It must, though, be remarked that angle of lighting, angle of projection and shadowing resulting from both of these all play important roles, i.e. a different result would be expected if the subject was to tilt his head sideways or forward/backward. But, however important this role is, we

still find the result *technically* pleasing because it tells us where to look for the
wanted signal. In the flatter palm area, angles will play less of a role but still have
an influence. Here (fig. 6.13) we see higher pulsatility in the "meatier" parts of
the hand, but the pattern is less clear than in the face.

(ii)   This is done in 6.3.1 and 6.3.2 and is successful, at least what the heart rate
and a few harmonics are concerned. This is seen in figs. 6.14(a) and 6.16(a). When
these spectrograms are compared to their more naively extracted counterparts in
figs. 6.14(b) and 6.16(b) we even perceive some overall improvement – "cleaning-
up" of noise between the harmonics – this is a likely result, but hard to judge
accurately from the spectrograms (see also figs. 6.18 and 6.19). Important is that
it is at all possible to extract the signal in this way, a way that may prove more
generalizable than separating out the forehead from a face detection, because of its
independence of face shape.

(iii)   In 6.5 we see an example of this. Although we in the initial (ii) Testing 1
register some instability for L≤6, for L=7 the method behaves well, with an RMSE
of only 2.3 bpm. It behaves similarly in Testing 2 when the number of subjects
and correspondingly the amount of data points increases, only with somewhat
higher spread and additional outliers. We take care when comparing the Bland-
Altman plot in fig. 6.23 to the ones in fig. 2.6, the latter being from tests where
the pulse is tracked continuously. The tests carried out here are "pointwise", i.e.
not taking advantage of the fact that changes in the pulse rate appear slow and
gradual, and may therefore be considered stronger. If a continuous tracking were
to be implemented using this method, acceptable results would be expected also
for L<7, where errors were dominated by a significant number of outliers.

7.5   Comparing PPG from different Locations

Such comparisons are shown to be possible in ch. 6.4, and also in [2] where phase
distribution is used to assess blood flow in portwine stains.

From the technical point of view, it is satisfying that we are able to acquire this
phase information with such a low equipment threshold. Practically, one might
find the positioning of both hand and face within the video frame (while staying
relatively still) physically awkward. This practice was used largely as an answer to
the need for synchronization, so if there did exist a simple way to record multiple
frame-synchronized videos, one might just as well use two cameras (which would
in turn make possible measurements on other locations, e.g. lower extremities).
Medically, there is also the question of whether raising one's hand to face height
changes the vascular compliance, blood pressure, heart rate or other parameters
that may influence the PWV, a question which is not answered here.

7.5.1 Phase Delay

We see that measuring the average phase delay (see 3.3) at the heart rate frequency between two locations is possible. The results of this for a two-minute recording of forehead/palm is shown in fig. 6.22(c), averages over 12 seconds. Regarding these results as an estimate of the face-hand PTTD: a more direct way would be to measure the distance between corresponding peaks in the two PPG signals[2]. Finding the peak distances may not be straightforward for a noisy signal whose morphology also varies, and the phase delay is believed to be a good and more stable estimate. Indeed, a crude manual inspection of the waveform in fig. 6.20 reveals that nearly half of the peaks[3] in the palm signal are found at the same sample point as their corresponding forehead peaks, and the rest are rather symmetrically distributed around that, though leaning slightly towards a positive lag, with an average of about 0.15 sample points (forehead peaks appear before palm peaks). This is equivalent to $0.15 \cdot 1000/30 = 5$ ms, compare the average estimated phase delay of 6.6 ms.

The accuracy of the results is also rightly questioned on grounds of the high (30 ms) sample period (low temporal resolution) of the signals. If, for example, the actual PTTD was a *steady* 6.6 ms, nothing would register at all. The fact that in this case the PTTD was not behaving that steadily but was varying from beat to beat, combined with averaging over 12-second time windows, gives the values higher accuracy.

7.5.2 Respiration Rates

The extracted signals for both forehead and palm show significant frequency peaks in the 0.055-0.75 Hz band, these are assumed to be a result of respiration (see 7.1). The content is shown filtered in fig. 6.21, and an estimate of its varying frequency was done, using FFT and a 12-second sliding window. The respiration rate estimated from the palm is seen in fig. 6.22(b) to generally be lower than that estimated from the forehead, their means are respectively 0.104 and 0.128 Hz or 6.2 and 7.7 "breaths per minute", an average disparity of 23%. (Had the values been closer to each other, a phase delay could have been found also at these frequencies.) We also see no clear shape correlation between them. It is uncertain whether this has to do with poor signal quality (affected by unphysiological disturbances) or differing blood flow velocity influencing the oxygen transport. A good coherence between actual respiratory rate (breaths taken per minute) and the value measured by PPG is reported [31]. The physiologic properties of the webcam-acquired signal could be investigated further with knowledge of the accurate RGB wavelength responses of the camera.

---

[2]or rather peaks in their polarity reversed versions, as the PPG is inversely related to blood pressure (see 2.1.1)

[3]of the ∼100 peaks in this signal with unambiguous positions

7.5.3   Heart Rates

The heart rates estimated from the two PPG locations are shown in fig. 6.22(a). We see that they are not exactly the same: the estimates from the palm have slightly more abrupt changes. All positive differences are evened out by negative, though, and they show the same average, as one would expect (68.6 bpm).
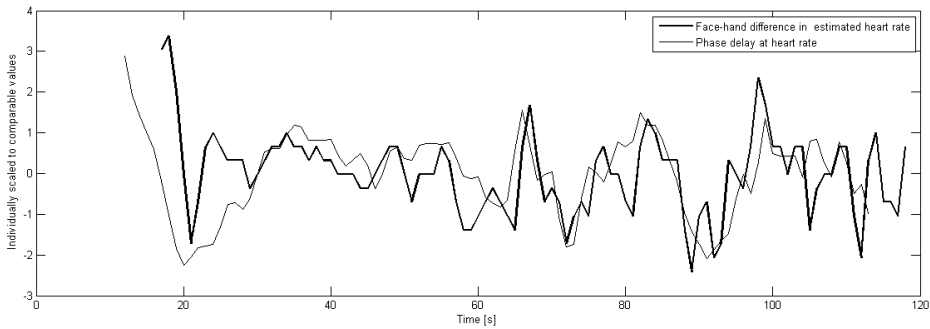

Figure 7.1: Phase delay variations compared to variations in heart rate difference

At the points where an above-average difference in heart rates between forehead and palm is measured, one would expect the phase delay to also be above average, i.e. some form of correlation between heart rate difference and phase delay. This is shown to hold some truth in fig. 7.1 and, though not further investigated, taken as evidence for the validity of both estimates.

7.6   Necessary Duration of Recording

In both [1] and [2] analysis is performed on time windows of 20-30 seconds duration[4]. In chapters 6.1-6.4 of this report, a duration of 12 seconds is shown to give a good balance between temporal resolution and unambiguity in the results. With a reasonably clean signal, however, it is natural to expect good results to be attainable even for durations long enough for only a few heart cycles. We see from the results in 6.5 an indication that this is true. Of the two methods, (i) shows a clearer pattern of convergence with increasing L, but they perform equally well when L reaches 7 seconds, in this limited testing scheme. Further testing on other subjects confirms this for method (ii); in fig. 6.23 we see that with an RMSE of 4.95 most measurement points are within a 5-10 bpm error, with only few outlier errors. The testing range is limited but covers much of the normal values for resting heart rate. As mentioned in 7.4, we expect that a continuous tracking of heart rate could remove some of the outliers and function well also with $L < 7$. Good temporal resolution in HR monitoring gives meaningful values for the HRV, this has several possible applications, as mentioned in ch. 2.6.

---

[4][2] mentions the use of recording durations as low as 10 seconds

7.7   Laparoscopy

From the plots in ch. 6.6 we see the frequency peak that is the most likely candidate for the heart rate at 1.87 Hz (112 bpm). This is visibly present in the JADE independent components from all 4 ROI's as well as in the RGB traces from ROI 3 and 4. We interpret this as JADE here being able to extract pulsatility components that are not clearly visible in the RGB traces. When they are clearly visible, however, we see no drastic change in the waveform. Peaks that may be consistent with respiratory rate are observed in a few of the plots at about 0.2 Hz.

The R trace from ROI 3 is seen in fig. 6.27 (top, right) to have a peak at 1.87 Hz of higher amplitude than that in the G and B traces. This is an indication that these pulsations are influenced by movement artifact[5] more than by true PPG. (This is also seen from the video itself.) We would expect signal traces extracted from a video of longer duration and higher stability to have content where the influence of true PPG outweighs this movement artifact, and that the different regions could be recognized (by phase, amplitude) more clearly in a comparison like the one in fig. 7.2, maybe even to the point of diagnostical usefulness, i.e. the ability to distinguish cancerous tissue from surrounding non-cancerous tissue.

Possible observed respiratory rate fluctuations are even more likely to be influenced by movement – consider the large influence of breathing mechanics on the abdomen. The mixing matrices were excluded from these results as they gave no further insight.
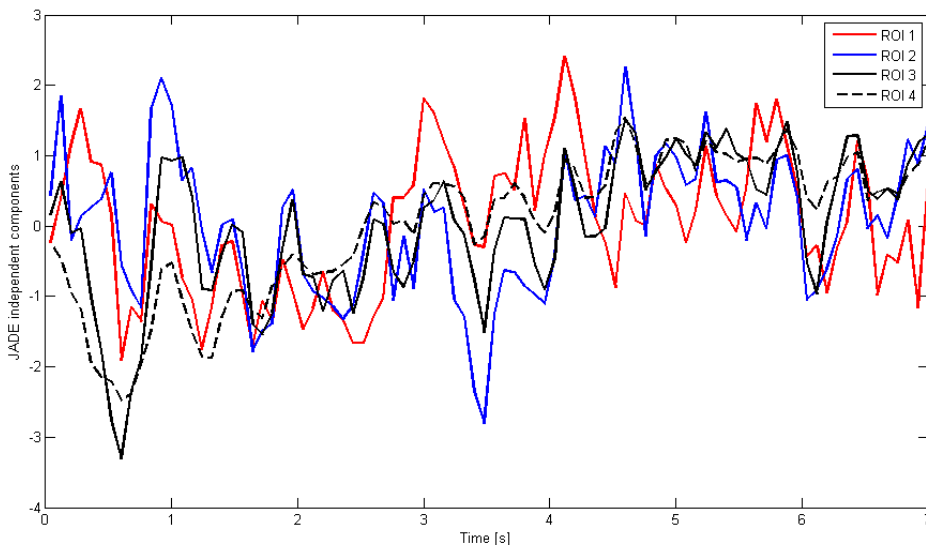

Figure 7.2: JADE components from laparoscopy compared

---

[5]though the heart rate may still be correct

7.8   Problems and Artifacts

Several parameters were seen to have an influence on the quality of the acquired pulsatile signal. Among them, interference from fluorescent lighting was found especially disruptive at one recording location, but also the correct choice of white balance[6] has seemed vital for obtaining a high quality signal.

All video used was uncompressed, but it was found that no information was lost neither through MJPEG (intra-frame only) compression nor several different modes of heavy H264 (inter-frame) compression. A Sony Handycam using another mode of MPEG-4, however, was found to produce video where any and all physiological pulsatility was masked completely by other spectral lines, suspectedly with relation to the keyframe interval of the codec. (One might comment that a good video compression scheme *should* remove such information, as it is not perceived by anyone.) A few movie scenes with face footage of sufficient length were found and analyzed but revealed no pulsatile signal. However well lit a film set is, this is not an unexpected result, taken into account heavy use of makeup and post-processing (dynamic compression, data compression).

A quantification of signal-to-noise ratio (SNR) is never attempted in this report. There would be certain difficulties involved in defining an SNR, i.e. what is to be defined as signal and what as noise? (High/low frequency noise, noise between the different harmonics, modulation products etc.) In many cases the *difference* in SNR is very clear, and quantification is not seen as necessary. In other cases it is less clear, and no strict conclusion is drawn. Signal quality is also assessed by how well the signal tracks the heart rate, and visually by spectrograms. Further investigations may warrant a definition of SNR.

---

[6]notorious for being chosen wrongly by automatic adjustments, especially indoors

Chapter 8

Conclusions

It was known that heart and respiratory rate can be extracted by using a web camera to film someone's exposed skin, especially his forehead, even with daylight as the only source of illumination. To expand on this we have investigated ways in which to do this extraction, as well as looked into what other information can be gathered from such a recording. To finish we briefly assess possible applications.

8.1   Extracting the best possible Signal

The arguably cleanest signal we have seen in this report is the one spectrographed in fig. 6.18(a) which was extracted by use of pulsatility mapping (ch. 3.4) from a video without movement. Compared to the output from a finger pulse oximeter in fig. 2.4(b) we see radical difference. The clarity of the non-contact ambient-light photoplethysmogram is severely restricted by noise and lack of illumination control.

As long as the subject is keeping relatively still, we extract a similarly clear signal by use of simple manual segmenting of the forehead (fig. 6.2(a)), a scheme which we were able to apply also when movement and automatic face tracking was introduced (ch. 6.2.2). The signals extracted from videos with movement using face tracking are, however, vulnerable to the nature of the tracker. Filtering of the coordinates alleviates some of this problem (ch. 6.2.3).

JADE source separation (ch. 2.3.1) is admitted some merit in revealing the heart rate where it was not visible (ch. 7.7) and in one case we see JADE causing an improvement in the detectability of the heart rate by removing lower-frequency disturbances (ch. 7.3). In addition to this there is rather conclusive evidence from [1] that JADE improves the heart rate detectability, in the case where the whole face of the subject is taken as the ROI. However, when it comes to obtaining the clearest possible signal, we show more evidence in favour of the more naively thought out procedures of forehead-segmenting and pulsatility mapping. All in all we have reason to assume that JADE works better when the starting point is a signal of lower quality (it can indeed uncover a signal in very high noise levels, as in ch. 4.2).

8.2   What Information can be acquired?

We give examples of **heart rate** estimations with as low as 0.2 bpm average error, 0.6 bpm maximum error and 0.2 bpm error standard deviation (compared to finger pulse oximeter), when the subject is not moving, from a video of duration 2 minutes (ch. 6.1.1). When the subject is moving (slow horizontal movements within the 100 middle pixels of a 360×360 video frame) the numbers are 0.1, 2.1 and 0.7 bpm, accordingly; from a video of similar duration (ch. 6.2.3).

Many of the spectrograms in this report clearly show several harmonics to the base HR frequency: the signals are recovered with more than a minimum of phase information. This means we can find the **heart rate variability** with more than a minimum of accuracy (fig. 6.10). The difference in pulse transit time between two distal locations (forehead and palm) is found in ch. 6.4.1. This shows that the phase information contained in the signals, both from forehead and palm, is robust enough to make use of the **phase distribution**.

For both face and hand **amplitude distribution** is found (ch. 6.3), not only interesting in itself but also as a means of excluding parts that are less or not at all pulsatile. Differing results from face to hand give us reason to doubt the values obtained for **respiratory rate** (ch. 6.4.2), but previous work has concluded that the information is present in the signals.

8.3   Possible Applications

Even the very limited results of this report show the clear possibility for creating an application (e.g. smartphone, laptop with webcam) that tracks heart rate (HR and HRV) from face video in real time, concurring with results from previous work [1, 2]. (To our knowledge this does not exist.) Such an application may be of use in a multitude of situations, from clinical patient monitoring to sports. We suggest applying pulsatility mapping or at least some way of automatically confirming the pulsatility of the signal in such an application, seeing as this works reasonably well even for short durations (ch. 6.5). Though movement can be compensated for to some degree, abrupt displacement will cause disturbances in the signal – the user would basically be required to face the camera and limit his movement to what is normal when interacting with the computer.

Other information that may prove medically relevant includes PTTD (see ch. 2.5) and respiratory rates. Applications involving these are thinkable in a remote monitoring situation where cost and availability is important. On the topic of laparoscopy (ch. 6.6) and other types of endoscopic recordings we are more reluctant to conclude; it is likely that PPG signals are available also in these cases, but this was not clearly demonstrated in our work.

Bibliography

[1] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard, *Non-contact, automated cardiac pulse measurements using video imaging and blind source separation*, Optics Express **18(10)** (May 2010).

[2] Wim Verkruysse, Lars O. Svaasand, and J. Stuart Nelson, *Remote plethysmographic imaging using ambient light*, Optics Express **16(26)** (December 2008).

[3] *The British Medical Association Illustrated Medical Dictionary*, 2nd ed., 2007.

[4] Andrews Reisner, Phillip A. Shaltis, Devin McCombie, and H. Harry Asada, *Utility of the Photoplethysmogram in Circulatory Monitoring*, Anesthesiology **108(5)** (May 2008), 950–958.

[5] M. Hirai, S. L. Nielsen, and N. A. Lassen, *Blood pressure measurement of all five fingers by strain gauge plethysmography*, Scandinavian Journal of Clinical & Laboratory Investigation **36(7)** (1976), 627–632.

[6] J Ando, A Kawarada, M Shibata, K Yamakoshi, and A Kamiya, *Pressure-volume relationships of finger arteries in healthy subjects and patients with coronary atherosclerosis measured non-invasively by photoelectric plethysmography.*, Japanese Circulation Journal **55(6)** (1991), 567–575.

[7] A.A.R. Kamal, J.B. Harness, G. Irving, and A.J. Mearns, *Skin photoplethysmography — a review*, Computer Methods and Programs in Biomedicine **28(4)** (1989), 257–269.

[8] S. Wendelken, S. McGrath, G. Blike, and M. Akay, *The feasibility of using a forehead reflectance pulse oximeter for automated remote triage*, Proceedings of the IEEE 30th Annual Northeast Bioengineering Conference, 2004 (April 2004), 180–181.

[9] Azumio Inc., *Instant Heart Rate*, June 2012. Smartphone application for Android/iOS, `http://www.instantheartrate.com/`.

[10] S López-Silva, M Dotor, and R. Giannetti, *On laparoscopic photoplethysmography and pulse oximetry*, Journal of Clinical Monitoring and Computing **24(3)** (2010), 219-220.

[11] W.G. Zijlstra, A. Buursma, and W.P. Meeuwsen-van der Roest, *Absorption spectra of human fetal and adult oxyhemoglobin, de-oxyhemoglobin, carboxyhemoglobin, and methemoglobin*, Clinical Chemistry **37(9)** (September 1991), 1633–1638.

[12] W. Cui, L.E. Ostrander, and B.Y. Lee, *In vivo reflectance of blood and tissue as a function of light wavelength*, IEEE Transactions on Biomedical Engineering **37(6)** (June 1990), 632–639.

[13] Kevin K. Tremper, *Pulse Oximetry*, Anesthesiology **70(1)** (April 1989), 98–108.

[14] Pierre Comon, *Independent component analysis, A new concept?*, Signal Processing **36(3)** (1994), 287–314.

[15] A. Hyvärinen and E. Oja, *Independent component analysis: algorithms and applications*, Neural Networks **13(4–5)** (2000), 411–430.

[16] James V. Stone, *Independent component analysis: an introduction*, Trends in Cognitive Sciences **6(2)** (2002), 59–64.

[17] Wei Lu and J.C. Rajapakse, *Approach and applications of constrained ICA*, IEEE Transactions on Neural Networks **16(1)** (January 2005), 203-212.

[18] J.F. Cardoso, *JADE in Matlab*, May 2005. Implementation of independent component analysis algorithm, `http://perso.telecom-paristech.fr/~cardoso/guidesepsou.html`.

[19] J. F. Cardoso, *High-order contrasts for independent component analysis*, Neural Computation **11** (1999), 157–192.

[20] S. Krishna, *Open CV Viola-Jones Face Detection in Matlab*, May 2008. Implementation of face detection algorithm, `http://www.mathworks.com/matlabcentral/fileexchange/19912`.

[21] Inc. Biopac Systems, *Pulse Transit Time and Velocity Calculation*, March 2006. Application note, `http://www.biopac.com/Manuals/app_pdf/app117.pdf`.

[22] W. Lu, H. Li, S. Tao, D. Zhang, Z. Jiang, L. Cui, J. Tu, and D. Gou, *Research on the main elements influencing blood pressure measurement by pulse wave velocity.*, Frontiers of medical and biological engineering **4(3)** (1992), 189–199.

[23] R. A. Payne, C. N. Symeonides, D. J. Webb, and S. R. J. Maxwell, *Pulse transit time measured from the ECG: an unreliable marker of beat-to-beat blood pressure*, Journal of Applied Physiology **100** (2006), 136–141.

[24] M. Nitzan, B. Khanokh, and Y. Slovik, *The difference in pulse transit time to the toe and finger measured by photoplethysmography*, Physiological Measurement **23** (February 2002), 85–93.

[25] Robert E. Kleiger, J.Philip Miller, J.Thomas Bigger Jr., and Arthur J. Moss, *Decreased heart rate variability and its association with increased mortality after acute myocardial infarction*, The American Journal of Cardiology **59(4)** (1987), 256–262.

[26] M.T. La Rovere, J.T. Bigger, F.I. Marcus, A. Mortara, and P.J. Schwartz, *Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction. ATRAMI (Autonomic Tone and Reflexes After Myocardial Infarction) Investigators.*, The Lancet **351(9101)** (February 1998), 478–484.

[27] Conny M. A. van Ravenswaaij-Arts, Louis A. A. Kollee, Jeroen C. W. Hopman, Gerard B. A. Stoelinga, and Herman P. van Geijn, *Heart Rate Variability*, Annals of Internal Medicine **118(6)** (1993), 436–447.

[28] Polar Electro, *Polar Fitness Test™and OwnIndex®*, 2012. `http://www.polar.fi/about_polar/who_we_are/research/fitness_test_with_OwnIndex`; `http://www.polar.fi/en/support/Polar_Fitness_Test_and_OwnIndex`.

[29] A.E. Aubert, B. Seps, and F. Beckers, *Heart Rate Variability in Athletes*, Sports Medicine **33(12)** (2003), 889–919.

[30] Mikko P. Tulppo, Arto J. Hautala, Timo H. Mäkikallio, Raija T. Laukkanen, Seppo Nissilä, Richard L. Hughson, and Heikki V. Huikuri, *Effects of aerobic training on heart rate dynamics in sedentary subjects*, Journal of Applied Physiology **95(1)** (July 2003), 364–372.

[31] Yue-Der Lin, Wei-Ting Liu, Ching-Che Tsai, and Wen-Hsiu Chen, *Coherence Analysis between Respiration and PPG Signal by Bivariate AR Model*, World Academy of Science, Engineering and Technology **53** (2009).

Appendix A

Matlab Script Samples

A.1 Extracting Signals, Face Detection

```matlab
% clc
% clear all,close all force

% Doing face recognition on each frame and storing coordinates

FILENAME = 'facevid.avi';
OBJ = VideoReader(FILENAME);

numFrames = get(OBJ, 'NumberOfFrames');
disp(numFrames);

h = waitbar(0,'Please wait...');
noface=zeros(1,numFrames);
Face=zeros(4,numFrames);
RGB=zeros(3,numFrames);

for k = 1 : numFrames, waitbar(k/numFrames,h)

    % Read frame, recognize face, make sure everything is ok or else using
    % coordinates from previous frame.
    A = read(OBJ,k);
    Img = double (rgb2gray(A));
    AFace = FaceDetect('haarcascade_frontalface_alt2.xml',Img);clear Img
    if all(size(AFace)==[1 4])
        Face(:,k)=AFace;
    elseif all(size(AFace)==[2 4])
            d1=abs(AFace(1,1)-Face(1,k-1));
            d2=abs(AFace(2,1)-Face(1,k-1));
            if d1<=d2, Face(:,k)=AFace(1,:);
            else
                Face(:,k)=AFace(2,:);
            end
    elseif all(size(AFace)==[3 4])
            d1=abs(AFace(1,1)-Face(1,k-1));
            d2=abs(AFace(2,1)-Face(1,k-1));
            d3=abs(AFace(3,1)-Face(1,k-1));
            if d1<=d2&&d1<=d3, Face(:,k)=AFace(1,:);
            elseif d2<=d1&&d2<=d3, Face(:,k)=AFace(2,:);
            else
                Face(:,k)=AFace(3,:);
```

```
            end
    else
        Face(:,k)=Face(:,k−1);
        noface(k)=1;
    end
    clear AFace

    % Reducing width of face rectangle by 40%
    reduceWidthBy=.4*Face(3,k);
    Face(1,k)=round(Face(1,k)+.5*reduceWidthBy);
    Face(3,k)=round(Face(3,k)−reduceWidthBy);

    % Extracting RGB averages
    f1=Face(1,k);f2=Face(2,k);f3=Face(3,k);f4=Face(4,k);
    for ii=1:3
        RGB(ii,k)=(1/(f3*f4))*sum(sum(A(f2:f2+f4,f1:f1+f3,ii)));
    end
end

close all force
save('facevid.mat','RGB', 'Face', 'noface')
```

A.2   Heart Rate Calculation

```
function [time pBPM] = compute_pulse(ppg,Fs,pMin)%,wL, wD, pMin, pMax)

% FUNCTION compute_pulse(ppg,Fs,wL, wD, pMin, pMax)
% ppg:  ppg or similar signal containing heart rate pulsations
% Fs:   sampling rate in samples per second
% wL:   window length in seconds
% wD:   window displacement in samples
% pMin: minimum pulse frequency in bpm
% pMax: maximum pulse frequency in bpm

wL=12;
wD=Fs;
% pMin=45;
pMax=200;

wL=wL*Fs;
w=hamming(wL);
% w=ones(wL,1);
wN=round((length(ppg)−wL)/wD);
fN=Fs*60*5;
pMin=round(pMin*fN/Fs/60);
pMax=round(pMax*fN/Fs/60);
pI=zeros(1,wN);

for ii=1:wN
    start=wD*(ii−1)+1;
    stop=start+wL−1;
    ppgSegmentWindowed=stdnrm(stdnrm(ppg(start:stop)).*w');
    ppgSegmentPSD=abs(fft(ppgSegmentWindowed,fN)).^2;
    ppgSegmentPSD([1:pMin pMax:end])=0;
    [¬,pI(ii)]=max(ppgSegmentPSD);
    if ii>1
```

```
        while pI(ii)>1.1*pI(ii−1)||pI(ii)<.9*pI(ii−1)
            ppgSegmentPSD(pI(ii))=0;
            [¬,pI(ii)]=max(ppgSegmentPSD);
        end
    end
end

pBPM=pI*Fs*60/fN;
% time=linspace(0,Fs*length(pBPM)/wD,length(pBPM));
time=.01*round(100*linspace(wL/Fs,wL/Fs+Fs*length(pBPM)/wD,length(pBPM)));
```

A.3 Pulsatility Mapping

```
%% Extract the G channel for each square in a grid

% clear all
% close all force
FILENAME = 'C:\Users\asmundr\Videos\01\01crop.avi';
OBJ = VideoReader(FILENAME);

numFrames = get(OBJ, 'NumberOfFrames');
% disp(numFrames)
frame1 = read(OBJ,1);
res=size(frame1);
yres=res(1);
xres=res(2);

image(frame1)
axis equal
hold on

dx=4; % x−dim of square
dy=4; % y−dim of square
numBlocks=xres*yres/(dx*dy);

% create vector of square coordinates on form
% [x−pos top left, y−pos top left, dx, dy]
b=zeros(4,numBlocks);
b(1,:)=mod(1:dx:dx*numBlocks,xres);
b(2,:)=sort(mod(1:dy:dy*numBlocks,yres));
b(3,:)=dx*ones(1,numBlocks);
b(4,:)=dy*ones(1,numBlocks);

% display grid on video frame
for ii=1:numBlocks
rectangle('position',b(:,ii),'edgecolor','g')
end
hold off

%% Average pixel value for each square and each frame
R=zeros(numBlocks,numFrames);
G=zeros(numBlocks,numFrames);
B=zeros(numBlocks,numFrames);
h = waitbar(0,'Please wait...');
for ii=1:numFrames
    waitbar(ii/numFrames,h)
```

```matlab
    frameii=read(OBJ,ii);
    for jj=1:numBlocks
        rect=frameii(b(2,jj):b(2,jj)+dy-1,b(1,jj):b(1,jj)+dx-1,1);
        R(jj,ii)=sum(sum(rect))/(dx*dy);
        rect=frameii(b(2,jj):b(2,jj)+dy-1,b(1,jj):b(1,jj)+dx-1,2);
        G(jj,ii)=sum(sum(rect))/(dx*dy);
        rect=frameii(b(2,jj):b(2,jj)+dy-1,b(1,jj):b(1,jj)+dx-1,3);
        B(jj,ii)=sum(sum(rect))/(dx*dy);
    end
end
for ii=1:numBlocks, G(ii,:)=zmean(G(ii,:)); end
save('01crop4x4rgb.mat','G','R','B')
close all force




%——————————————————————————————————————————————
load 01crop4x4rgb.mat

Gd=zeros(size(G));
for ii=1:size(G,1)
    Gd(ii,:)=limitdiff(G(ii,:));
end
G=Gd;

%compute PSDs
N=2^nextpow2(length(G));
GF=abs(fft(G',N)).^2/N;
GF=GF(1:N/2,:);
idx28bpm=round(28*N/(2*15*60));
idx200bpm=round(200*N/(2*15*60));%2*15*60*71/N
GF(1:idx28bpm,:)=0;
GF(idx200bpm:end,:)=0;
f=linspace(0,15*60,N/2);


%find peaks
for ii=1:size(G,1)
    [Y(ii) I(ii)]=max(GF(:,ii));
end

Icorrectpulse=find(I==mode(I));
Iwrongpulse=find(I≠mode(I));
Y(Iwrongpulse)=0;
Ibest=find(Y>3*std(Y));      % red squares
Igood=find(Y>2*std(Y));      % yellow squares
Iok=find(Y>std(Y));          % green squares
Ilessok=find(Y>.5*std(Y));   % blue squares
% figure,plot(Y)
% figure,hist(Y)


GZ=stdnrm(sum(G(Ibest,:))); % summing best signals
figure,plotbands(stdnrm(double(GZ)),'k')
myspectrogram(double(GZ),30,'log')
% plot(G(145,:))


% show previously computed image averaged over all frames to indicate
% video is not completely stable
load hand7c_averageimage
frame1 = frameaverage;
res=size(frame1);
yres=res(1);
xres=res(2);
figure
imagesc(frame1)
axis equal
```

```
hold on
dx=8;
dy=8;
numBlocks=xres*yres/(dx*dy);

b=zeros(4,numBlocks);
b(1,:)=mod(1:dx:dx*numBlocks,xres);
b(2,:)=sort(mod(1:dy:dy*numBlocks,yres));
b(3,:)=dx*ones(1,numBlocks);
b(4,:)=dy*ones(1,numBlocks);

% display colored squares overlayed video frame:

for ii=Ilessok
rectangle('position',b(:,ii),'facecolor','b')
end

for ii=Iok
rectangle('position',b(:,ii),'facecolor','g')
end

for ii=Igood
rectangle('position',b(:,ii),'facecolor','y')
end

for ii=Ibest
rectangle('position',b(:,ii),'facecolor','r')
end

hold off

% load('hand7cc.mat')
% myspectrogram(limitdiff(G),30,'log')
% figure,plotbands(stdnrm(double(G)),'k')
```

A.4   Phase Delay Calculation

```
function [time ptt] = compute_ptt(ppg1,ppg2,Fs,pMin)%,wL, wD, pMin, pMax)

% Computes PULSE TRANSIT TIME, i.e. the phase delay between the (assumed to
% have common frequency) pulsations of ppg1 and ppg2

% FUNCTION compute_pulse(ppg,Fs,wL, wD, pMin, pMax)
% ppg:  ppg or similar signal containing heart rate pulsations
% Fs:   sampling rate in samples per second
% wL:   window length in seconds
% wD:   window displacement in samples
% pMin: minimum pulse frequency in bpm
% pMax: maximum pulse frequency in bpm

wL=12;
wD=Fs;
% pMin=45;
pMax=200;

wL=wL*Fs;
```

```matlab
w=hamming(wL);
% w=ones(wL,1);
wN=round((length(ppg1)-wL)/wD);
fN=Fs*60*5;
pMin=round(pMin*fN/Fs/60);
pMax=round(pMax*fN/Fs/60);
p1I=zeros(1,wN);
p2I=zeros(1,wN);
ptt1=zeros(1,wN);
ptt2=zeros(1,wN);

for ii=1:wN
    start=wD*(ii-1)+1;
    stop=start+wL-1;

    % find frequency peak in ppg1, as in compute_pulse.m
    ppg1SegmentWindowed=stdnrm(stdnrm(ppg1(start:stop)).*w');
    ppg1SegmentPSD=abs(fft(ppg1SegmentWindowed,fN)).^2;
    ppg1SegmentPSD([1:pMin pMax:end])=0;
    [¬,p1I(ii)]=max(ppg1SegmentPSD);
    if ii>1
        while p1I(ii)>1.1*p1I(ii-1)||p1I(ii)<.9*p1I(ii-1)
            ppg1SegmentPSD(p1I(ii))=0;
            [¬,p1I(ii)]=max(ppg1SegmentPSD);
        end
    end

    % find frequency peak in ppg2, as in compute_pulse.m
    ppg2SegmentWindowed=stdnrm(stdnrm(ppg2(start:stop)).*w');
    ppg2SegmentPSD=abs(fft(ppg2SegmentWindowed,fN)).^2;
    ppg2SegmentPSD([1:pMin pMax:end])=0;
    [¬,p2I(ii)]=max(ppg2SegmentPSD);
    if ii>1
        while p2I(ii)>1.1*p2I(ii-1)||p2I(ii)<.9*p2I(ii-1)
            ppg2SegmentPSD(p2I(ii))=0;
            [¬,p2I(ii)]=max(ppg2SegmentPSD);
        end
    end

    % take the angle of the fft's
    ppg1SegmentAngle=angle(fft(ppg1SegmentWindowed,fN));
    ppg2SegmentAngle=angle(fft(ppg2SegmentWindowed,fN));

    % find difference in angle at peak points. two versions of this
    % difference are found, assuming the peak point are not exactly the
    % same for the two signals.
    ptt1(ii)=(ppg1SegmentAngle(p1I(ii))-ppg2SegmentAngle(p1I(ii)))/Fs;
    ptt2(ii)=(ppg1SegmentAngle(p2I(ii))-ppg2SegmentAngle(p2I(ii)))/Fs;

end

% "averaging" ptt1 and ptt2. where "discontinuities" are found in one but
% not in the other, use other. where "discontinuities" are found in both,
% use previous value.
c=.5*(ptt1(1)+ptt2(1));
d1=diff(ptt1);
d2=diff(ptt2);
d3=zeros(size(d2));
s1=2*std(d1);
s2=2*std(d2);
for ii=1:length(d1)
    if all([abs(d1(ii))>s1 abs(d2(ii))<s2])
        d3(ii)=d2(ii);
    elseif all([abs(d2(ii))>s2 abs(d1(ii))<s1])
        d3(ii)=d1(ii);
    elseif all([abs(d2(ii))>s2 abs(d1(ii))>s1])
        d3(ii)=0;
    else
```

```
        d3(ii)=.5*(d1(ii)+d2(ii));
    end
end
ptt=[c cumsum(d3)+c];

time=.01*round(100*linspace(wL/Fs,wL/Fs+Fs*length(ptt)/wD,length(ptt)));
```

A.5 Short-Time Heart Rate Estimation (ii)

```
%%
clear all,close all

% initialize camera
fps=10; %frames per second
snapshot=2; %number of seconds to record
numFrames=snapshot*fps;

%construct webcam object 3*8bit RGB @ res 640*360
obj=videoinput('winvideo',1,'RGB24_640x360');
triggerconfig(obj, 'manual');
obj.FramesPerTrigger=numFrames;
src=getselectedsource(obj);

%set properties of exposure etc.
%all auto—behavior set to off to avoid unwanted fluctuations,
%meaning settings must be adjusted to environment
src.BacklightCompensation = 'off';
src.ExposureMode = 'manual';
src.WhiteBalanceMode = 'manual';
src.Brightness = 90;
src.Exposure = -9;
% src.WhiteBalance = 6800;
src.WhiteBalance = 7200;
% src.WhiteBalance = 7500;
src.FocusMode = 'manual';
src.Focus = 18;
src.Sharpness = 20;
src.Saturation = 100;
src.FrameRate = num2str(fps);
%preview(obj)

%%
%get 1—frame snapshot for face recognition
start(obj)
%    tic,pause(2),toc,disp(' (pause 2 sec)')%time to stabilize
    trigger(obj)%sets startpoint of following recording
    frame=getsnapshot(obj);
stop(obj)

%face recognition works faster with a reduction in resolution
reduce=2;
img=double(rgb2gray(imresize(frame,1/reduce)));
Face=FaceDetect('haarcascade_frontalface_alt2.xml',img);
Face=reduce*Face;
```

```matlab
%set RegionOfInterest to expanded face—rectangle
%make sure dimensions are integer multiples of dx
dx=16; % x—dim of square
dy=dx; % y—dim of square
reducewidth=5/5;
increaseheight=6/4;
xres=dx*ceil(Face(3)*reducewidth/dx);
yres=dy*ceil(Face(4)*increaseheight/dy);
Face(1)=Face(1)+.5*(Face(3)—xres);
Face(2)=Face(2)+.5*(Face(4)—yres);
Face(3)=xres;
Face(4)=yres;
Face=round(Face);
obj.ROIPosition = Face;
frame=frame(Face(2):Face(2)+Face(4),Face(1):Face(1)+Face(3),:);
numBlocks=xres*yres/(dx*dy);

% create vector of square coordinates on form
% [x—pos top left, y—pos top left, dx, dy]
b=zeros(4,numBlocks);
b(1,:)=mod(1:dx:dx*numBlocks,xres);
b(2,:)=sort(mod(1:dy:dy*numBlocks,yres));
b(3,:)=dx*ones(1,numBlocks);
b(4,:)=dy*ones(1,numBlocks);

%%
%record numFrames number of frames
start(obj)
    tic,pause(2),toc,disp(' (pause 2 sec)')%time to stabilize
    trigger(obj)%sets startpoint of following recording
    tic,frames=getdata(obj,numFrames);toc,disp(' (data acq)')
stop(obj)
g=squeeze(frames(:,:,2,:));%2=green

%%
%compute average for each block and frame
G=zeros(numBlocks,numFrames);
for ii=1:numFrames
    for jj=1:numBlocks
        rect=g(b(2,jj):b(2,jj)+dy—1,b(1,jj):b(1,jj)+dx—1,ii);
        G(jj,ii)=sum(sum(rect))/(dx*dy);
    end
end
for ii=1:numBlocks, G(ii,:)=zmean(G(ii,:)); end
%%
close all

%compute PSDs
N=256;%2^nextpow2(numFrames);
f=60*linspace(0,fps,N);
p=abs(fft(G',N)).^2/N;
f=f(1:end/2);p=p(1:end/2,:);

%find peaks in interval (i45,i200)
i45=find(f>=45,1,'first');
i200=find(f>=150,1,'first');
for ii=1:numBlocks,[Y(ii) I(ii)]=max(p(i45:i200,ii));end
Ibpm=f(i45+I—1);
figure,hist(Ibpm(Ibpm≠0),500)

%removing histogram peaks at endpoints is found to improve result
I2=Ibpm(Ibpm≠min(Ibpm));
I2=I2(I2≠min(I2));
I2=I2(I2≠max(I2));
I2=I2(I2≠max(I2));

%FOUND PULSE
pulse=mode(I2);
```

```matlab
disp(round(pulse))

%display blocks where pulse was detected correctly
Icorrectpulse=find(round(Ibpm)==round(pulse));
Iwrongpulse=find(round(Ibpm)≠round(pulse));
Y(Iwrongpulse)=0;
Ibest=find(Y>2*std(Y));
Igood=find(Y>1*std(Y));
Iok=find(Y>0*std(Y));

figure, hold off
image(frame),axis equal,hold on

for ii=Iok
rectangle('position',b(:,ii),'facecolor','g')
end

for ii=Igood
rectangle('position',b(:,ii),'facecolor','y')
end

for ii=Ibest
rectangle('position',b(:,ii),'facecolor','r')
end
```
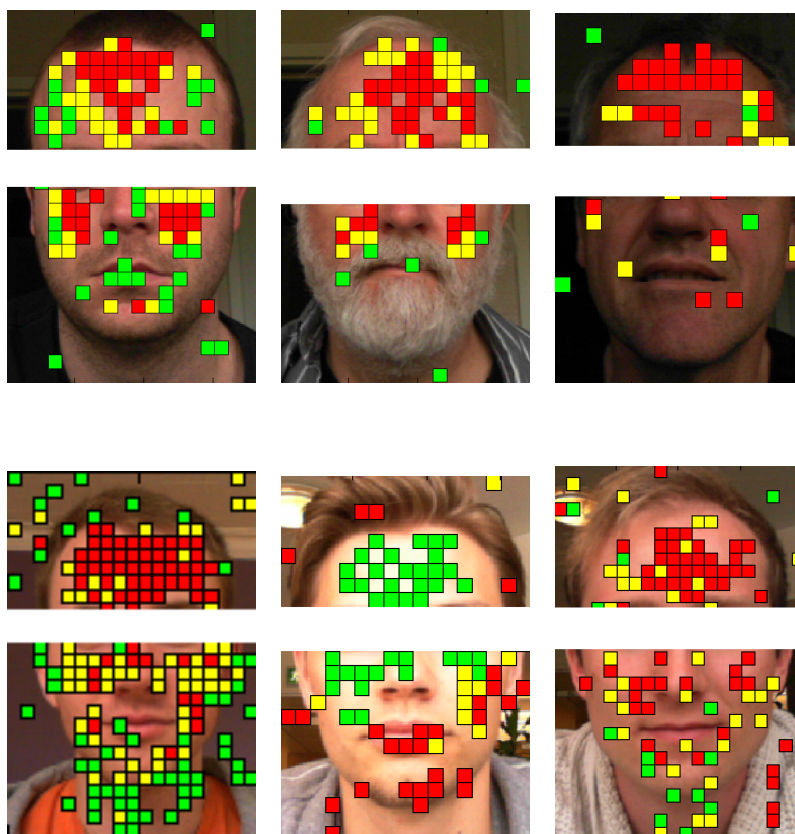
Appendix B

Other

B.1   Face Maps, other Subjects

B.2   Pulse Oximeter Specifications

Contec CMS50E, Contec Medical Systems Co., Ltd.   The pulse oximeter records a waveform with samplerate 60 Hz to a CSV file when connected to a computer. The CSV format was then converted to Matlab format.

Main Parameters (from user manual)

   Measurement of SpO$_2$ Measurement range: 0-100%, accuracy: 70-100%, $\pm$ 2%; 0-69%, unspecified.

   Measurement of pulse rate   Measurement range: 30-250 bpm, accuracy: ± 2 bpm or $\pm$ 2% (select larger).

   Resolution   SpO$_2$: 1%, pulse rate: 1 bpm.

   Resistance to surrounding light   The deviation between the value measured in the condition of man-made light or indoor natural light and that of dark room is less than ± 1%.

   Optical sensor   Red light (wavelength is 660 nm, 6.65 mW). Infrared (wavelength is 880 nm, 6.75 mW).

B.3   Web Camera Specifications

Microsoft LifeCam Studio

Imaging features (from technical data sheet)

   Sensor   CMOS sensor technology

   Resolution   Sensor resolution: 1920×1080, video resolution: up to 1920×1080.

   Field of view   75° diagonal field of view

   Frame rates   Supported frame rates are 7.5, 10, 15, 20 and at most 30 frames per second, depending on resolution.

   Other   Digital only pan, tilt and zoom. Auto focus from 10 cm. Automatic image adjustment with manual override.

B.4   Lists: Figures, Tables and Acronyms

List of Figures

List of Tables

Acronyms

ABP Arterial Blood Pressure

HR Heart Rate

HRV Heart Rate Variability

ICA Independent Component Analysis

JADE Joint Approximate Diagonalization of Eigenmatrices

MA Moving Average

PPG Photoplethysmography

PTT Pulse Transit Time

PTTD Difference in PTT

PWV Pulse Wave Velocity

RMSE Root Mean Square Error

ROI Region Of Interest

RR Respiratory Rate