



Norwegian University of
Science and Technology

Preprocessing of Arterial Pressure Waves for Later Use in Modeling of Individual Heartbeats

Tina Danielsen

Master of Science in Cybernetics and Robotics

Submission date: June 2018

Supervisor: Harald Martens, ITK

Co-supervisor: Nils Kristian Skjærvold, NTNU, St. Olavs
Kristin Tøndel, ITK

Norwegian University of Science and Technology
Department of Engineering Cybernetics

Problem description

Much of the research in product development in heartbeats have focused on modeling the heart rate, and in this study, the focus is modeling the shape of the heartbeats. Measurements of heartbeats can be performed in many different ways; Electrocardiography and Arterial Pressure waves (APW) (blood pressure curves). One part of this problem is to model APW and then try to find variations between persons and variations for one person. This study focus on one part of this problem. The focus is to perform preprocessing techniques to find baseline variations and finding peaks in APW. Furthermore, try to find the respiratory rate based on the APW.

Summary

Arterial pressure waves (APW) contain essential information, where the different shape of the waveform can indicate illness or essential information. Furthermore, there can be differences between subject independent of they are healthy or sick. The motivation for this thesis is to make individual models, such that the differences between subjects and the normality for one person is taken into account. In this thesis preprocessing of APW is performed and the purpose is to later use these preprocessing techniques before modeling individual heartbeats. The preprocessing have the goal of removing the baseline variations, remove the differences caused by different heart rate (HR), and separate the data into individual heartbeats.

The measurements of the data for seven healthy subjects in age 22-27 were performed by using The Human NIBP Set from ADInstruments, which is used to perform non-invasive blood pressure measurements from finger arterial pressure. The first step in the preprocessing was removing the baseline variations, and for this purpose, two different techniques were tried. The first technique was Savitzky-Golay (SG)-filtering and the second one was locally regression (LOESS). The second step was to interpolate the signal to get the same HR for all the subjects. The third step was to identify the systolic peaks and then at last separate the data into individual heartbeats. Besides, multi scatter correction (MSC) was tried as a preprocessing technique for removing the additive and multiplicative effect. The study ends with some principal component analysis (PCA) to study the possibilities of using the preprocessing techniques in further work.

The preprocessing with both LOESS and SG-filtering removes the baseline variations almost identically, and both methods give a satisfying result. However, the advantage with LOESS is the possibility to find several frequency components and that LOESS is a modeling method and not a filtering method. Furthermore, the result of the interpolation method shows a good result for the subjects with high HR, but not optimal for subjects with a HR around 1.2 Hz. The systolic peaks function work correctly, and it never fails to find the systolic peaks or classify other peaks to be the systolic peak. The separation function works good, and in most of the cases, the next and the last heartbeat are excluded from the heartbeat. The result of the MSC shows a good result for removing the additive and multiplicative effect, as long as there is not too much peak shifting.

Finding the respiration rate based on the baseline variations are hard since the baseline variations are influenced by the respiration rate, but are also other factors. This makes it hard to find the respiration rate only based on the baseline variations.

After the preprocessing a further analysis of the signals was performed, and then PCA was performed to check if individual difference appeared in the PCA. The PCA shows that there are variations between subjects and inside subjects. It also shows that outliers like premature ventricular contraction (PVC) heartbeats appear in the PCA model, which shows an opportunity to work further with the problem of finding an individual model for heartbeats.

Sammendrag

Arterielle trykkbølger (arterial pressure waves (APW)) inneholder viktig informasjon, hvor forskjellig bølgeform kan indikere sykdom eller annen viktig informasjon. Videre kan det være forskjeller mellom personer uavhengig av om de er friske eller syke. Motivasjonen for denne oppgaven er å lage individuelle hjertemodeller, slik at forskjellene mellom personer og normalitet for en person tas hensyn til. I denne oppgaven utføres preprosessering av arterielle trykkbølger og hensikten er å senere bruke disse preprosesseringsteknikkene før modellering av individuelle hjerteslag. Preprosesseringen har som mål å fjerne grunnlinjevariasjonene, fjerne forskjeller som skyldes forskjellig puls, og skille dataene i individuelle hjerteslag.

Målingene av dataene for syv friske personer i alderen 22-27 ble utført ved bruk av The Human NIBP Set fra ADInstruments, som brukes til å utføre ikke-invasivt blodtrykksmålinger fra finger arterielt trykk. Det første trinnet i forbehandlingen var å fjerne grunnlinjevariasjonene, og for dette formålet ble to forskjellige teknikker prøvd. Den første teknikken var Savitzky-Golay filtrering og den andre var lokal regresjon (LOESS). Det andre trinnet var å interpolere signalet for å få samme puls for alle personene. Det tredje trinnet var å identifisere de systoliske toppene og siste steg var å skille dataene inn i individuelle hjerteslag. Dessuten ble Multi Scatter Correction (MSC) prøvd som en preprosesseringsteknikk for fjerning av additiv og multiplikativ effekt. Studien avsluttes med prinsipal komponent analyse (PCA) for å se mulighetene for å bruke preprosesseringsteknikkene i videre arbeid.

Preprosesseringen med både LOESS og Savitzky-Golay filtrering fjerner grunnlinjevariasjonene nesten identisk, og resultatet er bra for begge teknikkene. Fordelen med LOESS er imidlertid muligheten til å finne flere frekvenskomponenter, og at LOESS er en modelleringsmetode og ikke en filtreringsmetode. Videre viser resultatet av interpoleringsmetoden et godt resultat for personene med høy puls, men ikke optimal for personer med puls rundt 1.2 Hz. Funksjonen for å finne systoliske toppe fungerer bra, og den klarer alltid å finne de systoliske toppene og klassifisere aldri en annen topp for å være den systoliske toppen. Separasjonsfunksjonen fungerer bra, og i de fleste tilfeller er det neste og det forrige hjerteslag ekskludert fra ønskede hjerteslag. Resultatet av MSC viser et godt resultat for å fjerne additiv og multiplikativ effekt, så lenge det ikke er for mye skift i toppene.

I denne studien ble det også forsøkt å finne respirasjonsfrekvensen basert på grunnlinjevariasjonene. Konklusjonen er at grunnlinjevariasjonene påvirkes av respirasjonsfrekvensen, men er også andre faktorer. Dette gjør det vanskelig å finne respirasjonsfrekvensen bare basert på grunnlinjevariasjonene.

Etter preprosesseringen kan en analyse av signalene utføres, og det ble dermed utført en PCA for å sjekke om den de ulike forskjellene mellom hjerteslag oppstod i PCA. PCA viser at det er forskjeller mellom personer og mellom hjerteslag for en person. Det viser også at ventrikulære hjerteslag skiller seg ut fra resten av hjerteslagene i PCA modellen, og dette viser store muligheter til å jobbe videre med dette problemet.

Preface

This is my master thesis spring 2018 at Norwegian University and Technology of Science as a part of the study program master in Cybernetics and Robotics with specialization Biomedical Cybernetics. The Department of Engineering Cybernetics have a project called Big Data Cybernetics, and based on this there was an idea from Harald Martens, my supervisor, to use the same techniques for heart modeling. Harald Martens is Adjunct Professor at Department of Engineering Cybernetics, and I want to thank Harald Martens for his inspiration and all his ideas.

The project and master thesis has been in collaboration with the master student Mikael Dyb Wedeld, where he has been working with electrocardiography (ECG), and I have been working with APW. The partnership was to collect data in fall 2017 for the project thesis, and we have shared our experiences in our work with the different signals. I want to thank him for sharing ideas and for all the support through this thesis.

This thesis is further work from the project thesis *Principal Component Analysis of the Arterial Pressure Wave* fall 2017. In this project, the data in this thesis was collected, and there was performed a PCA. Furthermore, there was stated some topics for future work, and some of these problems have been worked within this thesis. The peak finding algorithm was first used in the project thesis, but are further developed in this thesis. In the project thesis, the problem was that the parameters needed to be changed between different subject, and that was not ideal. Furthermore, the separation algorithm was developed in the project thesis, but are also further developed in this thesis. The problem with this algorithm used in my 2017 project was also the problem that the separation length needed to be changed for different subjects, due to various heart rate. The conclusion in the project thesis was that the PCA shows the difference in different subjects, however more preprocessing of removing outliers and baseline shift was needed to be addressed before more analysis could be performed.

I want to thank Nils Kristian Skjærvold, my co-supervisor, Researcher and Doctor at Department of Anesthesia and Intensive Care Medicine at Trondheim University Hospital for helping by lending equipment from Trondheim University Hospital to be able to collect data and help with medical questions. He has also been helpful to give me feedback on the master thesis.

Measurements of the data were performed by using The Human NIBP Set from AD-Instruments, which is used to perform non-invasive blood pressure (NIBP) measurements from finger arterial pressure. Matlab is used as the development tool for the preprocessing techniques, while The Unscrambler X have been the software to analyze the data with PCA and the preprocessing technique MSC.

I would also like to thank Silje S. Fuglerud, Ph.D. Candidate at the Department of Electronic Systems for her help with giving feedback to the thesis and sharing her experiences with writing a master thesis.

Tina Danielsen
Trondheim 4. June 2018

Innhold

Summary	i
Sammendrag	ii
Preface	iii
Table of Contents	vi
List of Figures	vii
Acronyms	viii
1 Introduction	1
1.1 Background and Motivation	1
1.2 Previous Work	2
1.3 Outline of the thesis	3
2 Theory	5
2.1 Ideal and Nonideal Spectral Measurements	5
2.1.1 Additive Problem	6
2.1.2 Multiplicative problem	7
2.1.3 Additive and Multiplicative Problem	7
2.1.4 Shifts	8
2.2 Preprocessing	10
2.2.1 Local Regression	10
2.2.2 Savitzky-Golay Smoothing Filters	11
2.2.3 Interpolation	14
2.3 Continuous Wavelet transform	15
2.3.1 The Analytic Wavelet Transform	15
2.4 Blood Pressure and Heart Theory	17
2.4.1 The Cardiac cycle	17

2.4.2	Arterial Pressure Waves	17
2.4.3	Methods of Blood Pressure Measurements	17
2.4.4	Premature Ventricular Contraction	19
2.5	Multivariate Data Analysis	19
2.5.1	Principal Component Analysis	20
2.5.2	Mean-Centering	20
2.5.3	Cross-validation	21
3	Approach	23
3.1	Data collection	23
3.1.1	Study population	23
3.1.2	Measurements	23
3.2	Preprocessing	24
3.2.1	Baseline correction	25
3.2.2	Finding Systolic Peaks	26
3.2.3	Interpolation to change the heart rate	26
3.2.4	Signal Separation	26
3.2.5	Multi Scatter Correction	26
3.3	Principal Component Analysis	27
4	Analysis	29
4.1	Preprocessing	29
4.1.1	Baseline Removal	29
4.1.2	Find Systolic Peaks	37
4.1.3	Interpolation and Heartbeat Separation	38
4.1.4	Heartbeat separation	40
4.1.5	MSC	40
4.2	PCA	40
5	Conclusion and Further Work	49
5.1	Conclusion	49
5.2	Future Work	50
	Bibliography	51
A	Appendix	55
A.1	Matlab Code	55
A.1.1	Blood Pressure Separation	55
A.1.2	Locate Systolic Peak	55
A.1.3	LOESS	56
A.1.4	SG-filtering	56
A.1.5	Interpolation	56

Figurer

2.1	Ideal and Nonideal Spectral Measurements	6
2.2	Mean versus signal with additive effect	7
2.3	Mean versus signal with multiplicative and additive effect in the data	9
2.4	Mean versus signal with shifts in the data	9
2.5	Normal Arterial Line Waveforms for blood pressure (BP)	18
2.6	Premature Ventricular Contraction in ECG (top plot) and BP (bottom plot)	19
4.1	Savitzky-Golay Filtering of subject 1	30
4.2	Savitzky-Golay Filtering of subject 2	31
4.3	The BP before and after SG-filtering.	32
4.4	LOESS model for subject 2 by 20 seconds of measurement 1	34
4.5	continuous wavelet transform (CWT) after LOESS model are subtracted for subject 2, 20 seconds in measurement 1	35
4.6	BP curve plotted before and after LOESS.	36
4.7	LOESS model for the whole 10 min dataset for subject 1	37
4.8	Find systolic peaks	38
4.9	Separated heart beats of 20 seconds.	41
4.10	Separated heart beats after interpolation for 10 minutes signal.	42
4.11	Separated heart beats after MSC.	43
4.12	PCA of subject 1.	45
4.13	PCA of subject 1 after MSC.	46
4.14	PCA of subject 2.	47
4.15	PCA of subject 2 after MSC.	48

List of Acronyms

BP blood pressure

HR heart rate

LOESS locally regression

CWT continuous wavelet transform

SG Savitzky-Golay

FIR finite impulse respons

PCA principal component analysis

MSC multi scatter correction

EMSC extended multi scatter correction

PLSR partial least square regression

APW arterial pressure waves

ICU intensive care uint

ECG electrocardiography

MVA multivariate data analysis

PC principal component

PVC premature ventricular contraction

Introduction

In this chapter the background and motivation for this thesis will be presented, followed by a brief description of previous work performed on the topic and the last part consist of the outline of the thesis.

1.1 Background and Motivation

BP measurement has clinically been used since around 1900. In the doctors' office a cuff on the arm is the standard way of measuring the BP and in the intensive care unit (ICU) invasive BP measurement, known as the APW at the monitor, is standard [6]. Even though the APW are measured, there is not much focus on the shape of the heartbeats. In [31] it is discussed the advantage of focus and understanding arterial pressure waves. The main point mentioned is that one should have a full understanding of normal waveform components, their relation to the cardiac cycle and the difference in waveforms recorded from different body sites. Besides, there are pointed out many different types of waveforms which origin in different diagnostics.

One of the challenges with the hospital today is the lack of individual models and control of the patients. Today monitoring has alarms based on what is healthy based on an average of parts of the population. However, all humans are different and what is normal for one person can be abnormal for another person. Another problem with today's alarms setting is that each of the monitoring channels, i.e., BP, ECG, temperature, respiration, HR, and oxygen saturation have separate alarm settings. There should be possible instead to use a technique that uses the different channels and make an individual model.

Making the future patient monitor with integrated clinical decision support and intelligent

alarms settings will both solve the problems of easier monitoring and control of patients. Included in this problem is follow the track of individual patients by unsupervised learning and compare diagnosis by supervised learning. Individual models with a smart alarm system will improve advanced health care with better control, minimizing adverse events, guide treatment and early diagnose detection with early targeted treatment. Included in this there should be an automatic outlier-detector. To achieve, a method like different machine learning techniques and continuous multivariate analysis are techniques that can solve these problems.

This thesis focuses on a small part of the problem with individual-based modeling of heartbeats an the study if it is possible to find individual differences in the shape of the APW. The thesis focus of preprocessing APW and this consists of removing baseline variations, interpolation and separate the data into individual heartbeats. There is a small part at the end of the thesis, where some PCA is performed to observe how individual differences will appear in a PCA model.

1.2 Previous Work

There are some few research which combining multivariate analysis and APW. However, there are some few research which uses PCA in parts of the signal. In [23] They used data from participants with normal blood pressure and participants with hypertensive blood pressure, and then use PCA to analyze the data. They used a sphygmomanometer to measure the blood pressure. In addition to the blood pressure, they had variables like weight, height, and BMI. In addition to the blood pressure, they had variables like height, weight, BMI, age and arm circumference. Another study [5] used PCA to find cardiovascular risk traits in three generations cohort among Indian Punjabi population. Both of these studies used PCA to find which factors are the most important concerning cardiovascular disease. However, these two studies have used the BP value and not the APW. Furthermore, in [39] and [20], they have used wave separation analysis to separate the pressure waves into forward and backward waves, while in [2] they investigate the accuracy of the wave equation in predicting arterial pulse propagation. Both of these studies are focusing on mathematical models.

In [3] they did research where the primary purpose was to use machine learning algorithms to deal with vectorized features extracted from APW. They performed some preprocessing work on the dataset to reduce the number of input features and increase the model accuracy by selecting the most relevant ones. For classification of the dataset, they used four machine learning algorithms: Random Forest, BayesNet, J48(decision tree) and RIPPER. One of the primary goals was to provide a multi-parametric approach capable of retrieving valuable information from arterial blood pressure waveforms. In opposite to traditional single parameter analysis where they focus on one component, instead of several components.

In [36] it is explained the problem with alarms in the ICU and that the problem is a sig-

nificant amount of false alarms. Artifacts cause many of the alarms, which are caused by example movements or that something is wrong with the monitor equipment. The problem is that most of these artifacts influence the measurements, which then triggers the alarm. In [22] they discuss the problems with different alarm algorithms and explain three levels where the alarms can be improved and which methodological criteria the monitoring alarm must fulfill. One observation is that most of the alarms only use a univariate alarm algorithm. However, when monitoring a patient, several different measurements are performed at the same time. Therefore they suggest using multivariate alarms. Multivariate can mean that either the results of univariate analysis are combined or that multivariate analyses like PCA or factor analyses are performed.

There is much research on using different types of artificial intelligence approach for construction of alarms systems. In [24] a knowledge-based alarms system during heart surgery for monitoring patients was tested. The overall performance was promising with low false positive rate, but it was only tested at few patients. Furthermore, in [16] a knowledge discovery based intelligent decision support system for intensive care medicine was presented. They used the technique to predict organ failure and mortality assessment.

In [9] they have performed a study with a wearable Multisensor in an outpatient glucose monitoring. It showed that the Multisensor and the algorithmic routine could track glucose trends in all patients both in controlled and uncontrolled conditions. One crucial part they discussed in this study is the importance of using multiple sensors to track activity and health parameters day and night over an extended period. Besides this, an important part is large battery capacity and future capability to monitor multiple parameters, including HR and its derivatives. Therefore, even though they do not use the ability for using multiple sensors yet, they can that this is the future and the importance of what this information together can give.

1.3 Outline of the thesis

This thesis is organized into five chapters with the following contents:

Chapter 2 is the background theory to understand the next chapters. The chapter starts with ideal and nonideal spectral measurements focused on how this can be used in the APW analysis. The next topic is preprocessing, where the first part is the theory behind the baseline removal techniques LOESS and SG-filtering and the next part is the interpolation. The next topic is the theory of BP and different methods for BP measurement. The last topic is multivariate data analysis (MVA), where the focus is PCA.

Chapter 3 is the approach chapter where the first part is the data collection approach. The next part is the preprocessing and the analysis and which methods was used in each of these parts. The preprocessing part starts with the baseline correction methods SG-filtering and LOESS, then it is a description of finding the systolic peaks function. Furthermore, the signal separation algorithm is described, and the last part is MSC. The chapter ends

with a description of how PCA is performed in Unscrambler.

Chapter 4 is the analyze chapter with results and discussion, where the first part is the preprocessing, and the last part is the PCA. First, there is the result and discussion about the baseline removal techniques SG-filtering and LOESS, then there is the result of finding systolic peak function. Furthermore, the result of the interpolation algorithm is shown together with the signal separation algorithm. The preprocessing parts end with the result of the MSC. The chapter ends with the result of the PCA before and after MSC.

Appendix A shows parts of the Matlab code developed in this thesis.

Kapittel 2

Theory

2.1 Ideal and Nonideal Spectral Measurements

In quantitative measurements, there are some different phenomena, which can create problems if they are not handled correctly. In spectral measurements, these effects can be minimized if the instruments and the procedures are optimized. However, sometimes this can be too expensive, too time-consuming or can introduce new artifacts. In that case, there are some different techniques to do this computationally. Different effects are additive effects, multiplicative effects, both additive and multiplicative effects, response curvature due to signal saturation, sideways shift and random noise. The different kinds of phenomena are presented in the next section based on [7]. Furthermore, these problems can also appear in APW, when each heartbeat are plotted on top of each other. To find out if the differences in individuals independent of additive and multiplicative effect, this is one way of understanding the problems.

In figure 2.1, the reference spectrum r and the unknown spectrum z are in each case compared, and it shows the different types of the unwanted phenomena. We assume for simplicity that $z_{True} = r$, is unknown to us and that the reference spectrum is equal to the true spectrum z_{True} . Further, we assume that the measured spectrum z has been affected by various unwanted and unknown phenomena during the measurement process:

$$z = f(z_{True}). \quad (2.1)$$

Then the correction can be found by estimating $f(\cdot)$ and inverting this function:

$$z_{Corr} = f^{-1}(z). \quad (2.2)$$

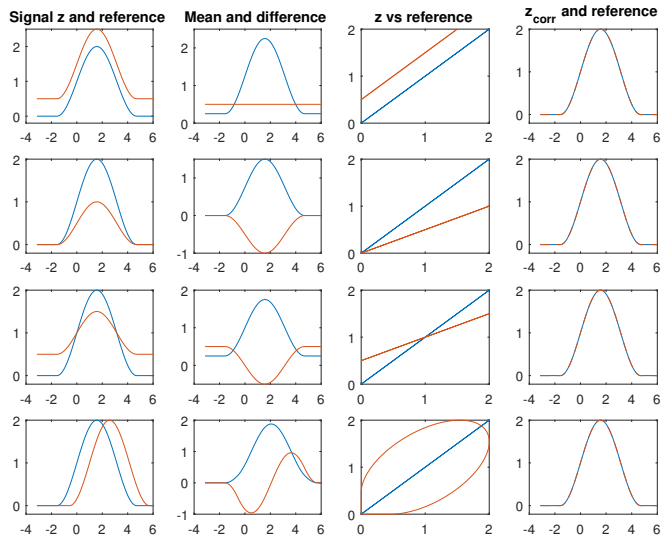


Figure 2.1: Ideal and Nonideal Spectral Measurements. The figure are reproduced in Matlab based on figure 1 in [7].

2.1.1 Additive Problem

The additive problem is due to baseline variations, which can be caused by an incorrect calibration before measurements or normal variations for a person, in BP measurements. In row one in figure 2.1 the theory behind the additive problem is shown and the model generating the measured z from the true z adds a fixed baseline effect of size a in z :

$$z = f(z_{True}) = z_{True} + a. \tag{2.3}$$

After the additive correction,

$$z_{Corr} = z - a \tag{2.4}$$

r and z_{corr} fall on top of each other as seen in figure 2.1 row 1 and column 4 as they should.

In the project thesis, it was showed that there is an additive effect in the APW due to variations in the baseline. However, the code for this is improved further in this thesis by adding the mean signal in the line plot. The plot constructed in the master thesis for showing the additive effect is shown in figure 2.2, and this shows almost a straight line in parallel with the mean, which shows an additive effect.

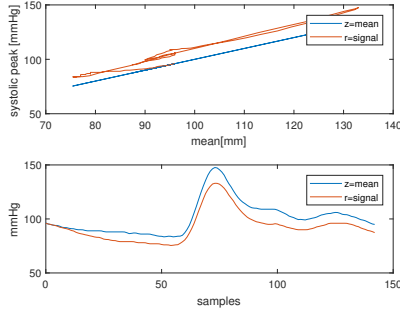


Figure 2.2: Mean versus signal with additive effect

2.1.2 Multiplicative problem

The Multiplicative problem is due to scaling variations, which can be caused by normal variations for a person or disturbances in the measurements. The theory behind this is shown in figure 2.1 row 2. The z in figure 2.1 row two and column one, were generated with a fixed scaling parameter b by

$$z = f(z_{True}) = bz_{True}. \quad (2.5)$$

The difference $(z - r)$ is shown in figure 2.1 row 2 and the blue curve, and it has the same shape as z and r

After the multiplicative correction

$$z_{Corr} = b^{-1}z, \quad (2.6)$$

where the r and z_{corr} spectra fall on top of each other in figure 2.1 row 3 and column 4 as they should.

2.1.3 Additive and Multiplicative Problem

The additive and multiplicative problem are due to both baseline and scaling variations occurring at the same time.

$$z = f(z_{True}) = bz_{True} + a \quad (2.7)$$

To deal with this problem a preprocessing technique MSC can be used, or if this is not good enough more advanced preprocessing like extended extended multi scatter correction (EMSC) can be used. The MSC is a transformation method used to compensate for additive and multiplicative effects in spectral data. This method is capable of correcting differences in the baseline and the trend and has an advantage that the transformed spectra are similar to the original spectral(most of the time) and that an optical interpretation is, therefore, more easily accessible [29]. If MSC is used the data z is represented by

$$z = f(z_{True}) = br + a, \quad (2.8)$$

and the spectrum are then corrected for by reordering equation 2.8:

$$z_{Corr} = b^{-1}(z - a). \quad (2.9)$$

A way to not remove the relevant information when using MSC is to use EMSC instead, which is an extension to conventional MSC. The performance of the methods can be improved if the user has prior knowledge about which parts of the signal contains essential information. Then these regions can be removed from the MSC correction to reduce the risk of removing relevant information. The way this is done is to down-weight the essential parts of the signals [10]. How multiplicative and additive effect occurs in APW is shown in figure 2.3, where it can be observed that the line is not parallel with the mean, but it is oblique with the mean.

2.1.4 Shifts

Figure 2.1 row 4 illustrates the problem with shifts in the signal. A sideways positional shift from spectrum z_{True} to spectrum z at channel j by $\delta(j)$ channels units is attained by

$$z(j + \delta(j)) = z_{True}(j). \quad (2.10)$$

If the size of the shift $\delta(j)$ is known or can be estimated, a shift correction can then be made by for example

$$z_{corr}(j) = z(j + \delta(j)) \quad (2.11)$$

A more precise shift may be by using continuous interpolation between the signals. Furthermore, shifts are a problem in APW, because the peaks are compressed and stretched based on the HR of the subject. How shifts appear in APW are shown in figure 2.4 and it can be observed that it is a small shift in the systolic peak.

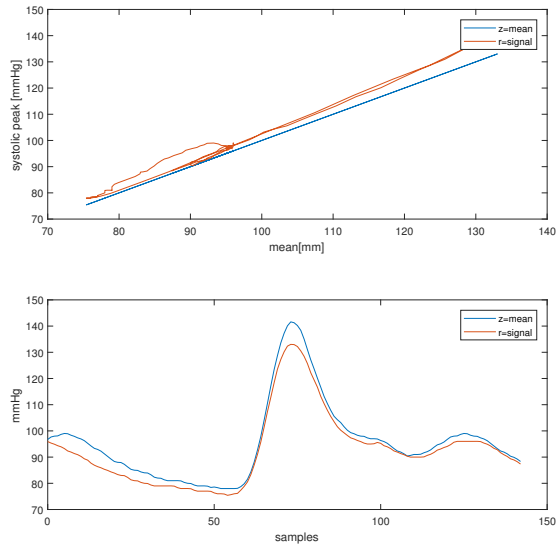


Figure 2.3: Mean versus signal with multiplicative and additive effect in the data

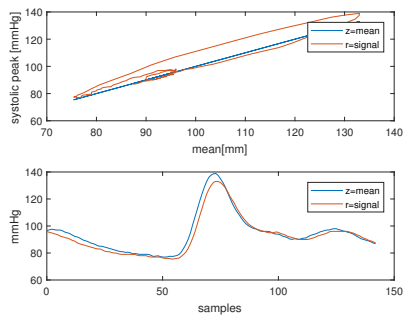


Figure 2.4: Mean versus signal with shifts in the data

2.2 Preprocessing

2.2.1 Local Regression

LOESS is a modeling method that is built on linear and nonlinear least squares regression. LOESS combines the simplicity of linear squares regression with the flexibility of nonlinear regression. LOESS does this by fitting simple models to localized subsets of the data and then find a function that describes the deterministic part of the variation in the data [14].

Definition of a LOESS model

LOESS was originally proposed in [11] and further developed in [12] where it was called locally weighted regression. At each point in the data set a low degree polynomial is fit to a subset of the data with an explanatory variable value near the point whose response is being estimated. An explanatory variable means that the variable is not independent for certain. The polynomial is fit by using weighted least squares, where points near the point estimated get more weight and point further away gets less weight. The value of the regression function for the point is obtained by evaluating for the local polynomial using explanatory for value variable for that data point. The locally fit of the data is done for each data point, and when this is completed for every n data point, and a LOESS model is completed. LOESS is a flexible method in the way that both the degree of the polynomial and the weight can be changed. How this is performed is discussed further in the text.

Localized Subsets of Data

The subsets of the data used for each weighted least squares fit in LOESS are determined by the nearest neighbor algorithm. A smoothing parameter, which is chosen by the user, determine how much of the data are used to fit each local polynomial. The smoothing parameter q is a number between $(d + 1)/n$ and 1, where d is denoting the degree of the local polynomial. Large values of q produce the smoothest function, where there are few wiggles. A small q produces a function with many wiggles since it then captures the small variation in the data. Too small q is not desirable since the regression function will then start to capture random noise in the data.

Degree of Local Polynomials

The local polynomials to fit the data are almost always the first or second degree, hence locally linear or locally quadratic. A zero degree polynomial turns LOESS into a weighted moving average. A higher degree of polynomial would work in theory, but this is not the

spirit of LOESS. LOESS is based on the idea that a low degree polynomial can approximate the function in a small neighborhood and the simple model can be fit to data easily.

Weight Function

The weight function gives more weight to the data points close to the point estimated and less weight to data points further away.

The traditional weight function used for LOESS is the tri-cube weight function,

$$w(x) = \begin{cases} (1 - |x|^3)^3, & \text{for } |x| < 1 \\ 0, & \text{for } |x| \geq 1. \end{cases} \quad (2.12)$$

In Cleveland(1979) properties for a weight function W satisfying the properties is listed;

1. $W(x) > 0$ for $|x| < 1$
2. $W(-x) = W(x)$
3. $W(x)$ is a nonincreasing function for $x \geq 0$
4. $W(x) = 0$ for $x \geq 1$.

Any function which satisfies these properties could be used as a weight function. The weight for a specific point in any localized subset of data obtained by evaluating the weight function at a distance between that point and the point of estimation, after scaling the distance so that the maximum absolute distance over all of the point in the subset of data is precisely one.

2.2.2 Savitzky-Golay Smoothing Filters

SG finite impulse responses (FIR) smoothing filters, also known as polynomial smoothing, or least-square smoothing filters, are a generalizations of the FIR average filter that can preserve better than the high-frequency content of the desired signal, at the expense of not removing as much noise as the average. These filters are typically used to smooth out a noisy signal whose frequency span(without noise) is significant [32].

The next theory is based on the book by Gander and Matt(1993) [17]. The fundamental idea for this filter is that the smoothed value g_i in the point x_i is obtained by taking an average f of its neighboring data. The simplest method consists in computing a moving average of a fixed number of f_i 's(the function to be smoothed). Furthermore, this can be performed more generally by fitting a polynomial through a fixed number of points. Then

the value of the polynomial at x_i gives the smoothed value g_i . Let n_L denotes the number of points to the left of x_i and n_R denotes the number of points right to x_i . By $p_i(x)$ we denote a polynomial degree of M which is fit in the least square through the $n_L + n_R + 1$ points. Then we have $g_i = p_i(x_i)$.

Filter Coefficients

The polynomial $p_i(x)$ of degree M , which is to be fitted through the data f_i , can be written as

$$p_i(x) := \sum_{k=0}^M b_k \left(\frac{x - x_i}{\Delta x} \right)^k. \quad (2.13)$$

It is assumed that x_i are uniformly spaced with $x_{i+1} - x_i \equiv \Delta x$. In order to fit $p_i(x)$ in the least squares sense through measured data, we have to determine the coefficients b_k such that

$$\sum_{j=i-n_L}^{i+n_R} (p_i(x(j)) - f_j)^2 = \min. \quad (2.14)$$

Let us define the matrix

$$A := \begin{bmatrix} (-n_L)^M & \cdots & -n_L & 1 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \\ \vdots & & \vdots & \vdots \\ n_M^R & \cdots & n_R & 1 \end{bmatrix} \in \mathbb{R}^{(n_L+n_R+1) \times (M+1)} \quad (2.15)$$

and the two vectors

$$\mathbf{b} := \begin{bmatrix} b_M \\ \vdots \\ b_1 \\ b_0 \end{bmatrix} \in \mathbf{R}^{M+1} \quad (2.16)$$

and

$$\mathbf{f} := \begin{bmatrix} f_{i-n_L} \\ \vdots \\ f_i \\ \vdots \\ f_{i+n_R} \end{bmatrix} \in \mathbf{R}^{(n_L+n_R+1)}. \quad (2.17)$$

Using these definitions, we can restate the least squares problem in equation 2.14 in matrix terms as

$$\|A\mathbf{b} - \mathbf{f}\| = \min. \quad (2.18)$$

It would be possible to solve equation 2.14 for \mathbf{b} by means of the QR-decomposition of A . For our filtering purpose, however, we only need to know $g_i = p_i(x) = b_0$. The solution of 2.14 can also be expressed as the solution of the normal equations

$$A^T A \mathbf{b} = A^T \mathbf{f}. \quad (2.19)$$

Thus we get

$$g_i = \mathbf{e}_{M+1}^T (A^T A)^{-1} \mathbf{f}, \quad (2.20)$$

where \mathbf{e}_{M+1} denotes the $(M+1)$ st unit vector.

We can then represent g_i as linear combinations of the f_i 's. We define the vector

$$\mathbf{c} := A(A^T A)^{-1} \mathbf{e}_{M+1} \quad (2.21)$$

containing the filter coefficients c_{-n_L}, \dots, c_{n_R} . Since \mathbf{c} does not depend on x_i and Δx , it only needs to be evaluated once. All the smoothed values g_i can be computed by simple scalar product

$$g_i = \mathbf{c}^T \mathbf{f} = \sum_{j=-n_L}^{i+n_R} c_{j-i} f_j. \quad (2.22)$$

The calculation of the vector \mathbf{c} according to Equation 2.21 may be inaccurate for large values of M . This loss of accuracy can be attributed to the fact that the condition number of A is squared in forming $A^T A$. The condition number of a function with respect to an argument measures how much output value of the function can change for a small change

in the input argument. On the other hand, it is well known that the least squares system in equation 2.14 can be stably solved using the QR-decomposition

$$A = QR. \quad (2.23)$$

By Q , we denote an orthogonal $n_L + N_R + 1$ -by- $(M + 1)$ matrix, and by R we denote an upper triangular $(M + 1)$ -by- $(M + 1)$ matrix. If we substitute decomposition 2.23 into of equation 2.21, we get for the vector c the expression

$$c = \frac{1}{r_{M+1, M+1}} Q e_{m+1} \quad (2.24)$$

2.2.3 Interpolation

In [35] interpolation in the field of mathematics is defined as a method of constructing new data points within the range of a discrete set of known data points. Furthermore, in engineering and science, one often have many data points, which represents the values of a function for a limited number of values of the independent variable. It is often required to interpolate the value of that function for an intermediate value of the independent variable.

Interpolating a set of data points can be done using polynomial, spline function or Fourier series. In [4] polynomial interpolation and interpolation by splines, i.e., piecewise polynomial interpolation are described. However, here the focus is on polynomial interpolation.

For $n \in \mathbb{N} \cup \{0\}$, we denote by P_n the linear space of polynomials

$$p(x) = \sum_{k=0}^n a_k x^k \quad (2.25)$$

for a real (or complex) variable x with real (or complex) coefficients a_0, \dots, a_n .

A polynomial $p \in P_n$ is said to be of degree n if $a_n \neq 0$. P_n is a subspace of the linear space $C[a, b]$ of continuously differentiable real (or complex) valued functions on $[a, b]$, where $a < b$. For $m \in \mathbb{N}$ we denote $C^m[a, b]$, the linear space of m times continuously differentiable real (-or comlex) valued functions on $[a, b]$.

Given that $n + 1$ distinct points $x_0, \dots, x_n \in [a, b]$ and $n + 1$ values $y_0, \dots, y_n \in \mathbb{R}$, there exists a unique polynomial P_n with the property

$$p_n(x_j) = y_j, j = 0, \dots, n \quad (2.26)$$

In the Lagrange representation, this interpolation polynomial is given by

$$p_n = \sum_{k=0}^n y_k l_k \tag{2.27}$$

with Lagrange factors

$$l_k(x) = \prod_{i=0, i \neq k}^n \frac{x - x_i}{x_k - x_i}, k = 0, \dots, n. \tag{2.28}$$

Lagrange discovered the representation in equation (2.27) in 1764, and it is simple and easy for theoretical investigations. However, it is not computationally stable enough for large n . However Newton obtained a quadratic formula for interpolation in 1676, and this is more practical for computational purposes. [4]

To define Newton's theorem, we state the following definition.

Definition 2.2.1. Given $n+1$ distinct points $x_0, \dots, x_n \in [a, b]$ and $n+1$ values $y_0, \dots, y_n \in \mathbb{R}$, the divided differences D_j^k of order k at the point x_j are recursively defined by

$$D_j^0 := y_j, j = 0, \dots, n, \tag{2.29}$$

$$D_j^k := \frac{D_{j+1}^{k-1} - D_j^{k-1}}{x_{j+k} - x_j}, j = 0, \dots, n - k, k = 1, \dots, n. \tag{2.30}$$

Newton's theorem states that for $n \geq 1$ the uniquely determined interpolation polynomial p_n of the Lagrange representation in equation (2.27)-(2.28) is given by

$$p_n(x) = y_0 + \sum_{k=1}^n D_0^k \prod_{i=0}^{k-1} (x - x_i). \tag{2.31}$$

2.3 Continuous Wavelet transform

2.3.1 The Analytic Wavelet Transform

A wavelet $\psi(t) \in L^2(\mathbb{R})$ is an analyzing function used to localize a signal simultaneously in time and frequency. By definition, a wavelet has zero mean and finite energy, and additionally satisfy the "admissibility condition"[27]

$$\int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \quad (2.32)$$

where $\Psi(\omega)$ is the Fourier Transform of the wavelet. The wavelet transform of a signal $x(t) \in L^2(\mathbb{R})$ is a series of projections onto the rescaled and translated version of $\psi(t)$,

$$W_\psi(t, s) \equiv \int_{-\infty}^{\infty} \frac{1}{s} \psi^* \left(\frac{\tau - t}{s} \right) x(\tau) d\tau, \quad (2.33)$$

where t is the time parameter, s is the scale parameter, and $*$ represents the complex conjugate. Furthermore, analytic wavelets are defined for positive frequencies, such that $\Psi(\omega) = 0$ for $\omega < 0$. In the frequency domain, this is represented by

$$W_\psi(t, s) = \int_0^{\infty} \Psi^*(s\omega) X(\omega) e^{i\omega t} d\omega \quad (2.34)$$

A family of analytic wavelets are the generalized Morse wavelets. They are useful for signals that both vary in time and frequency.

The Fourier transform of the generalized Morse wavelet is

$$\Psi_{P,\gamma}(\omega) = U(\omega) a_{P,\gamma} \omega^{\frac{P^2}{\gamma}} e^{-\omega^\gamma}, \quad (2.35)$$

where $U(\omega)$ is the unit step, $a_{P,\gamma}$ is the normalizing constant, P is the time-bandwidth product, and γ is the symmetry of the Morse wavelet. The two parameters symmetry γ and time-bandwidth P can be adjusted to obtain analytic wavelets with different properties and behavior. These two parameters determine the wavelet shape and affect the behavior of the transform. γ controls the symmetry of the wavelet in time through the demodulate skewness, and P is proportional to the wavelet duration in time. The duration determines how many oscillations can fit into the time-domain wavelet's center window at its peak frequency. The peak frequency is

$$\left(\frac{P^2}{\gamma} \right)^{\frac{1}{\gamma}}.$$

The demodulate skewness of Morse Wavelet is the normalized third central moments $\alpha_{3;\phi}$ and for generalized Morse wavelets this becomes

$$\alpha_{3;\beta,\gamma} = i \frac{\gamma - 3}{P_{\beta,\gamma}}, \quad (2.36)$$

where $P_{\beta,\gamma} = \sqrt{\beta\gamma}$. We can see that when $\gamma = 3$ and $\alpha_{3;\beta,\gamma} = 0$ [26]. In study CWT are used for showing which frequency component we have for a certain time, and then it is important that both frequency and time are optimal.

2.4 Blood Pressure and Heart Theory

2.4.1 The Cardiac cycle

The cardiac events that occur from the beginning of one heartbeat to the beginning of the next are called the cardiac cycle. The cardiac cycle consists of a period of a relaxation called diastole, during which the heart fills with blood, followed by a period of contraction called systole [19].

2.4.2 Arterial Pressure Waves

In figure 2.5 the arterial line waveform is shown, and this is the APW, which travels much faster than the actual blood which is ejected. It represents the impulse of left ventricular contraction [40]. The systolic upstroke is the ventricular ejection, and the peak systolic pressure (also known as the systolic BP) is the maximum pressure generated during the systolic ejection. The systolic decline is the rapid decline in arterial pressure as the ventricular contraction comes to an end and the dicrotic notch represents the closing of the aortic valve. The diastolic runoff is the rapid decline in arterial pressure as the ventricular contraction comes to an end, while the end-diastolic pressure is the pressure exerted by the vascular tree back upon the aortic valve (know as the diastolic pressure). The different parts of the arterial wave contain important information, that can explain much of the cardiac health. In [31] there is explained some diagnostic, where the waveform is different from the normal shape shown in figure 2.5 [19]. The diagnostic are for example aortic stenosis, aortic regurgitation, hypertrophic cardiomyopathy.

2.4.3 Methods of Blood Pressure Measurements

There is both invasive and non-invasive method for measuring the BP, and in the next section, the invasive and the non-invasive method with finger cuff is explained. However, there is a lot of other non-invasive methods like ultrasound techniques, auscultatory method, and the oscillometric technique.

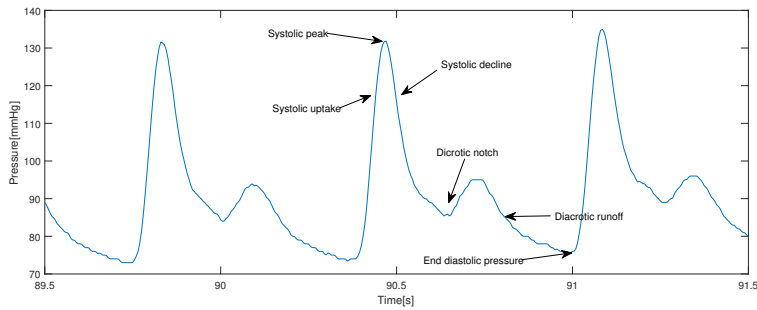


Figure 2.5: Normal Arterial Line Waveforms for BP

Invasive Method

Invasive blood pressure monitoring is a commonly used technique in the ICU and is also often used in the operating theatre. This technique involves direct measurement of arterial pressure by inserting a cannula needle in a suitable artery. The cannula must be connected to a sterile, fluid-filled system, which is connected to an electronic patient monitor. The advantage of this system is that a patient's BP is constantly monitored beat-by-beat, and the APW are shown on the monitor [30].

The finger cuff method

A noninvasive recording of BP at finger level was first developed in [34] and was further developed in [38]. That is the FINAPRES method, which was compared with the invasive method by [33]. They discovered that the beat-to-beat changes in finger systolic and diastolic BP were on average similar to beat-to-beat changes measured intra-arterial. Furthermore, the beat-to-beat BP recording via FINAPRES provides an accurate estimate of means and variability of radial BP. In most cases, this is an acceptable alternative to invasive BP monitoring during laboratory studies.

Factors that can affect blood pressure readings

During a BP measurement, many factors can influence the BP measurements, which can result in the wrong result. When measuring the BP, the patient needs to sit entirely still and sit correctly. Factors that can affect the BP are the emotional state, talking, alcohol, smoking, etc. Since all these factors can influence the BP, it is essential to take these factors into account when measuring BP [15].

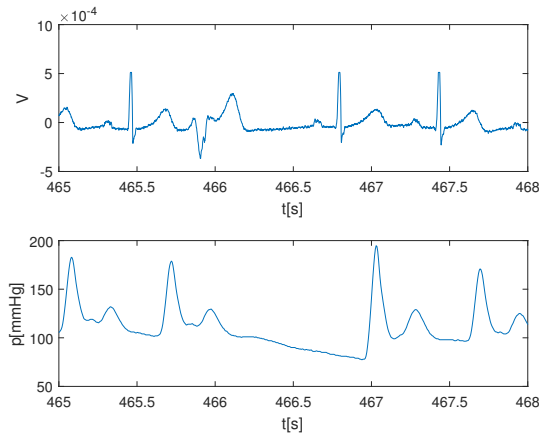


Figure 2.6: Premature Ventricular Contraction in ECG (top plot) and BP (bottom plot)

2.4.4 Premature Ventricular Contraction

PVC are premature heartbeats originating from the ventricles of the heart. Premature ventricular contractions are premature because they occur before the regular heartbeat. PVC is common, and many people have premature ventricular contractions without any symptoms. Premature ventricular contractions may be more common among older patients, patients with high BP, and patients with heart disease. Premature ventricular contractions can also occur in young healthy individuals without heart disease or high BP [25]. How this appears in the ECG and in the APW is shown in figure 2.6.

2.5 Multivariate Data Analysis

MVA is the investigation of many variables, simultaneously, to understand the relationship between them. MVA can analyze data with variables from two to millions of variables. So MVA gives us an opportunity to analyze a significant amount of data in a short time. This type of technique is different from the traditional statistical way of doing it, where one variable is compared with one variable. With MVA relationships of the data that was not known beforehand can be found.

MVA is soft modeling tool, where a model is based on the data and no assumptions are made. As opposed to mathematics and physics, where a model is made based on a simplified version of the world with many assumptions. The procedure is executed, and the result is analyzed based on the graphical view. The advantage of MVA is first that it identifies variables that contribute most to the overall variability in the data. Second, it helps to isolate the variables that are related, so they co-vary with each other. Third, the possibility of

graphical presentation, which for humans means a lot. Further, in the chapter, the different types of tools in MVA are presented [8].

2.5.1 Principal Component Analysis

PCA is the analysis of variability in a particular set of data. It is an essential part of MVA and its the first step of MVA. PCA can be presented graphically for understanding the relationship between samples and variables [8]. PCA can be used for any data set which has more than one variable, but it is best suited for data set with more than 10 variables and 100 samples [8]. PCA has a broad range of applications like exploratory data analysis, classification, and identification, variable reduction, etc.

PCA use principal component (PC) to separate the data. A PC is a latent variable, which means that the variable cannot be measured directly. Latent variables are computed as linear combinations of a set of input variables, which can be computed directly [21]. A PC defines which variables most contribute to the greatest sources of variability, shows which variables are related to each other and which ones are not and describes whether a particular set of variables are describing the important structure or are just random noise [8]. The goal is to have the fewest possible PCs without loss of essential information. The first PC is the one that contributes the most to the variability and PCs reflect the largest eigenvalues of the co-variances in the data [21]. PCA tries to find as few PC as possible to represent the data, which gives PCA the opportunity to reduce the dimension of the data. Usually, each PC does not correspond to an individual physical phenomenon. However, seen together, the first few PCs usually span the important information space of how physical phenomena vary [21]. The equation for the PCA model is

$$\mathbf{X} = \mathbf{t}\mathbf{p}' + \mathbf{E}, \quad (2.37)$$

where \mathbf{t} are the scores, \mathbf{p} are the loadings and \mathbf{E} are the residuals and where \mathbf{t} , \mathbf{p} and \mathbf{E} are unknown [21]. Loadings are the correlation between variables, while scores are the correlation between the samples.

2.5.2 Mean-Centering

Mean-centering is necessary for variables with higher values, not are weighted higher than the rest. In multivariate data analysis, we are usually interested in the patterns of variations between the samples, not in absolute numbers. Mean centering is, therefore, a common practice in MVA. This means that the input data \mathbf{X} are first modeled regarding the mean for the variables, plus an initial residual \mathbf{E}_0 [21]. This is computed by

$$\mathbf{E}_0 = \mathbf{X} + 1\bar{\mathbf{x}}. \quad (2.38)$$

2.5.3 Cross-validation

Validation is important for ensuring the quality of the model and the method presented here is cross-validation. In cross-validation we leave out a part of the data, and make submodels to all of the data has been taken out once. We need to set a validation level, example 5% of the data. If there is no obvious groups of samples or background knowledge, random cross-validation is the best choice [21]. Other cross-validation methods are k-fold cross-validation and leave-k-out cross-validation.

Kapittel 3

Approach

3.1 Data collection

The data collection was done in November in work with the project thesis and is presented in the next section.

3.1.1 Study population

The study population was students in the age of 22-27, and they had no known heart problems or other important health problems. It was a total of 7 participants, where there were five male and two females. The study was approved by REK (Regional Committees for medical and health research ethics). All participants signed an informed consent agreement, and all of the subjects are identified with a number.

3.1.2 Measurements

Measurements of the data were performed by using The Human NIBP Set from ADInstruments, which is an equipment to perform non-invasive blood pressure (NIBP) measurements from finger arterial pressure. The equipment uses a finger cuff to measure the finger arterial pressure at the annular finger. There were three different sizes of the finger cuff, to fit all the subjects finger size [1]. The equipment used to measure is supposed to represent the pressure wave in the same way as the invasive measurement.

The measurements were performed for each of the subjects in the same way:

1. Test person at rest for 5 min before test start
2. Do an ECG and a blood pressure measurement at the same time for 10 min while the person rests, while sitting on a chair.
3. Repeat 4 times per person

Most of the measurements were performed on the same day and right after each other with a small break between each of the measurements. Before the measurements, there was no restriction about food, training, coffee, etc. We know for sure that some of the subjects was training and drinking coffee right before. All of these factors can influence the result, and since all the measurements were done in approximately one hour, there can be a difference between the last and the first measurement since they have relaxed longer at the last measurement.

We had 7 test persons, and each measurement was repeated four times per subject, and this resulted in 28 measurements. The APW were measured with a cuff on the finger as described in background chapter. The problem with the measurement in this experiment was that the calibration equipment for the BP did not work, so the measurement is not correct in value. This gave the problem that for each new experiment the reference level varied a lot from one measurement to the next. From one measurement to the next the middle value varied from 0-40 mmHg for some of the subjects, which resulted in an not accurate BP measurement at all. However, the problem was to analyze the shape of the heartbeat and not the actual value, and therefore the relative value is important an not the actual value.

3.2 Preprocessing

In all of the preprocessing steps, except for MSC, Matlab version R2017a is used to do the preprocessing, where parts of the code are shown in the appendix. MSC is performed in Unscrambler® X, which is a complete multivariate data analysis and experimental design software solution, equipped with powerful methods including PCA, partial least square regression (PLSR), clustering, and classification [10]. One of the goals is to do the preprocessing in real-time and therefore the preprocessing is in general performed in a 20 seconds window, but the result is also shown for the whole signal in some cases. This is not completely real-time, but it is better than using the whole signal. Some of the preprocessing (i.e., systolic peak separation and MSC) shows the result of the whole signal.

Preprocessing of the signals have been a large part of the analysis, and the following steps are performed:

1. Baseline correction by LOESS/SG smoothing filters
2. Interpolation to change heart rate

3. Finding systolic peaks
4. Signal separation
5. MSC

The different steps are explained in more detail in the next sections.

3.2.1 Baseline correction

The BP varies over time, as explained in theory chapter since it is dependent on where in the cardiac cycle the heart is at a particular time. In chapter 2.1 ideal and nonideal spectral measurements are explained, explicitly additive and multiplicative effect.

Removal of the additive effect, due to variations in the baseline caused by the respiration rate and other unknown factors, was performed different methods for trying to find the baseline variations and correct for this. LOESS and SG smoothing filters are the two methods used, and these are presented in the next sections.

LOESS

The Matlab function *smooth* with input parameters *data*, *span* and *method* is used to find the LOESS model. *Data* is which data to perform the smoothing at, the *span* is how much of the data to use to find the model, and LOESS was the method used. Like described in theory chapter large *span* gives the low-frequency components, including the baseline drift, and a low *span* gives the high-frequency components. The baseline correction with LOESS is shown in appendix A.1.3. The output of the algorithm is a baseline corrected BP, and this algorithm also gives a model for each of the frequency signals, i.e., LOESS model 1, 2 and 3. These signals are important to keep, such that we do not lose information we may want to use in later analysis. This information may keep interesting information such that heart rate, respiration rate or other relevant information. The *span* values are dependent on how large the original data is. If the full signal at ten minutes are smoothed a smaller value of *span* are used than if 20 seconds of the signal is used.

Savitzky-Golay smoothing filter

SG are another method to find the baseline variations and here are the Matlab function *sgolayfilt* with the input parameters *Data*, *Polynomial order* and *framelength* used. It works by finding the baseline by use of the *sgolayfilt* and then subtract this from the original raw data. The function is shown in the appendix A.1.4.

3.2.2 Finding Systolic Peaks

To find the systolic peaks, which is the highest peak as explained in Theory Chapter, the Matlab function *findpeaks* with input parameters *data* and *Minimum peak prominence* is used. The minimum peak prominence is the peaks that have a relative importance of at least the *Minimum peak prominence*-value. Also, to find the peaks the mean heart rate is found by finding the interval between peaks and taking the mean of this. The function is shown in appendix A.1.2.

3.2.3 Interpolation to change the heart rate

The HR varies from person to person and from situations to situations, which makes it complicated to compare the shape of the heartbeats independent of the HR. So to compare the heartbeats in different settings, the HR is interpolated such that the signals are not dependent on the HR and that all have the same constant HR, which is set to 72 beats/min (1.2 Hz). In section 2.1 there is explained about shift, and this is one way of trying to fix the shift problem. The interpolation was performed with the Matlab function *interp1*, which has the input *old time vector*, *data*, *new time vector* and *method*. The method used is a linear interpolation. In order to find the correct interpolation frequency the average HR is used to find the right new sampling frequency. The mean HR are found by finding systolic peaks and then finding the time between each peak. The interpolation method is shown in appendix A.1.5, where there can be seen that how to choose the new sampling frequency are dependent on the mean HR is below or above the wanted 1.2 Hz.

3.2.4 Signal Separation

A separation of the BP curve into one heartbeat put on top of each other needs to be performed before a PCA can be performed to do the analyze. This algorithm was developed in the project thesis, and are further developed in the master thesis. The algorithm, shown in appendix A.1.1, uses the *findpeaks* as described in section 3.2.2 and from each side of the systolic peak a constant number of samples included in that specific heartbeat.

3.2.5 Multi Scatter Correction

A description of how MSC works are explained in the theory section 2.1, and then it is explained that it can correct for additive and multiplicative effects. The MSC of the data was performed in Unscrambler® X.

3.3 Principal Component Analysis

PCA is performed in Unscrambler® X to analyze the data after preprocessing. NIPALS algorithm is used to find the PC in Unscrambler® X, and the validation method is the cross-validation method random 20 segments method used to validate the data.

Kapittel 4

Analysis

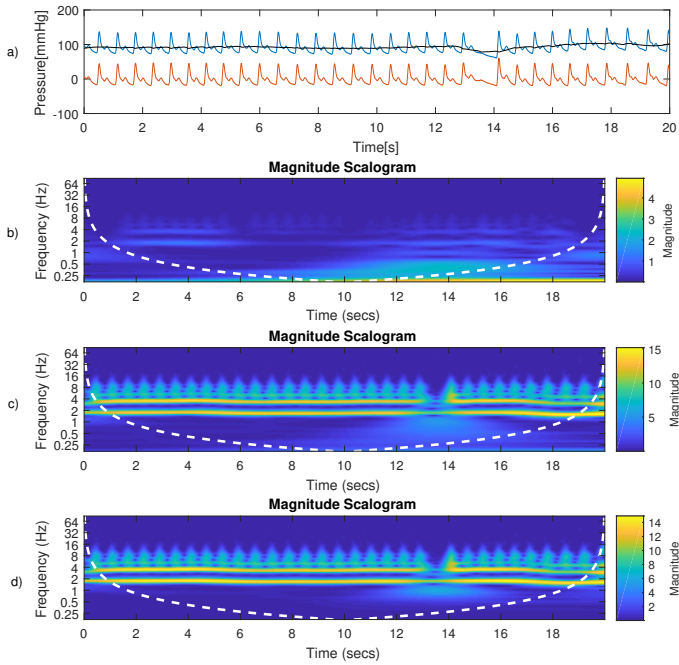
4.1 Preprocessing

To analyze a signal an important part is to preprocess the data. Different kinds of signals need different kinds of preprocessing, so to find the best and a good way to preprocess the data, different methods were used and tested. In the next sections, SG smoothing filters and LOESS is presented, and both are used for finding the baseline of the signal and correct for the baseline. CWT is used to illustrate the effect of removing the low frequencies in the signals. The Matlab function *cwt* with input parameters *data* and sampling frequency is used to find the the CWT. *cwt* use as default the analytic Morse Wavelet with symmetry parameter of 3 and a time-bandwidth of 60.

4.1.1 Baseline Removal

Savitzky-Golay Smoothing Filters

Figure 4.1 and figure 4.2 shows the APW before and after the SG filter is performed. How the SG filter work is presented in theory. In this analyze the *sgolayfilt* function in Matlab was used to find the baseline of the data. The input for this function is data, polynomial order, and frame length. Experimentation with different size of frame length and polynomial order was performed to find a good size. To not change the form of the signal the frame length needed to be high and frame length on 501 was used and a polynomial order of 3 was used. Furthermore in figure 4.3 the BP before and after SG filtering for each of the subjects in a 20 seconds window is shown. We can observe that the filter finds the baseline shift and then the baseline is subtracted from the BP signal. We can observe that this works



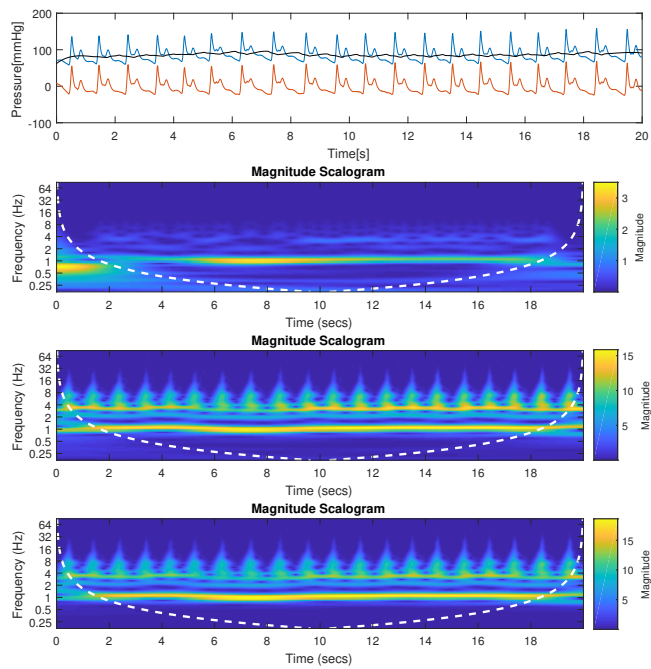
Figur 4.1: (a) 20 seconds of BP curve before and after SG filtering of measurement 1 of subject 1. The blue curve is the BP signal, the green curve is the baseline, and the red curve is after the baseline subtraction from the BP curve (b) shows CWT of the baseline curve (c) shows CWT of the BP curve before filtering (d) shows CWT of the BP curve after filtering

since we find the baseline and it is shifted down such that the baseline is at 0 instead of at 100.

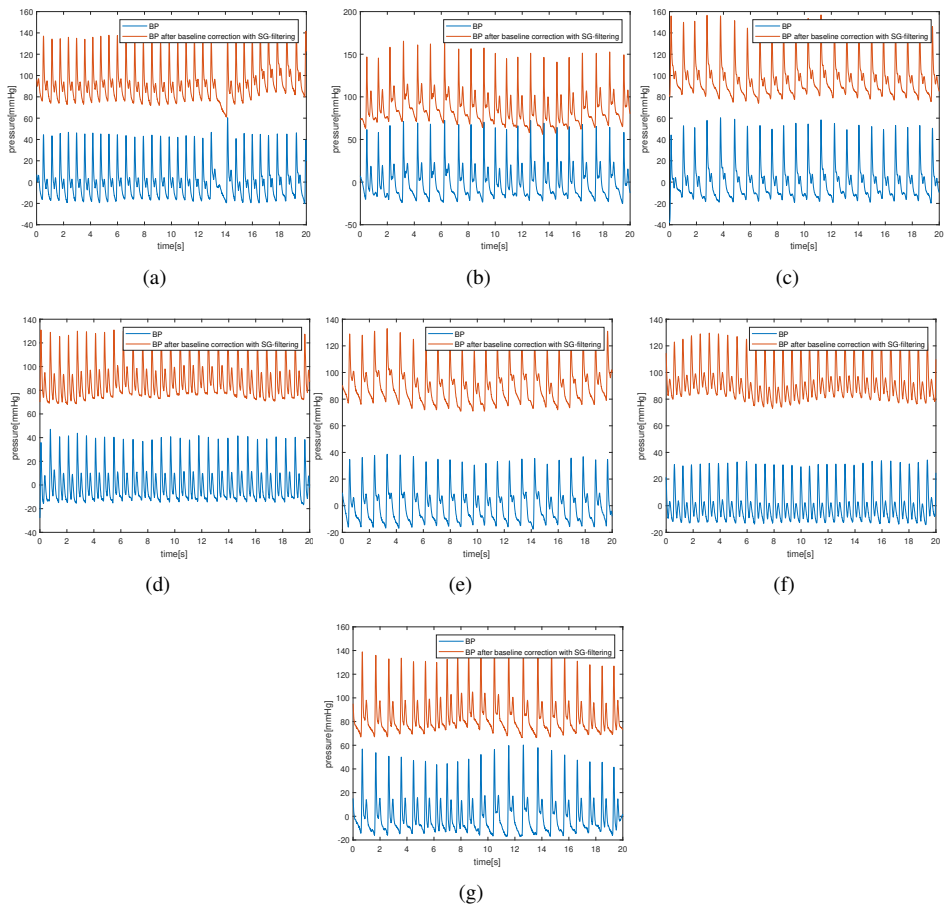
SG-filtering is an easy method to implement and gives a good result and do not change the signal in this case, but there are signals where the SG-filtering can change the signal. One of the downsides with SG-filtering is that it is a filtering method and not a modeling method.

LOESS

In the theory chapter the theory of how LOESS works are explained and here it is used to find different frequency components in the signal. Figure 4.4 shows the different LOESS models and figure 4.5 shows the corresponding CWT of the models. In able to find a good model it was much tuning of the span parameter. In figure 4.4 a span of respectively 0.5,



Figur 4.2: 20 seconds of the BP curve with SG-filtering for subject 1 measurement 1, a) BP curve (blue, the upper curve), the black curve is the baseline curve found by SG-filtering and the red curve (the bottom curve) is the BP curve after the baseline is subtracted from the blue BP curve b) the CWT of the baseline curve, c) CWT of the BP before filtering and d) CWT of the BP after the baseline is subtracted from the BP curve.



Figur 4.3: The BP before and after SG-filtering: (a) subject 1; (b) subject 2; (c) subject 3; (d) subject 4; (e) subject 5; (f) subject 6; and, (g) subject 7.

0.2 and 0.05 is used from the three LOESS models.

In the first LOESS model the desired result was to find the baseline variation and frequency, so therefore one half of the signal was used to find this trend. LOESS model 1 is as shown in figure 4.4 a low-frequency signal, and this low frequency could be the respiratory rate.

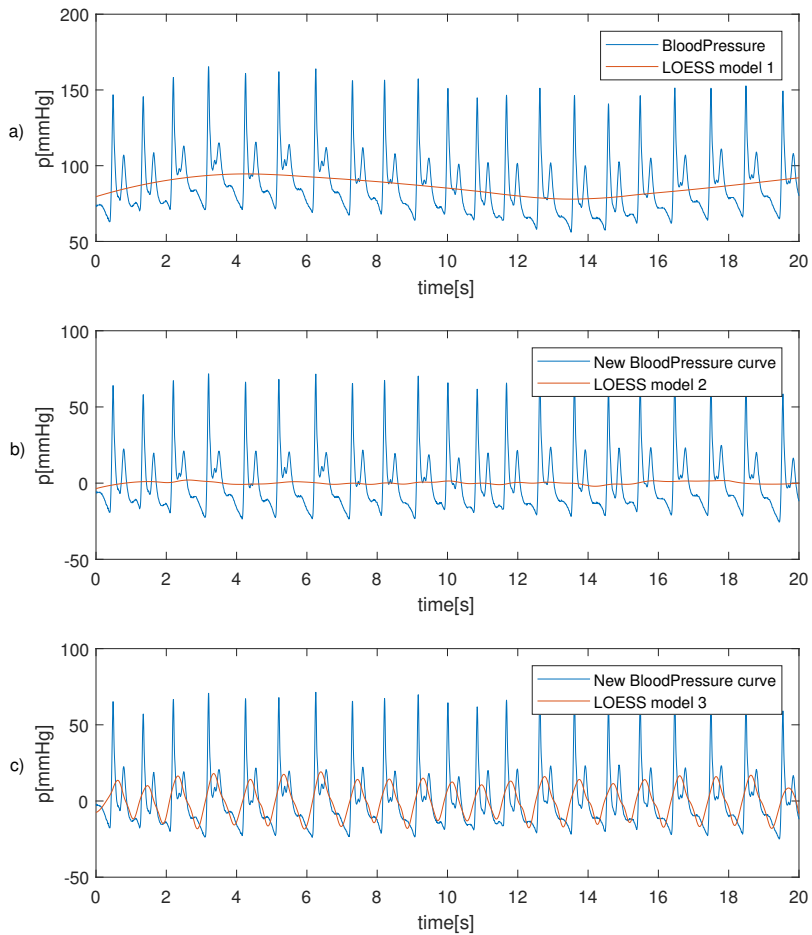
Further, the LOESS model 2 is also a low-frequency signal, however with a higher frequency. These two models are subtracted from the original model, such that the low-frequency components are no longer a part of the signal. In figure 4.5 CWT of the after the LOESS models are subtracted from the signal. The last plot shows that there are almost no longer frequency components lower than 0.5 Hz. The LOESS model 3 in figure 4.4 have a much higher frequency, and we can observe that this is a sinus wave with an almost perfect fit of the frequency of the heart rate. Figure 4.5 there is a strong yellow line around 1 Hz, which corresponds to the HR. Furthermore, this shows that by the LOESS technique there can found both high and low-frequency components, and remove the low-frequency components, which is the baseline variations.

In figure 4.6 the 20 second of the APW curve are plotted before and after LOESS. An observation is that the baseline is moved from around 100 mmHg to 0, which not cause a problem since we are not interested in the actual value, but the shape. If the value were necessary, this would cause some problems if the real value is not preserved, but we are preserving the model, such that it can be used for later analyzes. Furthermore, we can observe that the baseline for all the subjects are found and then subtracted and the result is a new BP curve without baseline shifts.

Figure 4.7 shows the ten minutes APW before and after LOESS for subject one. The observation is that LOESS finds the baseline variations, and remove these variations. The frequency of this baseline variate and are not consistent. One desired result was to find the respiratory rate from APW, but as observed in figure 4.7 this is hard. One of the problems is that the baseline varies with respiration and other factors like talking and movements. Besides, respiratory rate can be self-controlled, which makes it even harder to find it based on a curve.

LOESS versus SG-filtering

A comparison between LOESS and SG-filtering based on figure 4.6 and 4.3, shows that they give almost the same result of baseline removal without change the form of the signal, and they both compensate and find the PVC beats. The problem with both LOESS and SG-filtering is they both need a window to find the baseline components, and they can not do it in real time. However, when an analysis is performed, there will not be in absolute real time, and some time delay will appear. Both SG-filtering and LOESS has the advantage that they are simple methods to implement and easy computationally since they only use polynomials to find the baseline models. LOESS is computationally easy for the short signal 20 seconds signal, but finding a model for the whole 10 minutes signal was



Figur 4.4: LOESS model for subject 2 by 20 seconds of measurement 1: a) BP curve(the blue) before any preprocessing and the red curve is LOESS model 1, b) BP curve(the blue) before after LOESS 1 is subtracted from the original BP and the red curve is LOESS model 2 and c) The 'the blue curve is the BP after LOESS model 2 are subtracted from the BP curve and the red curve is LOESS model 3.

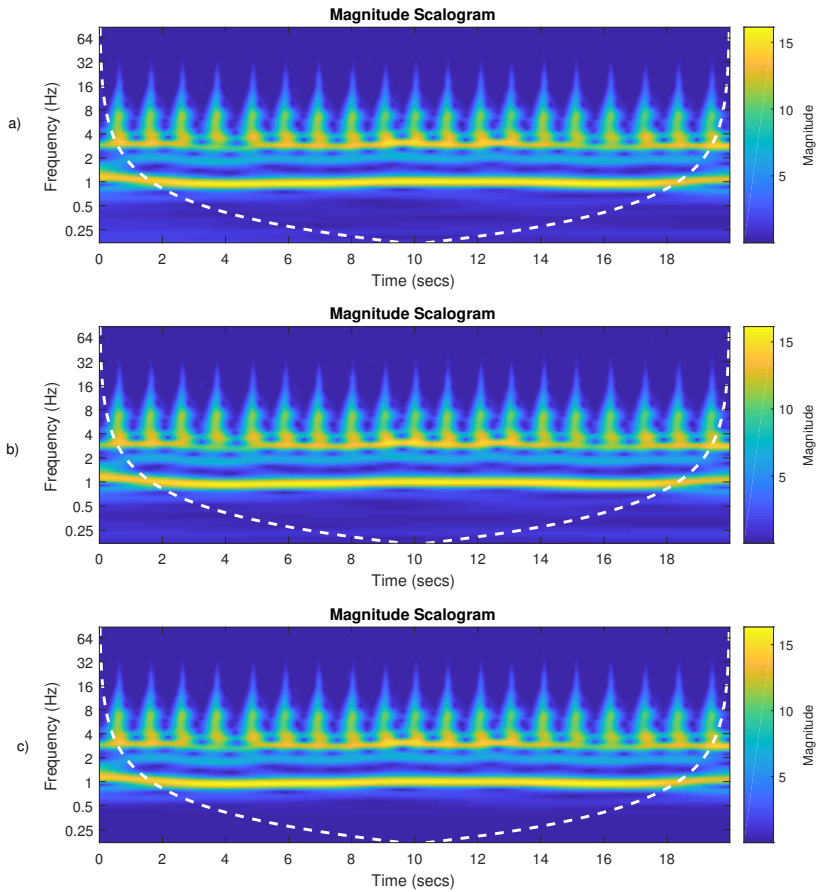


Figure 4.5: CWT after LOESS model are subtracted for subject 2, 20 seconds in measurement 1: a) BP before LOESS, b) BP after first LOESS model is subtracted and c) BP after second LOESS model is subtracted and this is the final model

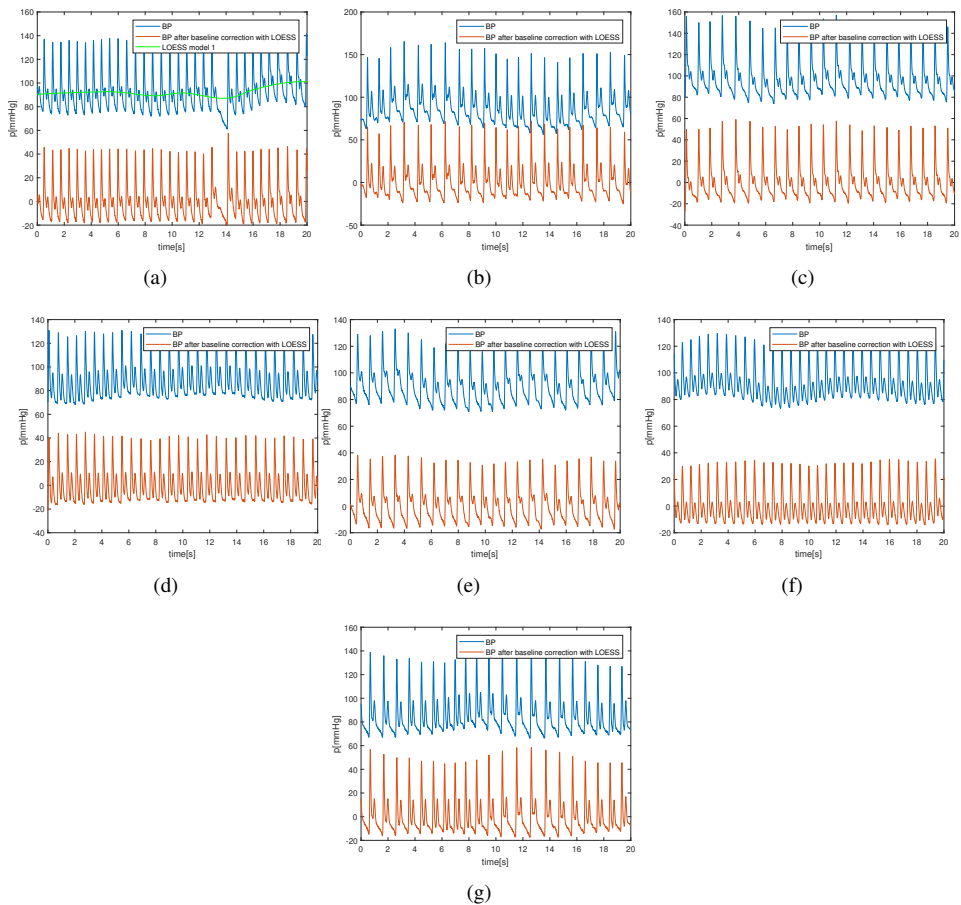


Figure 4.6: BP curve plotted before and after LOESS, where the blue (upper) curve are before LOESS and, red (bottom) curve is after LOESS and the green in (a) are the LOESS model 1: (a) subject 1; (b) subject 2; (c) subject 3; (d) subject 4; (e) subject 5; (f) subject 6; and, (g) subject 7.

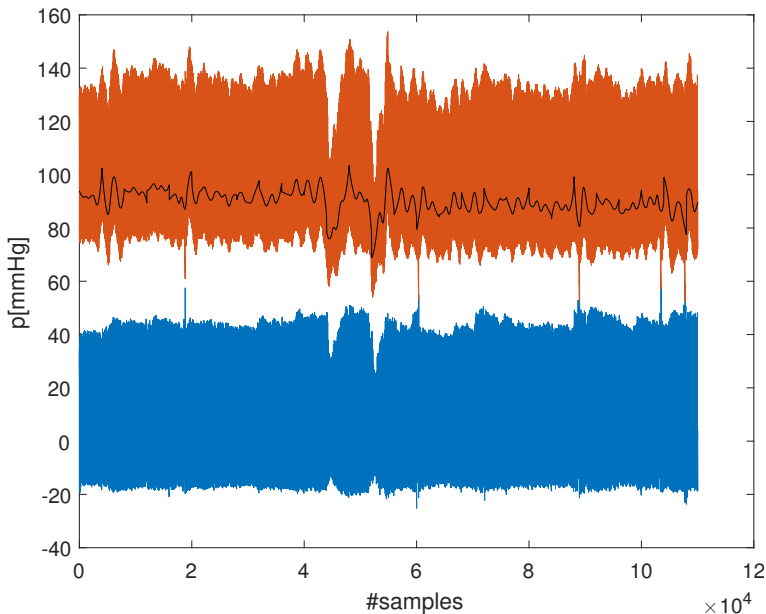
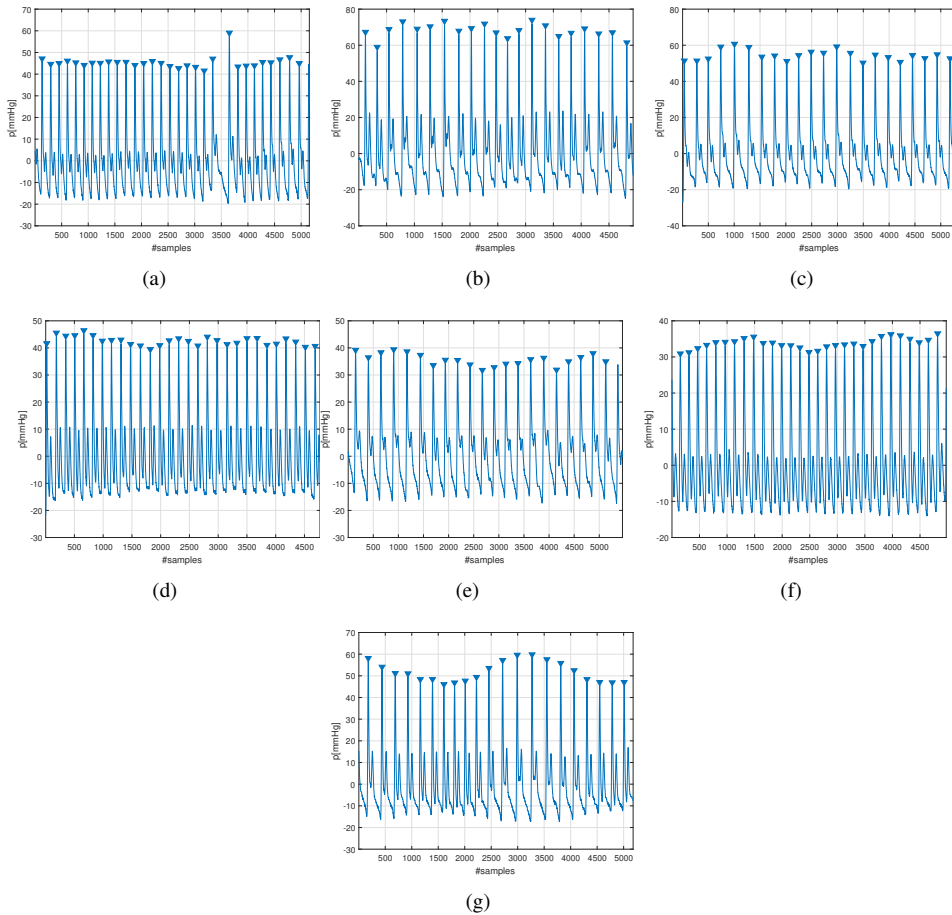


Figure 4.7: LOESS model for the whole 10 min dataset for subject 1, where the red curve (upper) is the original signal, the black curve is LOESS model 1 and the blue curve (bottom) is the signal after LOESS

time-consuming. An advantage with LOESS is the possibility to find different frequency components in the signal, not only the baseline. In figure 4.4 plot c) LOESS found the sine wave corresponds to the HR. The downside with LOESS is the possibility to find the wrong frequency components if the parameter tuning is bad, as the baseline and then the wrong baseline is subtracted from the BP curve. In this study, there was used some time to find a good value for the span, and this may be a disadvantage for LOESS. When the span value was found, there was no further tuning even for a new subject. However, in this study, there were only seven subjects, and they were sitting on a chair during the measurements, and there were few disturbances. This will may cause more problems when there are different subjects and more disturbances.

4.1.2 Find Systolic Peaks

Figure 4.8 shows a result of the function *findpeaks* for each of the subjects. The function never fails to find a systolic peak and never classify the peak dirotic notch peak as the systolic peak. In the project thesis, the function sometimes missed some peaks or misclassified a dirotic notch peak as a systolic peak, as described in the introduction. One of the reasons that it works better know is the removal of the baseline shift before finding the peaks. In this study only the peak minimum prominence was used as the input parameter, this is indifference from the project thesis [13], where both peak prominence and mini-



Figur 4.8: Find systolic peaks: (a) subject 1; (b) subject 2; (c) subject 3; (d) subject 4; (e) subject 5; (f) subject 6; and, (g) subject 7.

num peak distance was used as input parameters. In the project thesis [13] the minimum peak distance was used to use the knowledge about the HR, but this may not be the best idea if the HR suddenly changes. A result can be that some systolic peaks can be missed.

4.1.3 Interpolation and Heartbeat Separation

In the theory chapter, it is described that the HR varies between persons and that it also can change over time. However, to compare signals with different heart rate an interpolation was performed to compare different signals independent of the heart rate and compensate for shifts in the signals. Furthermore, are the heart beats separated into individual heart-

beats. Figure 4.9(a)-(g) shows the interpolation and the heartbeat separation for each individual subject, and figure 4.9(h) shows the interpolation for all the individuals. In section 2.1 there is explained about a shift in spectral measurements, and this is what is tried to fix here. In figure 4.9 it can be observed that the signals are stretched both before and after the middle peak and where the central peaks are still at the same point. Furthermore observe that the dicrotic notch peak is shifted to the left, which are the result we wanted such that we can perform further analysis independent of shifts. In addition to the individual subject in figure 4.9(a)-(g), we have figure 4.9(h) with all the subjects. The observation is that the dicrotic notch is almost fitted on top of each other. However, it is not a completely perfect fit as desired.

Figure 4.10 shows the interpolated, and separated heart beats for the whole 10 min signal for the seven subjects different subjects. We can observe in figure 4.10(a) that all the curves fit over the dicrotic notch and we can also observe that there are some few curves which are higher and are separated from the rest and these curves are the premature ventricular beats, explained and shown in the theory chapter. Furthermore in figure 4.10(b) it is more shift in the peak after the dicrotic notch and and last figure 4.10(c) it does not fit good. It can be observed that there are two different peaks. It was performed an investigation of the HR for each part of the simulation, to understand these differences in the separation. The result of this investigation was that when the pulse is a little bit higher than the constant 1.2 Hz HR as wanted, and the algorithm does not work that good.

Based on both the 20 seconds signals and the 10 minutes signals, we can observe that this algorithm does not perform optimal, when the HR changes over time, especially if the HR is near 1.2 Hz. Furthermore, we can observe that the subjects the algorithm have most problems with finding one constant frequency are the subjects shown in figure 4.9(c) and 4.9(e), which both separates into two different frequency components. A good observation is that both of these two subjects have respectively HR with mean frequency 1.08 Hz and 1.09 Hz, which means they have almost the same mean pulse. However, this is the mean pulse over the whole 10 minutes signal a further investigation of the result of each analysis of 20 seconds showed that the frequency is changing over time and are sometimes above the limit 1.2 Hz. The problem with the algorithm is that it finds the new sampling frequency based on if it is above or below the limit. However, when it just above the limit this will not change the sampling frequency much and then the signal is not stretched. On the other hand the signals with frequency above 1.4 Hz are shown in figure 4.10(a), 4.10(c) and 4.10(f), where all the peaks are almost on top of each other. However, this is not perfect, and there are some minor shifts in the peaks. At last, figure 4.10(b) and 4.10(g) shows two subject with HR below 1.08 Hz. Here we can observe that there are none of the heart beats, which separates clearly from the rest. However, there are more shifts in the dicrotic notch peak in the low HR than for the signals with HR above 1.4 Hz. The shift in the peaks shows that the algorithm performs well for subjects with a high HR, but not that good for subjects with lower HR.

4.1.4 Heartbeat separation

In figures 4.9(a)-(g) interpolation and heartbeat separation is shown for each individual subject, and in figure 4.9(h) all individuals are shown. We can observe from the different subjects that a difference in HR leads to a difference in how much of the heartbeat is included for each subject. Before interpolation subjects with subjects with a high HR (i.e., subject 1, 4 and 6) parts of the next and prior heartbeat are included in the separation. After the interpolation, the prior heartbeat is included, but not the next one. For subjects with a low HR (i.e., subject 2, 3, 5, and 7) neither the prior or the next heartbeat are included in the separation. The difference in how the separation works for different subjects creates a problem in the analysis, and this may be improved further to use the knowledge about the HR to find the separation instead of using a constant value. However, this creates a problem with comparing different subjects in the same analysis, but if the goal is to analyze individuals, this is not a problem.

4.1.5 MSC

Figure 4.9 shows the result of the separated heartbeats after the interpolation. The result of the MSC for each subject are shown in figure 4.11. A comparison of these two figures an observation is that the systolic peaks are closer to the MSC after the MSC than before. However, the dicrotic notch peak is not compressed that much. Furthermore, we can observe that after the MSC the few curves (due to PVC higher than the rest, are now together with the rest. This is because that the PVC are multiplicative since it is higher than the rest and additive since the baseline is at a low value. This shows that MSC can correct for both the additive and multiplicative effect in the systolic peak. However, MSC have not managed to compress much of the small peak except for the PVC. Furthermore, when use MSC this is a modeling tool and the information is not lost, and the correction of the multiplicative and the additive effect are kept in parameters and can be used for later analyzes.

4.2 PCA

Figure 4.12 shows the result of the PCA of subject one and the marked samples in figure 4.12(a) and (c) corresponds to the peaks that are higher because of the premature ventricular beats. An observation is that the marked samples in the influence plot are in the part where corresponds to high influence on the PCA model. In figure 4.12(b) the first PC shows that the first PC is where the value in the start of the signal. In figure 4.9(a) this corresponds to the value, which is almost a straight line and a low value. The second PC is mostly due to the value of the systolic peak. Furthermore, the explained variance plot in 4.12(d) needs seven PC to explain 95 % of the variance in the data.

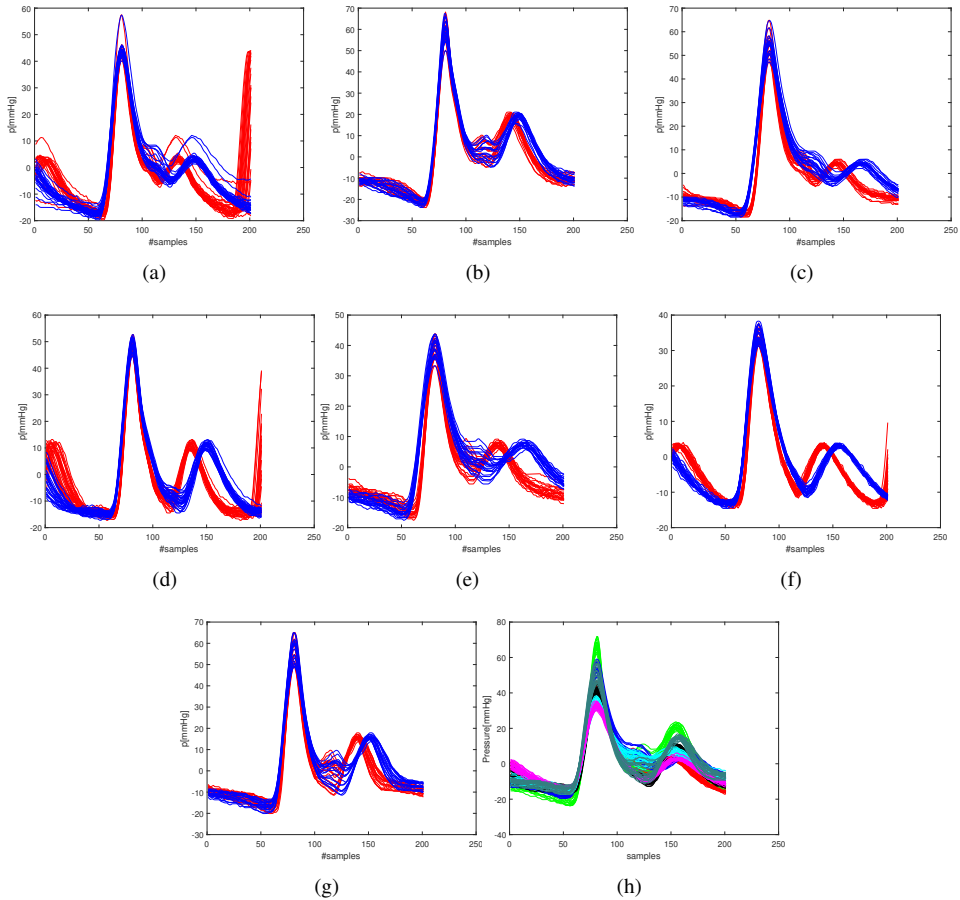
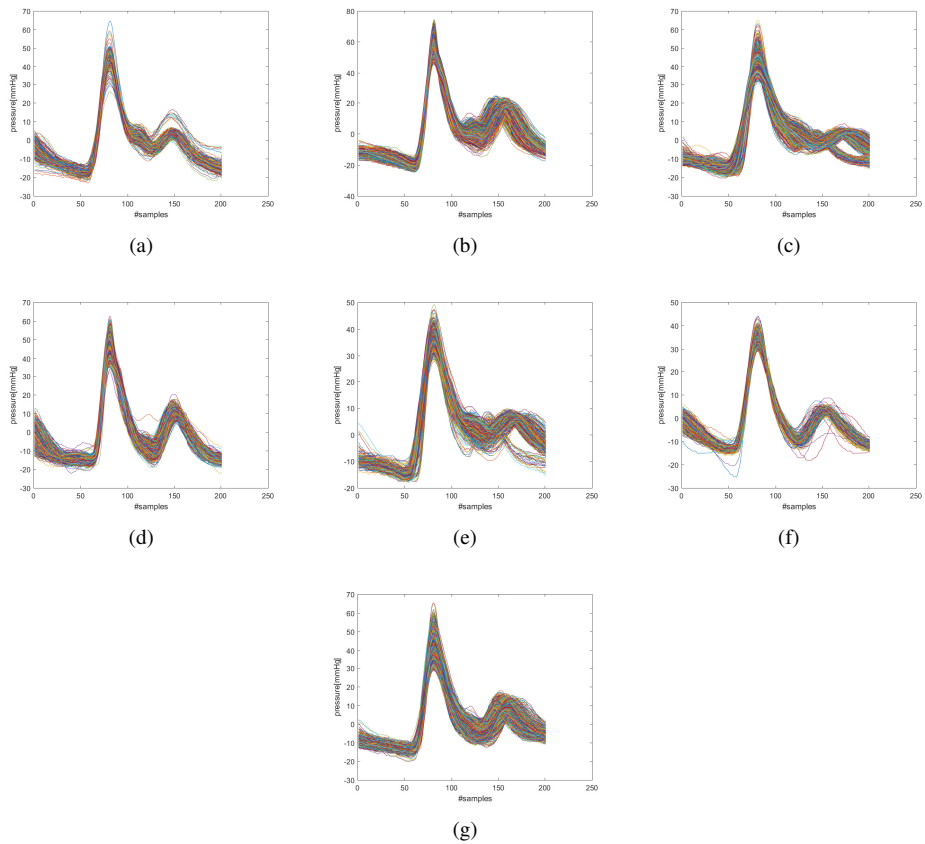
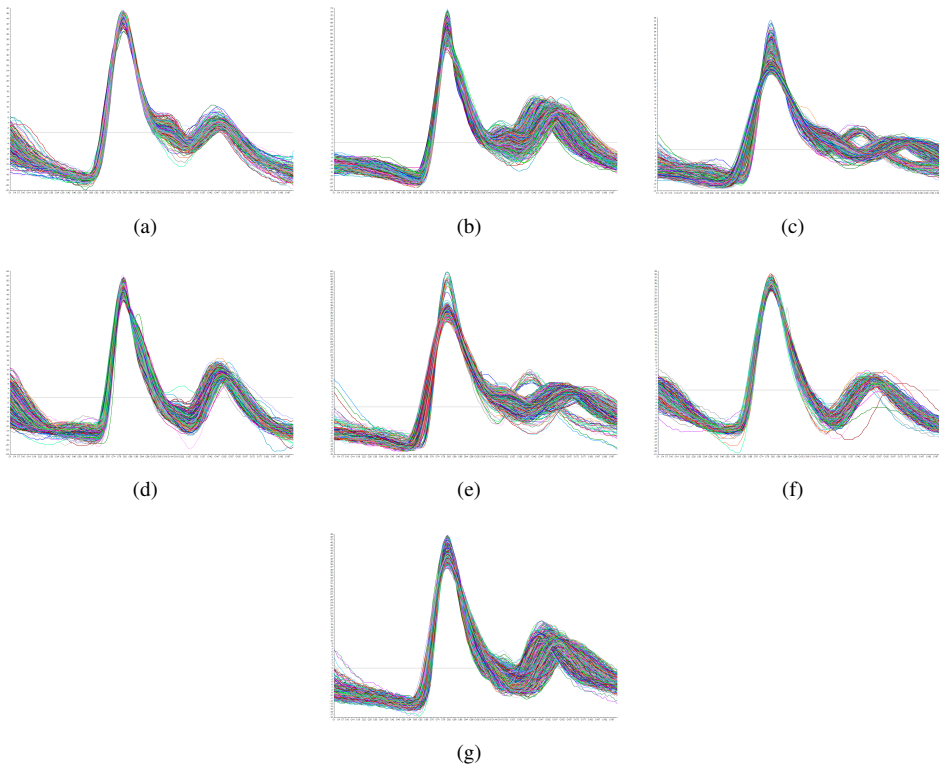


Figure 4.9: Separated heart beats of 20 seconds, where blue curves are after interpolation and red curves are before interpolation: (a) subject 1 with mean HR = 1.6 Hz; (b) subject 2 with mean HR = 1.04 Hz; (c) subject 3 with mean HR = 1.08 Hz; (d) subject 4 with mean HR = 1.4 Hz; (e) subject 5 with mean HR = 1.09 Hz; (f) subject 6 with mean HR = 1.49 Hz; (g) subject 7 with mean HR = 1.07 Hz; and, (h) shows all the different subjects after interpolation, where each color corresponds to different subjects.



Figur 4.10: Separated heart beats after interpolation for 10 minutes signal: (a) subject 1 with mean HR = 1.6 Hz; (b) subject 2 with mean HR = 1.04 Hz; (c) subject 3 with mean HR = 1.08 Hz; (d) subject 4 with mean HR = 1.4 Hz; (e) subject 5 with mean HR = 1.09 Hz; (f) subject 6 with mean HR = 1.49 Hz; and, (g) subject 7 with mean HR = 1.07 Hz.



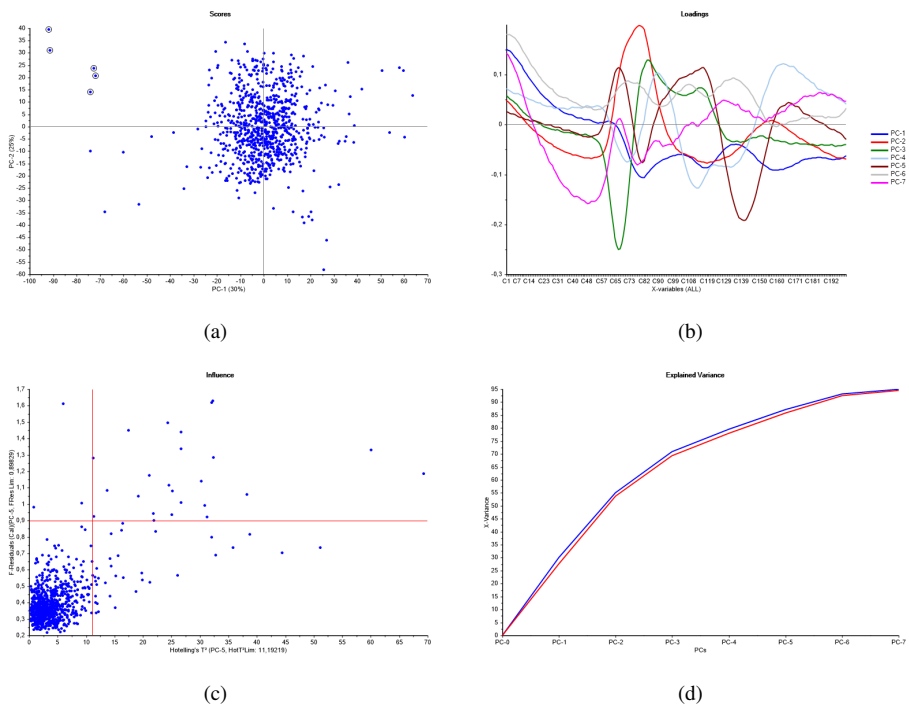
Figur 4.11: Separated heart beats after MSC: (a) subject 1 ; (b) subject 2 ; (c) subject 3 ; (d) subject 4 ; (e) subject 5 ; (f) subject 6 ; and, (g) subject 7.

Furthermore, figure 4.13 shows the PCA of subject one after MSC. As discussed in MSC section, the outliers caused by PVC are removed by the MSC. This is shown in figure 4.13(a), where there is a smaller variations in the APW curves than before the MSC in figure 4.12(a). Figure 4.13(b) shows the loading plot and the first PC shows that the most variance the value of the start of the heartbeat, which also was the case before the MSC. Based on this there may have been a better separation of the heart beats before the analysis.

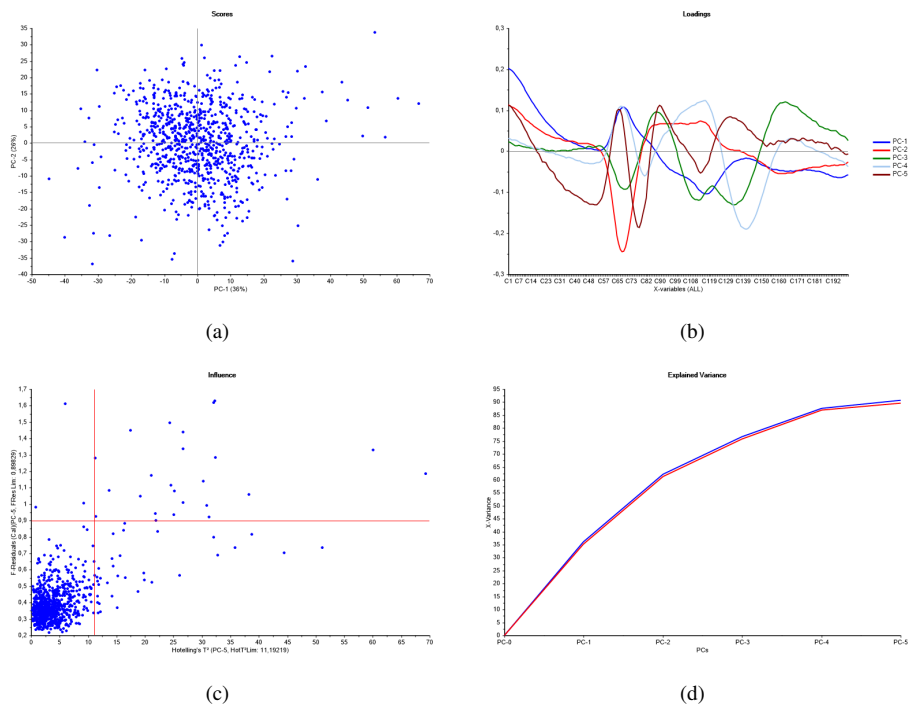
Figure 4.14 shows the result of PCA for subject two, and here we can observe that there is no clear outliers in figure 4.14(a). In figure 4.14(b) the first PC shows that most of the variance around the dirotic notch peak. A comparison with figure 4.9(b) shows that there is both a shift, the multiplicative and the additive effect in this area. PC two shows that most of the variances are at the start of the signal.

Figure 4.15 shows the result of PCA for subject two after MSC, and a comparison with before MSC there is almost the same result. The differences are small variations in the explained variance for each PC, and in both cases the PCA model can explain over 95% of the variance in the data with 7 PC. In the loading plot for the first 3 PC, there is almost no difference between them. This is confirmed by comparing figure 4.9(b) and 4.11(b), where there is just a small difference between them. In the loading plot for PC for PC 4 to 7, there are more variations between before and after the MSC. However, these components explain a much smaller part of the variations in the data.

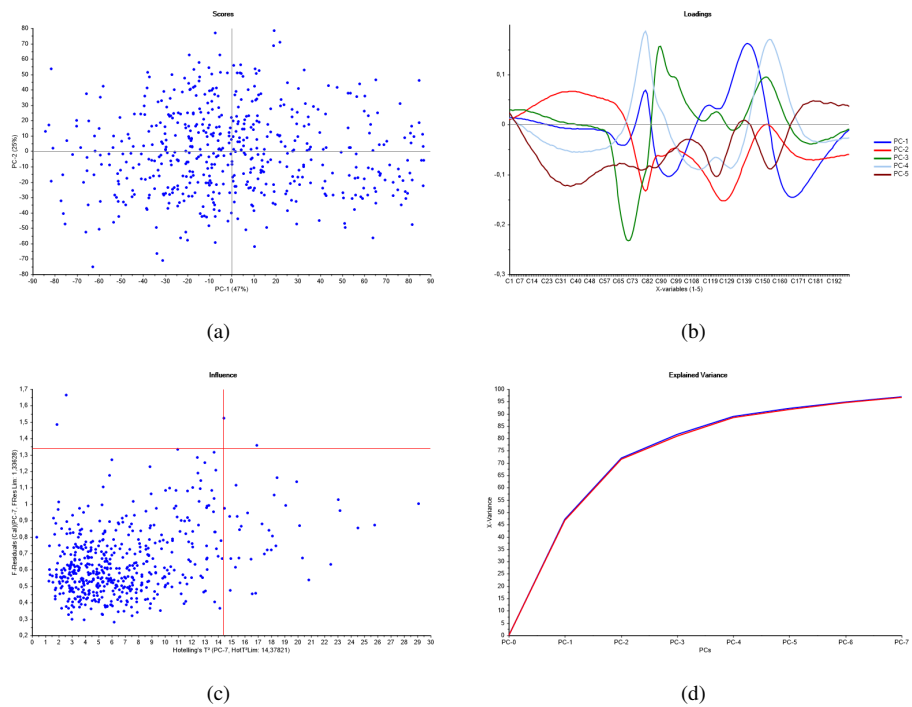
The PCA was as explained not the primary focus of this research, but used as an illustration of how the APW can be modeled after the preprocessing. We can observe that the model picks up differences in the APW.



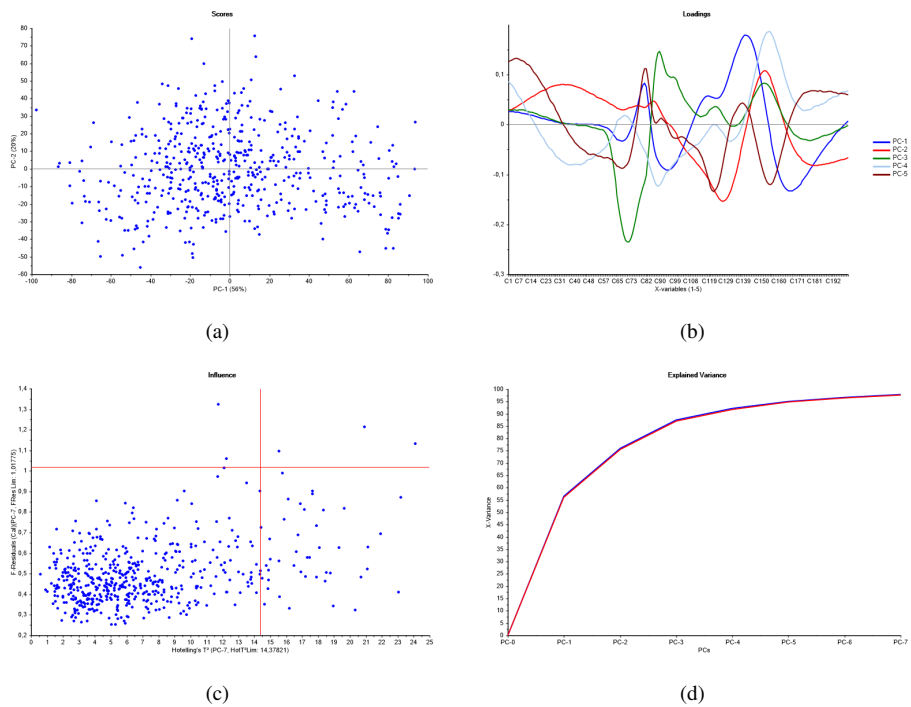
Figur 4.12: PCA of subject 1: (a) score plot; (b) loading plot; (c) influence plot ; (d) explained variance



Figur 4.13: PCA of subject 1 after MSC: (a) score plot; (b) loading plot; (c) influence plot ; (d) explained variance



Figur 4.14: PCA of subject 2: (a) score plot; (b) loading plot; (c) influence plot ; (d) explained variance



Figur 4.15: PCA of subject 2 after MSC: (a) score plot; (b) loading plot; (c) influence plot ; (d) explained variance

Conclusion and Further Work

5.1 Conclusion

The preprocessing with both LOESS and SG-filtering removes the baseline variations almost identically. However, the advantage with LOESS is the possibility to find several frequency components, such as both HR and the baseline variations. Also, LOESS is a modeling method and not a filtering method like SG-filtering. In this study, we want to model the heartbeats and not filtering the heartbeats. In every step of the way the data are modeled, and all the information in each of the step are preserved. Therefore, LOESS is the first and the best choice.

The interpolation algorithm works and stretches the signal such that the signals are possible to compare independent of the HR. However, it does not perform flawlessly, but good enough to be able to compare each independent of the HR. Furthermore, the separation algorithm works much better after the interpolation than before the interpolation, since it is then independent of the HR.

MSC removes the additive and multiplicative effect good in the systolic peak without changing the shape the form of the heartbeats, but not that good in the dicrotic notch peak. MSC have troubles with removing the additive and multiplicative effect in signals with peak shifting. However, MSC removes the outliers due to PVC, since this these heartbeats differs from the rest by higher BP during the whole heartbeats.

Finding the respiration rate based on the baseline variations are hard since the baseline variations are influenced by the respiration rate, but are also other factors. This makes it hard to find the respiration rate only based on the baseline variations.

After the preprocessing a further analysis of the signals can be performed, and as a part of that a small PCA was performed to check if individual difference appeared in the PCA. The PCA shows that there are variations between subjects and inside subjects. It also shows that outliers like PVC heartbeats appear in the PCA model, which shows an opportunity to work further with this problem.

5.2 Future Work

During this work, many issues have been discovered and there are many possibilities to work further with this study and the problem.

The interpolation algorithm should be further developed, such that it adapts and change when the HR changes independent of the HR is low or high.

In this study, Matlab is used as the tool for development, but for further development, this should be used in open source tools like Python or C/C++. Matlab is expensive and not an open source tool, which will make it hard to use this for commercial purposes. Furthermore, Unscrambler is not a tool for development it is more for graphical analysis looking for patterns.

After all the preprocessing the next step would be to model the heartbeats continuously and have a system which detects changes. One way of doing this could be to use the On-The-Fly (OTFP) tool as explained in [37]. The OTFP is a method for rational handling of time series of multi-channel measurements, which automatically detects any new co-variation patterns arising in the data and will remove much of the noise in the data. In this study, the focus has been APW. In further work it not only the APW should be included in this model, but also other signals like ECG, respiration, and temperature.

The study only involved seven subjects to test the algorithms at, so for further work, there should be new test persons There would be very interesting if the patients have some different disease or other known artifacts, this will give the chances to verify the results and further development.

There was performed an attempt to find the respiration rate based on baseline movement, but we had no actual measurements on the respiration to verify the results. With new data, there would be nice also to measure the respiration and then compare with the baseline movement. New data will give the opportunity to study what of the baseline movement is because of respiration and what is caused by other factors.

Other methods for finding the baseline in APW could also be tried, for example, discrete wavelet transform [18] or empirical mode decomposition [28]. Besides, other methods for finding systolic peaks could be tried, for example, [18] which uses empirical wavelet transform for locating systolic peak of the BP pulse.

Bibliografi

- [1] Human nibp set. [Online; accessed 30-April-2018].
- [2] A. Al-Jumaily and A. Lowe. Accuracy of the wave equation in predicting arterial pulse propagation. *Mathematical and Computer Modelling*, 57(3):460 – 468, 2013.
- [3] V. G. Almeida, J. Vieira, P. Santos, T. Pereira, H. C. Pereira, C. Correia, M. Pego, and J. Cardoso. Machine learning techniques for arterial pressure waveform analysis. *Journal of Personalized Medicine*, 3(2):82–101, 2013.
- [4] S. Axler, F. W. Gehring, and K. A. Ribet. *Graduate Texts in Mathematics Numerical Analysis*. Springer, 1 st edition, 1998.
- [5] Badaruddoza, R. Kumar, and M. Kaur. Principal component analysis of cardiovascular risk traits in three generations cohort among Indian Punjabi population. *Journal of advanced research*, 6(5):739–46, 9 2015.
- [6] J. Booth. A short history of blood pressure measurement. *Proceedings of the Royal Society of Medicine*, 70(11):793–799, 1977.
- [7] S. D. S. D. Brown, L. A. Sarabia, and J. Trygg. Standard Normal Variate, Multiplicative Singal Correction and extended Multiplicative Singal Correction Preprocessing in Biospectroscopy. In V. S. A.Kohler, M.Zimonja and H. Martens, editors, *Comprehensive chemometrics : chemical and biochemical data analysis*, chapter 2, pages 139–162. Elsevier, 2009.
- [8] C. S. B.Swarbrick. *Multivariate Data Analysis for Dummies*. England: Wiley, 1th edition, 2012.
- [9] A. Caduff, M. Zanon, P. Zakharov, M. Mueller, M. Talary, A. Krebs, W. A. Stahel, and M. Donath. First Experiences With a Wearable Multisensor in an Outpatient Glucose Monitoring Study, Part I: The Users' View. *Journal of Diabetes Science and Technology*, page 193229681775093, 1 2018.
- [10] Help in unscrambler. Under help, and contest in Unscrambler.

-
- [11] W. S. Cleveland. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979.
- [12] W. S. Cleveland and S. J. Devlin. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, 83(403):596–610, 1988.
- [13] T. Danielsen. Principal component analysis of the arterial pressure wave. project thesis, Norwegian University of Science and Technology, 2017. Not published.
- [14] Loess (aka lowess). webiste. [Online; accessed 29-mai-2018].
- [15] 10 factors that can affect blood pressure readings. [Online; accessed 30-April-2018].
- [16] P. Gago, M. F. Santos, A. Silva, P. Cortez, J. Neves, and L. Gomes. INTCare: a Knowledge Discovery Based Intelligent Decision Support System for Intensive Care Medicine. *Journal of Decision Systems*, 14(3):241–259, 1 2005.
- [17] W. Gander and U. von Matt. *Smoothing Filters*, pages 121–139. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.
- [18] A. Ghaffari, M. R. Homaeinezhad, M. Akraminia, and M. Davaeiha. Finding events of electrocardiogram and arterial blood pressure signals via discrete wavelet transform with modified scales. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(1):27–42, 2010. PMID: 20225455.
- [19] A. Guyton and J. Hall. *Textbook of Medical Physiology*. W.B. Saunders Company, 9nd edition, 1996.
- [20] B. Hametner, M. Schneider, S. Parragh, and S. Wassertheurer. Computational assessment of model-based wave separation using a database of virtual subjects. *Journal of Biomechanics*, 64(Supplement C):26 – 31, 2017.
- [21] H. Martens and M. Martens. *Multivariate Analysis of Quality An Introduction*. England: Wiley, 1th edition, 2001.
- [22] K. S. Imhoff M. Alarm algorithms in critical care monitoring. *Anesthesia & Analgesia*, 102(5):1525–37, 2006.
- [23] G. Kaur, A. S. Arora, and V. K. Jain. Using hybrid models to predict blood pressure reactivity to unsupported back based on anthropometric characteristics. *Frontiers of Information Technology & Electronic Engineering*, 16(6):474–485, 6 2015.
- [24] E. M. Koski, T. Sukuvaara, A. Mäkivirta, and A. Kari. A knowledge-based alarm system for monitoring cardiac operated patients—assessment of clinical performance. *International journal of clinical monitoring and computing*, 11(2):79–83, 5 1994.
- [25] D. L. Kulick and W. C. : Shiel. Premature ventricular contractions (pvc, pvc). [Online; accessed 30-April-2018].

-
- [26] J. M. Lilly and S. C. Olhede. Higher-Order Properties of Analytic Wavelets. *IEEE Transactions on Signal Processing*, 2 2008.
- [27] J. M. Lilly and S. C. Olhede. On the Analytic Wavelet Transform. *IEEE Transactions on Information Theory*, 56(8):4135–4156, 8 2010.
- [28] M.-T. Lo, K. Hu, Y. Liu, C.-K. Peng, and V. Novak. Multimodal Pressure Flow Analysis: Application of Hilbert Huang Transform in Cerebral Blood Flow Regulation. *EURASIP journal on advances in signal processing*, 2008:785243, 2008.
- [29] M. Maleki, A. Mouazen, H. Ramon, and J. D. Baerdemaeker. Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. *Biosystems Engineering*, 96(3):427 – 433, 2007.
- [30] Memscap — invasive blood pressure. website. [Online; accessed 23-mai-2018].
- [31] I. Moxham. Understanding Arterial Pressure Waveforms. *Southern African Journal of Anaesthesia and Analgesia*, 9(1):40–42, 2 2003.
- [32] S. J. Orfanidis. *INTRODUCTION TO Signal Processing*. Pearson Education, Incc, 1996.
- [33] G. Parati, R. Casadei, A. Groppelli, M. Di Rienzo, and G. Mancia. Comparison of finger and intra-arterial blood pressure monitoring at rest and during laboratory testing. *Hypertension (Dallas, Tex. : 1979)*, 13(6 Pt 1):647–55, 6 1989.
- [34] J. PENAZ. Photoelectric measurement of blood pressure, volume and flow in the finger. *Digest of the 10th international conference on medical and biological engineering-Dresden, 1973*, 104, 1973.
- [35] R. G. Rice and D. D. Do. *Applied mathematics and modeling for chemical engineers*. John Wiley & Sons, Inc, 2nd edition, 2014.
- [36] F. Schmid, M. S. Goepfert, and D. A. Reuter. Patient monitoring alarms in the icu and in the operating room. *Critical Care*, 17(2), 2013.
- [37] R. Vitale, A. Zhyrova, J. F. Fortuna, O. E. de Noord, A. Ferrer, and H. Martens. On-the-fly processing of continuous high-dimensional data streams. *Chemometrics and Intelligent Laboratory Systems*, 161:118 – 129, 2017.
- [38] K. Wesselin, B. De Wit, J. Settels, and K. W.H. On the indirect registration of finger blood pressure after pefiaz. *Funkt Biol Med*, 1:245–250, 1982.
- [39] N. Westerhof, P. Segers, and B. E. Westerhof. Wave separation, wave intensity, the reservoir-wave concept, and the instantaneous wave-free rationovelty and significance. *Hypertension*, 66(1):93–98, 2015.
- [40] A. Yartsev. Normal arterial line waveforms. [Online; accessed 30-April-2018].
-

Appendix

A.1 Matlab Code

A.1.1 Blood Pressure Separation

```
1 function [s] = BloodPressureSeparation(data,meanpeakprom)
2
3 sepDist = 120; %the separation distance after the systolic peak
4 sepDist2 = 80; % separation distance before the systolic peak
5 systolic = locate_syst(data,meanpeakprom);
6 s = zeros(length(systolic),sepDist+sepDist2+1);
7 for k = 1:length(systolic)-1
8     if (systolic(k)-sepDist2 ≥ 0)&&(systolic(k)+sepDist ≤ ...
9         length(data)) && ((systolic(k)≥sepDist) && ...
10            (systolic(k)≥sepDist2))
11         s(k,:) = data(systolic(k)-sepDist2:systolic(k)+sepDist);
12     end
13 end
14 end
```

A.1.2 Locate Systolic Peak

```
1 function [syst_loc] = locate_syst(data,meanpeakprom)
2 [-,syst_loc] = findpeaks(data,'MinPeakProminence',meanpeakprom);
3 end
```

A.1.3 LOESS

```
1 loess_model_1 =smooth(BloodPressure,0.5,'loess');
2 smooth_data_1 = BloodPressure -loess_model_1';
3
4 loess_model_2 =smooth(smooth_data_1,0.2,'loess');
5 smooth_data_2 = smooth_data_1 -loess_model_2';
6
7 loess_model_3 =smooth(smooth_data_2,0.05,'loess');
8 smooth_data_3 = smooth_data_2 -loess_model_3';
9 BP_After_LOESS = smooth_data_2;
```

A.1.4 SG-filtering

```
1 baseliney = sgolayfilt(BloodPressure,3,501);
2 detrendedY = BloodPressure - baseliney;
```

A.1.5 Interpolation

```
1 meanpeakprom=40;
2 [pk_systolic,location_systolic] = ...
    findpeaks(BP_After_LOESS,'MinPeakProminence',meanpeakprom);
3 pk_syst_interval = zeros(1,length(location_systolic)-1);
4 for m=2: length(location_systolic) % from m = 2 since the first is zero
5     pk_syst_interval(m) = location_systolic(m)-location_systolic(m-1);
6 end
7 meanCycle_syst = mean(pk_syst_interval(2:end));
8 pulse = Fs/meanCycle_syst;
9 if pulse <1.2
10     pulse_want = 0.9+(1-pulse);
11 else
12     b=1;
13     pulse_want = 1.2;
14 end
15 Fsq = pulse/pulse_want*Fs;
16 Tq = 1/Fsq; % Sampling period for interpolation
17 Lq = Fsq *L/Fs;
18 tq = (0:Lq-1)*Tq;
19 interpolate_BP = interp1(t,BP_After_LOESS,tq,'line');
```