Universitetet i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study program/specialization: Informasjonsteknologi - Automatisering og signalbehandling	Spring semester, 2017 Open / ~~Confidential~~
Author: Fredrik Mellemstrand	……Fredrik Mellemstrand…… (signature author)

Instructor: Trygve Eftestøl

Supervisor(s): Trygve Eftestøl, Helge Fossan

Title of Master's Thesis: Simulation of 12 Lead ECG
Norwegian title: Simulering av 12 avlednings EKG

ECTS: 30

Subject headings: ECG Medical simulation	Pages: 84 + attachments/other: code in .7z format Stavanger, 15.06.17 Date/year

Summary

Simulation is commonly used in the training of medical professionals; including nurses, emergency responders and doctors. By using real equipment with medical simulators it allows for a much more realistic experience, allowing the students to get experience with the actual equipment to be used later with real patients.

The idea behind this project is to develop methods to allow the simulation of 12 Lead ECG based on real recordings, such that users can connect devices such as patient monitors, defibrillators and electrocardiographs to a simulator and record realistic ECG.

In the report it is presented how a 12 Lead ECG with synchronized leads can be calculated into the voltage potentials, which in turn can be generated to the ECG electrodes. It is then showed how the QRS complexes and waves of the ECG is detected. This is used to segment the ECG recording into beats, and a representative beat is found as the median beat of all the segmented beats fulfilling different criteria. The representative beat of the ECG is then modeled by a simple parametric model and it is showed how this can be used for realistic simulation of normal sinus rhythms at different heart rates, as well as presenting signs of conditions like myocardial infarction.

In the project the simulations were based on the PTB Diagnostic ECG database, a database consisting of 549 ECG recordings from 290 different healthy volunteers or subjects with different heart diseases. In the report it is showed how the voltage potentials required for simulation could be calculated from the records in the database. The ECG was then reconstructed from the voltage potentials with low errors on most records in the database. The parametric model was fit to all the records of the database, and for many records the model seems to represent the ECG quite well. Finally it is showed how rate adjustment and myocardial infarction simulation would look on some of the records in the database.

Acknowledgements

I would like to thank my advisor for the project, Trygve Eftestøl, for excellent feedback and guidance throughout the whole project.

I would also like to thank Helge Fossan for suggesting the topic for this thesis, and for all your help along the way.

Finally I would also like to thank Leik Woie for your valuable feedback on the project.

Contents

List of Figures

List of Tables

Abbreviations

ECG	**e**lectro**c**ardio**g**ram
MGF	**m**odified **g**aussian **f**unction
SA node	**s**ino**a**trial node
AV node	**a**trio**v**entricular node
RA	**R**ight **A**rm
LA	**L**eft **A**rm
LL	**L**eft **L**eg
RL	**R**ight **L**eg
QTc	corrected QT interval
BPM	**B**eats **P**er **M**inute
ms	**m**illi**s**econds

1. Introduction

Simulation is commonly used in the training of medical professionals; including nurses, emergency responders and doctors. Increasing numbers of medical schools and health care institutions are now implementing simulation-based learning into their education. A simulation allows students to get valuable experience and confidence in their abilities without causing unnecessary risk or distress to patients.[2] While medical simulation can be done with relatively simple means, advanced patient simulators allows for a much more immersive experience. When students use real medical equipment with a patient simulator this adds to the experience, and provides useful experience with the specific equipment to be used later with real patients.

1.1 Problem description

The goal of this project will be to develop methods to allow simulation of 12 Lead electrocardiograms based on real recordings. This will allow users to connect patient monitors, defibrillators, electrocardiographs or other devices capable of recording an electrocardiogram to a simulator. The first goal is to show how the voltage potentials to be generated can be reconstructed from a recorded electrocardiogram with synchronized leads. A method is then developed to extract a representative beat from the ECG recording. This beat is fit to a simple parametric model. It is then demonstrated how this model can be used for a realistic simulation of normal sinus rhythms at different heart rates (also including tachycardia and bradycardia), as well as to present signs of myocardial ischemia and infarction such as ST-elevation, pathologic Q waves and inverted T waves.

1.2 Previous work

The project is based on the work of Helge Fossan, presented in the document called ECG Simulation[3]. In this document it is explained how ECG can be simulated using

real ECG data. This document presents the ECG relations and how these can be solved for each of the electrode voltages. It also suggests how to design the required electronics to play back the ECG.

In the initial phase of the project, several related works was found. ECGSYN[4] generates a realistic looking ECG with adjustable heart rate, number of beats, sampling frequency, waveform morphology, mean and standard deviation of the RR interval, frequency properties of the RR interval. ECGSYN uses a set of three ordinary differential equations to produce realistic ECG waveforms. It does not however provide the flexibility necessary to simulate a 12 lead ECG, but it is useful for evaluation of algorithms for QRS detection or wave delineation. The idea of using Gaussians to represent the ECG waveform originates from this work.

Several products also exists on the market for generating ECG waveforms. Vital signs patient simulators such as Fluke ProSim 8[5] are designed to test patient simulators. These are standalone units, providing many of the features wanted for this project. However, these are proprietary large standalone units with button inputs, which would be difficult to integrate in a system such as a high fidelity patient simulator. The HAL®S1020 Emergency Care Simulator[6] by Gaumard is a patient simulator implementing many of the features discussed in this project with a large ECG library and ability to add myocardial infarction symptoms to the ECG. It also allows for your own defibrillator to be connected to the simulator. This is a proprietary product, and additional ECG's can be added by manual editing in one of their tools, which is different to the automatic data based approach suggested in this project.

1.3 Structure of the report

The report is divided into five main parts. The first part will introduce the necessary background theory the solution is based on. This includes both theory about electrocardiograms and signal processing techniques. The next part describes the databases used in this project. Then the proposed method is presented. This is followed by a chapter that shows the results of the simulation. The results are then discussed in a final chapter where suggestions of future work and problems with the method are also presented.

In addition to the previously mentioned chapters there are two appendices. The first show some experiments performed on the database. The second appendix explains how to install the software used in this project to process the records, find a representative beat and simulate new ECG. It also contains a brief explanation of what the different files in the project do.

2. Background

This section will walk through the necessary theory for understanding an electrocardiogram and and some basics of its interpretation. First the physiology of the heart will be presented. Then it will be explained how this gives rise to the tracings of the electrocardiogram, and finally how this relates to simulation of an electrocardiogram. Then the necessary background information about the QRS detection algorithm, ECG wave delineation and signal processing are presented.

2.1 Physiology of the heart

The human heart is a muscular organ responsible for pumping blood throughout the cardiovascular system. The heart is divided in four different chambers. These are called

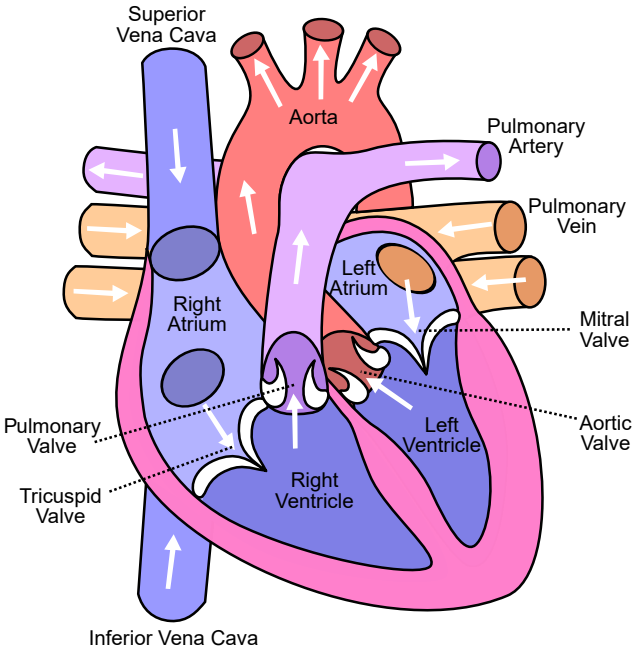

FIGURE 2.1: Diagram of the human heart. (`https://commons.wikimedia.org/wiki/File:Diagram_of_the_human_heart_(cropped).svg`)

the right and left atrium, and the right and left ventricle (figure 2.1). The atria are the upper chambers of the heart. This is where the blood enters the heart, and it is also where the heart beat normally is initiated. The ventricles are the lower chambers where blood gets pushed out[7]. The right atrium and ventricle is called the right heart, and the left atrium and ventricle is called the left heart. These function as two separate pumps. The right heart pumps blood through the lungs, while the left heart pumps blood through the peripheral organs[8]. The heart contracts rhythmically, and different parts of the heart has to contract at different times to pump efficiently. One such contraction and relaxation, is called a cardiac cycle.[9] The timing of the different parts of the heart is controlled by a elaborate system of cardiac muscle called the electrical conduction system of the heart.

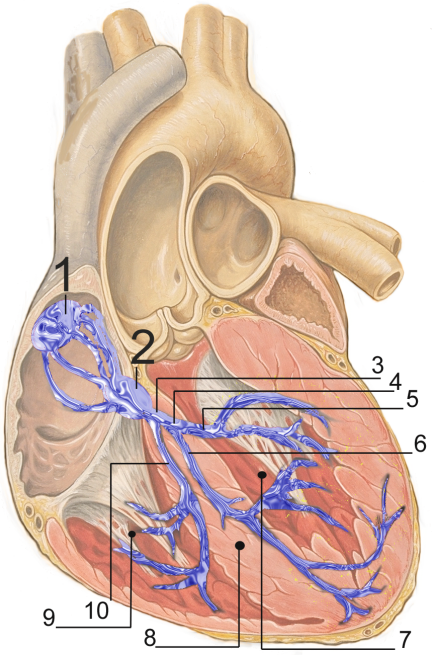

FIGURE 2.2: Diagram of the conduction system of the heart. 1: Sinoatrial node, 2: Atrioventricular node, 3: Bundle of His, 4: Left bundle branch, 5: Left posterior fascicle, 6: Left-anterior fascicle 7: Left ventricle, 8: Ventricular septum, 9: Right ventricle, 10: Right bundle branch. (By J. Heuser - self made, based upon Image:Heart anterior view coronal section.jpg by Patrick J. Lynch (Patrick J. Lynch; illustrator; C. Carl Jaffe; MD; cardiologist Yale University Center for Advanced Instructional Media), CC BY 2.5, `https://commons.wikimedia.org/w/index.php?curid=1734607`)

The electrical activity of the heart is generated by cardiac action potentials. When ions move through the cell membrane of a cell they give rise to action potentials. This will cause brief changes in the voltage across the cell membrane. Cardiac muscle cells are different from skeletal muscle cells because they can generate action potentials without being stimulated by nervous activity. This means that the cardiac cells have automaticity.[10] The cardiac cycle is normally initiated by a group of specialized cells in the atria called the sinoatrial node (SA node). During normal heart rhythms this will

be the pacemaker of the heart, and the SA node is therefore often called the primary pacemaker of the heart. The cells in the sinoatrial node will spontaneously depolarize at a rate of between 60 and 100 times per second, which will result in a heart rate of 60 to 100 beats per minute. This rate is constantly modified by the autonomic nervous system.

If the SA node is unable to initiate the cardiac cycle, the heart beat can also be initiated by the cells in the AV junction (atrioventricular junction). These cells will normally discharge at a rate of about 40-60 beats per minute. Because of this the AV junction is commonly called the secondary pacemaker. Further down it can also be initiated by the left and right branches of the His bundle, and the Purkinje fibres. These cells will produce a spontaneous action potential at a rate of about 30-40 beats per minute and can therefore function as the pacemaker[11].

The cells of the myocardium is connected to each other with electrically conductive structures. When one cell is stimulated it will quickly spread to the neighbooring cell, then to its neighbooring cell and so on. If the SA node function as the heart's pacemaker, it initiates the heart beat, causing the atria to depolarize. This fills the ventricles with blood. The atria are electrically separated from the ventricles. Therefore the cells in the AV node will delay the signal slightly while conducting it to the ventricles. This results in the atria contracting a short time ahead of the ventricles. This delay is essential for the hearts effectiveness[8, ch9]. Next the signal travel along the Bundle of His and to the bundle branches causing the venctricles to contract and pump blood into the cardiovascular system. The atria can be seen as a primer pump for the heart, increasing the volume which the ventricles can pump for each contraction[12].

2.2 The electrocardiogram

An electrocardiogram (ECG or EKG) is a commonly performed measurement of the electrophysiological activity of the heart. The ECG is useful in the detection of conditions such as rhythm disturbances, heart block, ischemia and infarction and other signs which might have diagnostic importance. An ECG is obtained by measuring the voltage potential between electrodes attached to the skin. Much less commonly the ECG can also include electrodes inside the body such as in the esophagus, which provides more information about the atria. For most ECGs between 3 and 10 electrodes attached to the surface of the skin are used. This includes 4 electrodes attached to the limbs, called limb leads. One of these, RL, is used to remove common mode noise, instead of being

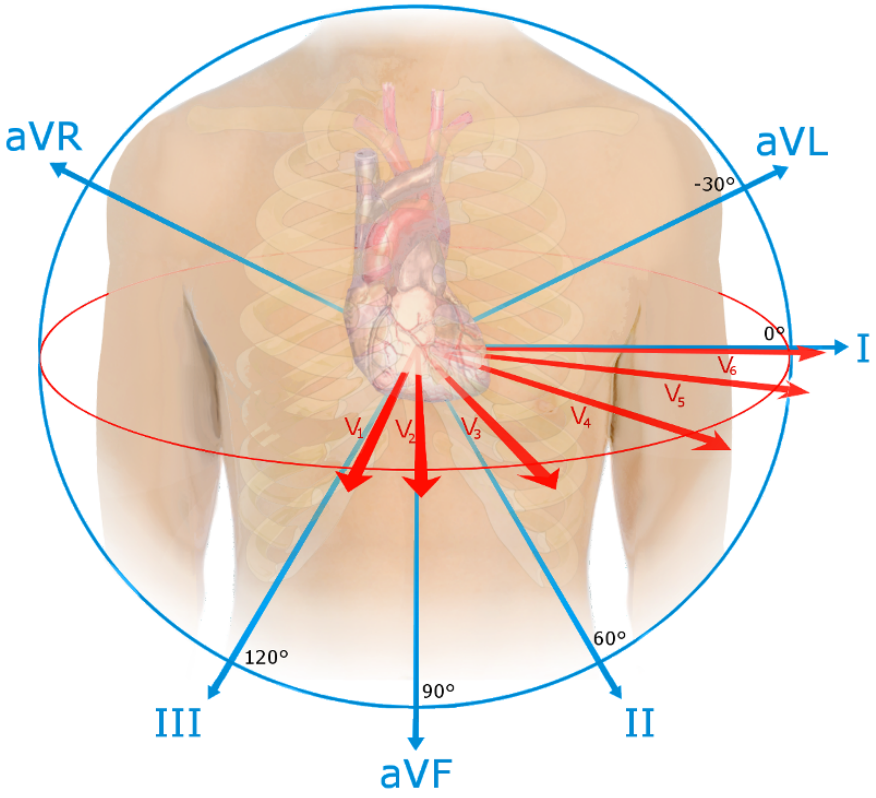

FIGURE 2.3: Spatial orientation of ECG leads. (By Npatchett - Own work, CC BY-SA 4.0, `https://commons.wikimedia.org/wiki/File:EKG_leads.png`)

used for measurements. The remaining 6 are called precordial electrodes and are attached across the chest. The limb leads are named: RA (Right arm), LL (Left leg), LA (Left arm), RL (Right Leg). The precordial leads are named V1-6[13].

The electrocardiogram is normally represented by tracings printed on a paper or showed on a display. The most commonly used diagnostic ECG contains 12 traces derived from measurements from 10 electrodes. These traces are denoted leads. Each of these leads is said to provide a different perspective to the electrophysiological activity of the heart described in chapter 2.1. The angles of the different leads can be seen in figure 2.3. To a trained clinician this provides a lot of information.

Each cardiac cycle gives rise to a characteristic pattern in the ECG (figure 2.4). These represent the different areas of the heart depolarizing and repolarizing. When a wave of depolarization moves towards the positive electrode and away from the negative it will result in a positive deflection of that lead. The opposite is true for the repolarization. The depolarizing of the sinoatrial node is not directly visible on the ECG. The waves of the ECG are normally denoted P, Q, R, S, T and U waves.

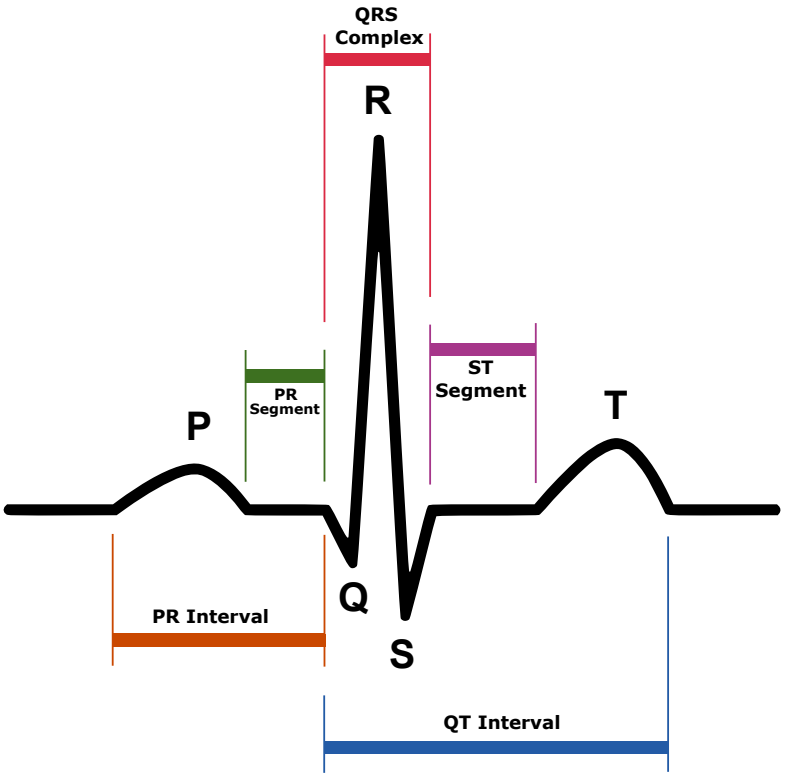

FIGURE 2.4: Schematic of a normal sinus rhythm ecg. (`https://commons.wikimedia.org/wiki/File:SinusRhythmLabels.svg`)

P Wave The P Wave is the first wave visible on the ECG. It represents the depolarization of both atria. This wave of depolarization is initiated in the sinoatrial node, and then moves from the right atrium to the left. The P wave is normally visible as a small upright deflection of the lead, but it can also be negative or biphasic (both positive and negative). The early part of the P-wave represents the depolarization of the right atrium and the later part represents the depolarization of the left atrium. Therefore in lead V1 the P-wave is commonly biphasic as the wave of depolarization moves in different directions in the right and left atria in relation to lead V1.[14] The P wave is typically less than 80 ms in duration.[15]

QRS complex The QRS complex is a name for the 3 common deflections of the ECG corresponding to the depolarization of the ventricles. The R wave is defined as the first positive deflection after the P wave. The Q wave is any negative deflection preceding the R wave. The S wave is a negative deflection succeeding the R wave.[16][17] The QRS complex is normally 80 to 100ms in duration[15].

T wave The T wave is normally a positive deflection after the QRS complex. It originates from the repolarization of the ventricles.[18]

Method	**Formula**
Bazett	$QT_c = QT/\sqrt{RR/1000}$
Frederica	$QT_c = QT/\sqrt[3]{RR/1000}$
Framingham	$QT_c = QT + 0.154(1 - RR/1000)$
Hodges	$QT_c = QT + 1.75(hr - 60)$

TABLE 2.1: Formulas for corrected QT intervals. From [1]

U wave The U wave is a small positive deflection after the T wave. The exact origins of this wave is unknown, but it might originate from the repolarization of the interventricular septum (which is the wall separating the left and right ventricles)

A segment in the ECG means the segment connecting two waves of the ECG without including any of them. An interval is a segment of the ECG including one or both of the waves. Some of the most commonly mentioned intervals are the RR interval, PR interval, ST segment, and the QT interval.[15]

RR Interval The distance between two consecutive R waves. This is dependant on the heart rate. If the RR interval is measured in milliseconds the heart rate in beats per minute can be found as $hr = (60 * 1000)/RR$. Thus based on the normal range of heart rates of adults between 60 and 100 bpm, we have a normal range of RR-intervals between 600 and 1000 ms.

PR interval The PR interval is the interval between the beginning of the P wave and the beginning of the R wave. This interval is normally between 120 to 200ms, and can be interpreted as the time it takes for the wave of depolarization takes to go from the sinus node and through the AV node.[15]

ST segment The segment connecting the end of the QRS complex with the beginning of the T wave. This is usually isoelectric, but a elevation or depression of this segment can be a important sign of different conditions.

QT interval The time from the start of the Q wave to the end of the T-wave. This interval varies with heart rate. Therefore a corrected QT interval called QTc are normally calculated. This is meant to estimate the QT interval at a heart rate of 60 BPM. QTc is normally below 440ms. There are several formulas for calculating the QTc. A selection of QTc interval formulas can be found in table 2.1.

2.3 ECG interpretation basics

An ECG often has to be interpreted quickly. A suggested systematic way of interpreting an ECG is suggested by Gerard Fennessy as the ECG "Rule of Fours"[19]. This consist of the four initial features to look for, the four waves or complexes and the four intervals. The four initial features are the following:

1. **History and clinical picture** - The ECG is a test, and should always first be considered in the context of the history and other clinical symptoms
2. **Rate** - For adults if rate is below 60 it is called bradycardia, if the rate is over 100 it is called tachycardia
3. **Rhythm** - Is it a sinus rhythm or other rhythm? Is the rhythm regular?
4. **Axis** - Calculate the axis of the ECG. This can be done on by taking the inverse tangent of the net deflection in lead aVF divided by the net deflection in lead I. The net deflection is found by taking the positive deflection minus the negative deflection of the QRS complex.

The four waves are:

1. **P-wave** - Examine morphology
2. **QRS-complexes** - Look for Q-waves, QRS amplitude and R wave progression in the precordial leads
3. **T-waves** - Look for T-wave inversion, T-wave flattening, concordance or discordance with QRS
4. **U-waves** - Are they present?

The four intervals:

1. **PR interval** - Find length between the onset of the P wave to the onset of the R wave.
2. **QRS width** - Find length between the onset of the Q wave to the end of the R wave.
3. **ST-segment** - Look for elevation or depression. Is the ST segment upsloping or downsloping?
4. **QT interval** - Find length between the onset of the Q wave and the end of the T wave.

2.4 Common artifacts in the ECG

An artifact in the ECG is parts of the ECG which does not originate from the electrical activity of the heart. This often originates from other muscles, such as with patients that have muscle tremors from for example Parkinson's disease or hypothermia. These artifacts, called muscle artifacts, can be reduced by low pass filtering.

Another example of a common ECG artifact is baseline wander. This means that the baseline of the ECG shifts over time. This may happen with respiration but are also commonly associated with loose or dry electrodes[20]. This can be reduced by high pass filtering.

2.5 ECG relations for simulation

Each of the leads in a standard 12 lead ECG is derived from measuring voltage differences between different electrodes, or combinations of electrodes on the surface on the skin. In clinical use the actual voltage potentials in each node is not of interest, and the analysis is only carried out on the leads. In the context of simulation, reconstructing these node voltages from the 12 leads might be of interest, and could enable a simulator to play

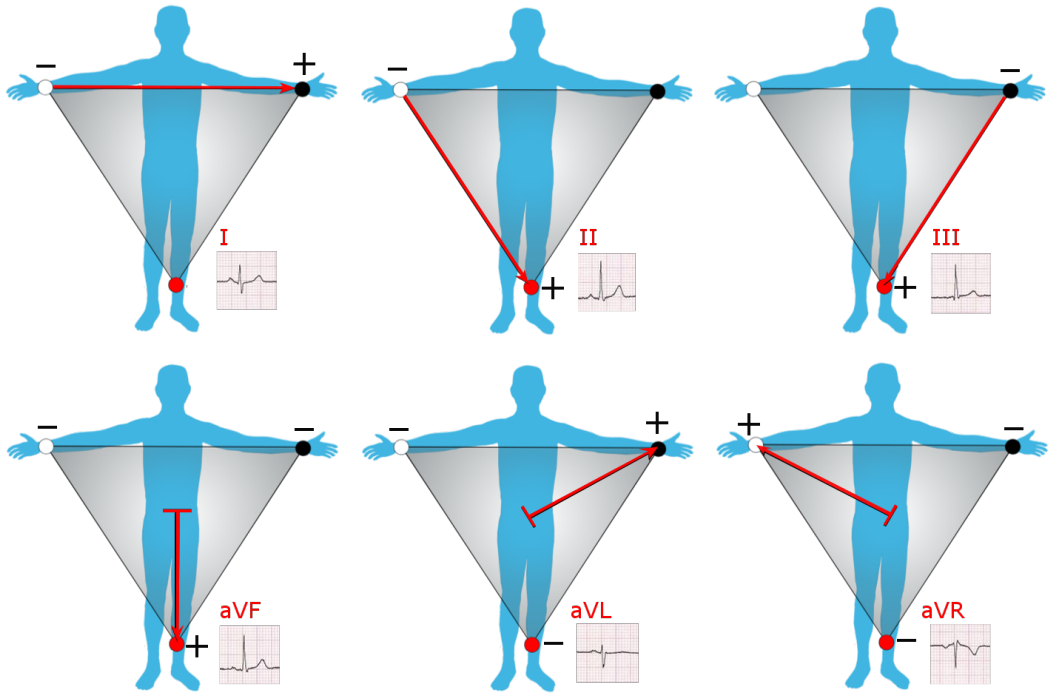

FIGURE 2.5: Visualisation of the different leads. (By Npatchett - Own work, CC BY-SA 4.0, `https://commons.wikimedia.org/w/index.php?curid=39235282`)

back a recorded ECG to a electrocardiograph or patient monitor. The following section will describe how the leads are derived from the electrodes.

The standard 12 leads are commonly divided into the limb leads, augmented limb leads and the precordial leads. The limb leads are voltages measured measured between different limb electrodes. The augmented limb leads are measured between a limb electrode and a combination of two other limb electrodes. Finally, the precordial leads are measured between electrodes V1-6 and a node called the Wilson terminal, a combination of 3 limb electrodes.

The limb electrodes are named RA (Right arm), LA (Left arm), LL (Left leg) and RL (Right leg). RL is only used for noise reduction (right leg drive)[21]. The limb leads (I, II, III) and the augmented limb leads(aVR, aVL, aVF) are linear combinations of the limb electrodes.

The limb leads are defined in the following way:

$$I = LA - RA \tag{2.1}$$

$$II = LL - RA \tag{2.2}$$

$$III = LL - LA \tag{2.3}$$

In figure 2.5 it is clear that these leads form a closed loop. Using Kirchhoff's second law, which states that at any instant, "The directed sum of the electrical potential differences (voltage) around any closed network is zero"[22], the following relationship can be found moving in the clockwise direction in figure 2.5 :

$$I + III - II = 0$$

Rearranging we get:

$$II = I + III \tag{2.4}$$

This equation is called Einthoven's law, meaning that if the three limb leads are measured simultaneously, lead II will be the sum of lead I and lead III at any instant.[23]

Due to the limb leads being measured between two electrodes, they are commonly called bipolar leads. Frank Norman Wilson wanted to investigate if it was possible to have a unipolar lead. For that a reference that was independent of the rest of the electrodes would be needed. Wilson proposed using the average of the limb electrodes (except RL) for this purpose. This was achieved by connecting a 5kΩ resistor to each of the

electrodes, RA, LA and LL. This terminal was used as the negative reference.[24] The Wilson central terminal is therefore defined as:

$$Wv = \frac{1}{3}(RA + LA + LL) \tag{2.5}$$

Wilson proposed 3 new leads called VR, VL and VF. These were defined as:

$$VR = RA - Wv = \frac{2RA - LA - LL}{3} \tag{2.6}$$

$$VL = LA - Wv = \frac{2LA - RA - LL}{3} \tag{2.7}$$

$$VF = LL - Wv = \frac{2LL - RA - LA}{3} \tag{2.8}$$

Goldberger realised that the resistor connected to the positive electrode is not strictly necessary, and results in decreased amplitude. Therefore he proposed using the average of the two other limb electrodes, to increase the amplitude by 50%.[24] This resulted in the augmented limb leads defined as[15]:

$$aVR = RA - \frac{1}{2}(LA + LL) = \frac{3}{2}(RA - V_w) = \frac{2RA - LA - LL}{2} \tag{2.9}$$

$$aVL = LA - \frac{1}{2}(RA + LL) = \frac{3}{2}(LA - V_w) = \frac{2LA - RA - LL}{2} \tag{2.10}$$

$$aVF = LL - \frac{1}{2}(RA + LA) = \frac{3}{2}(LL - V_w) = \frac{2LL - RA - LA}{2} \tag{2.11}$$

Note that aVR, aVL and aVF (and VR, VL, VF) are redundant with respect to I, II and III. No additional information is contained in augmented leads as they are all derived from the measurement of the same 3 points (RA,LA,LL). However for visualization they are very useful as they can be seen as viewing the electrical activity from different angles[25] (see figure 2.3).

Equations (2.1),(2.2),(2.3),(2.9),(2.10) and (2.11) can be described in matrix form as:

$$\underbrace{\begin{bmatrix} I \\ II \\ III \\ aVR \\ aVL \\ aVF \end{bmatrix}}_{L} = \underbrace{\begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} RA \\ LA \\ LL \end{bmatrix}}_{N} \tag{2.12}$$

2.6 QRS detection

For many types of ECG applications a reliable automatic detection of QRS-complexes is necessary. The QRS detector used in this project is described in detail in [26], and the implementation is available from `https://github.com/tru-hy/rpeakdetect`. The algorithm consist of the following steps:

1. Low pass and high pass filtering with zero phase implementation
2. Differentiation and squaring of the signal
3. Energy thresholding to remove impact of noise spikes from muscle artifacts
4. Shannon energy computation to reduce differences between succesive R-peaks
5. First order gaussian differentiatior to locate peaks in the feature signal

The algorithm have been tested on the MIT-BIH database in [26], and achieved an average sensitivity of 99.94% and a positive predictivity of 99.96%.

2.7 ECG Wave Delineation

An ECG wave delineator expands the detection of QRS complexes to detect the P and T waves, as well as the onset and end of these waveforms. This can be used for automatic detection of intervals such as ST segments and QT intervals.

The software used for detecting the waves of the ECG in this project was ecgpuwave [27], an automatic ECG wave delineator available from PhysioNet[28]. Ecgpuwave is

based on the algorithm described in [29]. Ecgpuwave returns a vector of sample values and a vector of the annotation type. It also returns a subtype, which contains more information of each wave, for example if a wave is biphasic, positive or negative.

The algorithm works by first filtering the signal and then detecting the QRS complexes in every lead. Alternatively a QRS annotation can be provided to the algorithm. Then the algorithm detects the waveform boundaries in each lead, and the annotations from the lead with the longest lasting electrical activity is used. For more details and an evaluation of the algorithm see [29] and [30] respectively.

2.8 Zero phase filtering

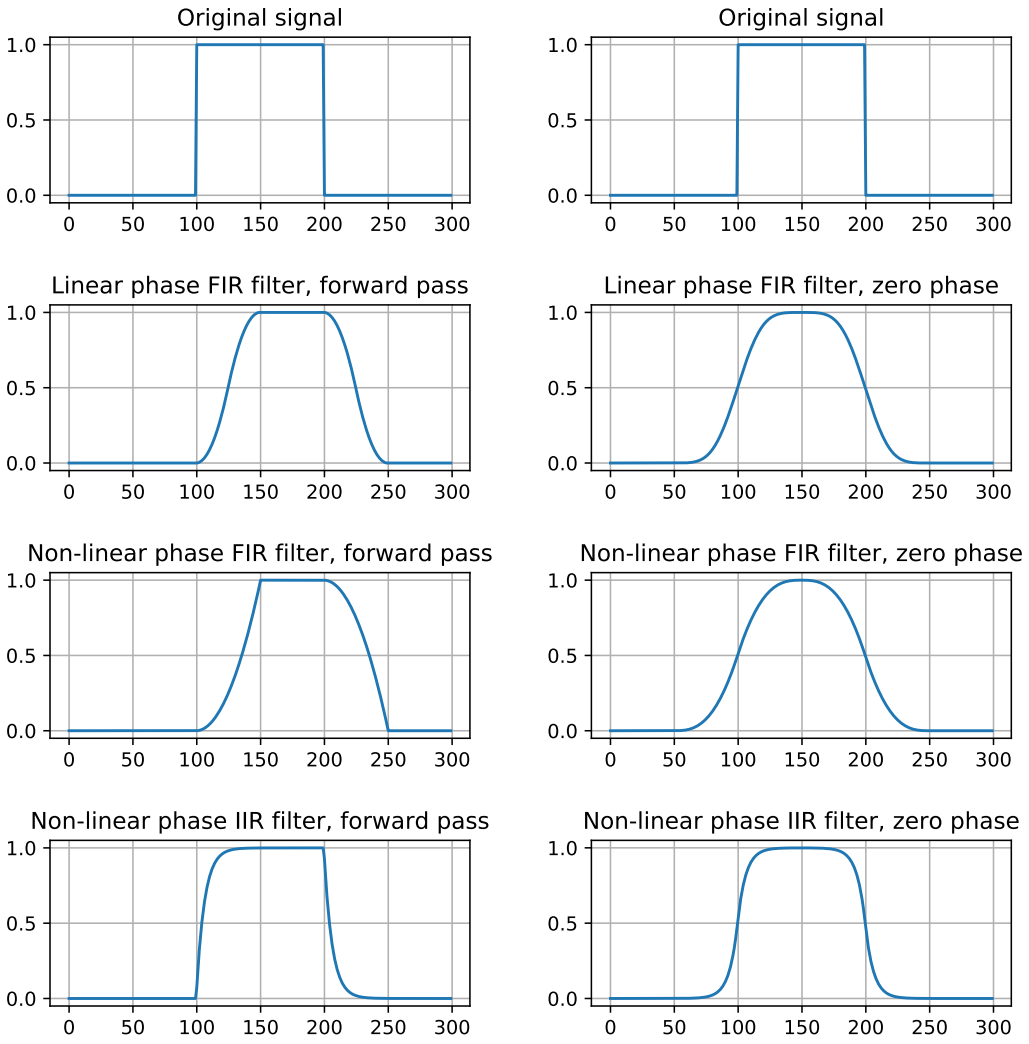

FIGURE 2.6: Demonstration of zero-phase filtering on a rectangular pulse.

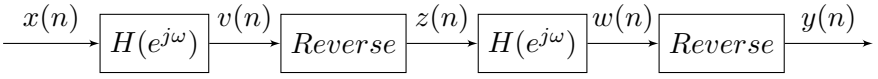

FIGURE 2.7: Block diagram of forwards-backwards filtering

When designing filters it is often desired to design a filter with a linear phase response. A linear phase response means that the the phase response of a filter is a linear combination of the frequency. This will normally result in a delay, though every frequency will be delayed the same amount. This delay is often called the phase delay[31]. While a non-linear phase filter will distort the appearance of a signal, a linear phase filter will only delay it. This is a very useful property when dealing with electrocardiograms, since even minor changes in the appearance may have significant impact on the interpretation.

For digital filters a linear phase filter can be easily designed by using a FIR (finite-duration impulse response) filter. This is achieved by restricting the impulse response to be symmetric or anti-symmetric[32]. In practise a IIR (infinite-duration impulse response) filter is preferred because it requires a lower order for the same magnitude response, however these filters are recursive, and will case phase distortion[33]. If real time performance is not required, a way to solve this is to use zero-phase or forward-backward filtering. This is not commonly discussed in many textbooks, but it is implemented in software like MATLAB [34] and SciPy[35].

The forward-backward filtering works by first filtering the signal with in the forward direction. Then the signal is time reversed and filtered again with the same filter. The result is then time reversed again. If the input signal is denoted x(n), the output y(n) and impulse response of the filter is h(n), we have the following[36]:

$$y(n) = filtfilt(x(n), h(n)) \tag{2.13}$$

$$v(n) = x(n) * h(n), \tag{2.14}$$

$$w(n) = v(-n) * h(n), \tag{2.15}$$

$$y(n) = w(-n) \tag{2.16}$$

By using the Fourier transform we can find the effective frequency response:

$$H_{eff}(e^{j\omega}) = \frac{Y(e^{j\omega})}{X(e^{j\omega})} \tag{2.17}$$

Using the time-reversal property of DTFT:

$$x[n] \overset{DTFT}{\leftrightarrow} X(e^{j\omega}) \tag{2.18}$$

$$x[-n] \overset{DTFT}{\leftrightarrow} X^*(e^{j\omega}) \tag{2.19}$$

We have the following (see figure 2.7)[36][37]:

$$V(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega}) \tag{2.20}$$

$$Z(e^{j\omega}) = V^*(e^{j\omega}) = H^*(e^{j\omega})X^*(e^{j\omega}) \tag{2.21}$$

$$W(e^{j\omega}) = H(e^{j\omega}) * Z(e^{j\omega}) = |H(e^{j\omega})|^2 X^*(e^{j\omega}) \tag{2.22}$$

$$Y(e^{j\omega}) = V^*(e^{j\omega}) = |H(e^{j\omega})|^2 X(e^{j\omega}) \tag{2.23}$$

Therefore the effective response of the forward-backward filtering is:

$$H_{filtfilt}(e^{j\omega}) = |H(e^{j\omega})|^2 \tag{2.24}$$

Equation (2.24) shows that the effective amplitude response after the $filtfilt$ operation is squared and the phase response is cancelled completely.

It should be noted that by using $filtfilt$, the original causal filter is turned into a non-causal filter. This means that a change in the original signal can be visible earlier in the filtered signal. This is visible in figure 2.6, where the signal does not change until n=100, but in the zero phase filtered versions the signal is visibly changed before n=75.

2.9 Modified Gaussian function

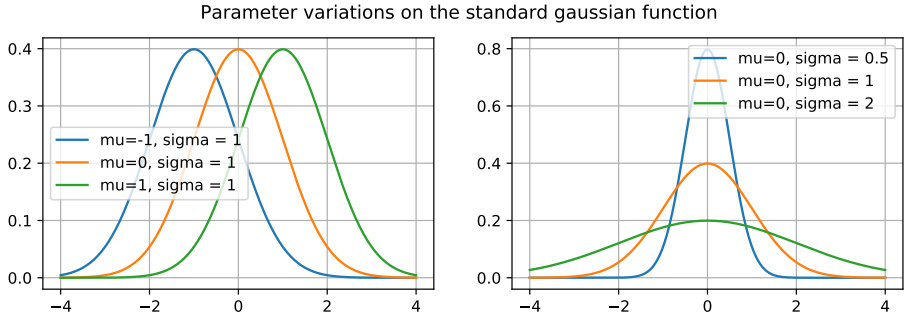

FIGURE 2.8: Gaussian shown with different parameter variations

To represent the beat of an ECG as a sum of few gaussians, some modifications to the standard gaussian function was made. Since the gaussian function is the probability density function of the normal distribution, it is often defined as[38]:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{2.25}$$

The function has 2 parameters, μ which determines the placement of the peak, and σ which determines the width. The gaussian function is plotted for different parameters in figure 2.8. The term $1/\sqrt{2/\pi}$ is a scaling parameter that forces the integral of f(x) to 1. Instead this term is called a and used as a parameter for the height of the wave:

$$f(x) = ae^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{2.26}$$

The function f(x) is defined for all x, however when $|x - \mu|$ is large it approaches 0. If we consider the position where the amplitude is $r_{th} \cdot a$ where r_{th} is the amplitude of the wave divided by the amplitude at the start and end of the wave, we can find the limits of the wave as:

$$ae^{\frac{-(x-\mu)^2}{2\sigma^2}} = r_{th}a \tag{2.27}$$
$$e^{\frac{-(x-\mu)^2}{2\sigma^2}} = r_{th} \tag{2.28}$$
$$\frac{-(x-\mu)^2}{2\sigma^2} = ln(r_{th}) \tag{2.29}$$
$$-(x-\mu)^2 = ln(r_{th})2\sigma^2 \tag{2.30}$$
$$x - \mu = \pm\sqrt{-ln(r_{th})2\sigma^2} \tag{2.31}$$
$$x = \pm\sqrt{-2ln(r_{th})}\sigma + \mu \tag{2.32}$$

The length of the wave is then $x_{max} - x_{min}$:

$$L_w = x_{max} - x_{min} \tag{2.33}$$
$$L_w = \sqrt{-2ln(r_{th})}\sigma + \mu - \sqrt{-2ln(r_{th})}\sigma - \mu \tag{2.34}$$
$$L_w = 2\sqrt{-2ln(r_{th})}\sigma \tag{2.35}$$

With $r_{th} = 0.04$ we have:

$$L_w = 5.0745\sigma \tag{2.36}$$
$$\sigma = 0.1971L_w \tag{2.37}$$

A bias parameter is also added. The modified gaussian(MG) function defined as:

$$f_{MG}(x) = ae^{\left(\frac{-(x-\mu)^2}{2(0.1971L_w)^2}\right)} + bias = ae^{\left(\frac{-(x-\mu)^2}{0.07767L_w^2}\right)} + bias \tag{2.38}$$

The modified gaussian function is plotted for different parameters in figure 2.9.

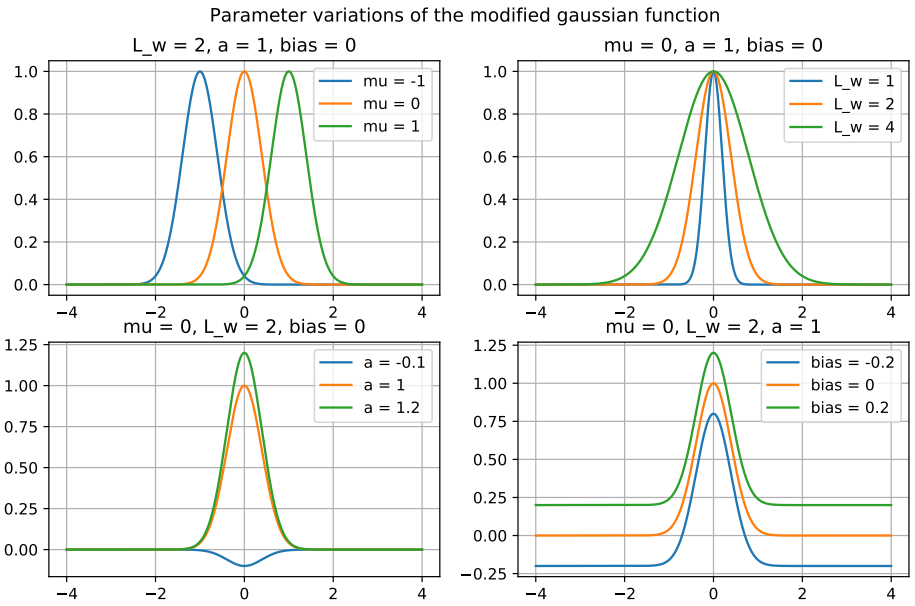

FIGURE 2.9: Modified Gaussian shown with different parameter variations

To represent more complicated waveforms a sum of several MG functions can be used. $bias_{tot}$ represents the average bias of all the MG functions.

$$f_{MG,tot} = \sum_k a_k e^{\left(\frac{-(x-\mu_k)^2}{0.07767 L_{w,k}^2}\right)} + bias_{tot} \tag{2.39}$$

2.10 Non linear least squares curve fitting

Non linear least squares curve fitting is a technique to fit parameterized nonlinear functions to data in a least squares sense. Often the problem can be solved by approximating the model as a linear one. The parameters are then gradually improved over several iterations[39].

The algorithms used in this projects are Levenberg–Marquardt[40] and Trust Region Reflective[41]. The Levenberg-Marquardt algorithm are used for unbounded problems and the Trust Region Reflective algorithm are used for the bounded problems. Both algorithms will find a local minimum.

The specific implementation used in this project was the scipy.optimize.curve_fit() function available from from SciPy[42].

2.11 K-means clustering

The k-means algorithm is a simple yet commonnly used clustering algorithm. It works by dividing the set of N samples x into k disjoint clusters C. Each cluster is described by the mean of the samples in the cluster. The k-means algorithm aims to find the clusters to minimize the within-cluster sum of squares criterion defined as[43]:

$$\min_{\mu_j \in C} \sum_{i=0}^{n} ||x_j - \mu_i||^2 \tag{2.40}$$

After initializing the means, this can be achieved with repeating two simple steps:

1. Assign each x to the cluster with the nearest mean (euclidian distance)
2. Calculate new mean for each cluster

Repeat step 1 and 2 until there are no changes in the clusters.

3. Data material

The data material used in this project was from two sources, the Life In The Fastlane (LITFL) ECG library[44] and the PTB Diagnostic ECG Database (PTBDB)[45] available from PhysioNet[28]. The LITFL ECG library was used to understand different parts of the ECG, while PTBDB was used for simulation. In the following two sections both databases will be described with more detail.

3.1 Life In The Fastlane (LITFL) ECG library

LITFL[46] is a free medical blog and website about emergency medicine and critical care medicine. The authors of the website are mostly emergency physicians and intensivists based in Australia and New Zealand. The website contains a very large searchable database of electrocardiograms including interpretation. The electrocardiograms are in image format and are therefore not used directly for simulation. Instead the electrocardiograms explained in the LITFL ECG database was used to help interpret and understand the signals observed in the PTB database.

3.2 PTB Diagnostic ECG Database (PTBDB)

The PTB diagnostic database is a large database of digitized ECGs by Physikalisch-Technische Bundesanstalt (PTB), the National Metrology Institute of Germany. The database consists of 549 records from 290 different subjects, where each subject is represented by one to five records. The subjects consists of healthy volunteers and patients with different heart diseases.

Each records include 15 synchronized channels. 12 of these are the leads of the standard 12 lead ECG (I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6), and the other 3 are called Frank leads (VX, VY, VZ). The channels are sampled at 1000 Hz, and has a 16

Diagnostic class	Subjects
Myocardial infarction	148
Cardiomyopathy/Heart failure	18
Bundle branch block	15
Dysrhythmia	14
Myocardial hypertrophy	7
Valvular heart disease	6
Myocarditis	4
Miscellaneous	4
Healthy controls	52
No clinical summary	22
Total	290

TABLE 3.1: Diagnostic class for the subjects in the PTBDB database

bit resolution in the range ± 16.384mV. For each record a detailed clinical summary is available. This includes both age, gender and diagnosis. Some records also includes clinical history, information about medication, interventions, coronary artery pathology, ventriculography, echocardiography and hemodynamics. The clinical summary is not available for 22 subjects. The diagnostic class of the different subjects in the database are presented in table 3.1.

The database was downloaded from PhysioNet[28], and read with wfdb-python[47]. The patient categories in the header files are called Reason of admission and are slightly different from the ones in table 3.1. Table 3.2 shows the number of records by reason of admission as it is presented in the header files.

Reason of admission	Records
Myocardial infarction	368
Healthy control	80
n/a	27
Cardiomyopathy	17
Bundle branch block	17
Dysrhythmia	16
Hypertrophy	7
Valvular heart disease	6
Myocarditi	4
Stable angina	2
Unstable angina	1
Palpitation	1
Heart failure (NYHA 4)	1
Heart failure (NYHA 3)	1
Heart failure (NYHA 2)	1
Total	549

TABLE 3.2: Reason of admission for the records in the PTBDB database

4. Method

The method used in this project is divided into 4 main steps:

1. Processing of records
2. Segmentation
3. Fit sum of modified gaussians to representative beat
4. ECG generation

Figure 4.1 show a simplified block diagram of the method presented in this chapter.

4.1 Processing of records

This section explains the preprocessing of the records and how the records can be converted into voltage potentials which could be generated to an ECG recorder.

4.1.1 Detection of QRS complexes and wave delineation

Before processing the records, a QRS detector was used on the whole database on lead I (channel 0). The QRS detector is decribed in chapter 2.6 and the evaluation of the QRS detections can be found in appendix A.2. The results of the detection was written to a PhysioNet annotation file for further use.

This annotator was then used as a input annotator to ecgpuwave [27], an automatic ECG wave delineator available from PhysioNet[28]. Ecgpuwave is based on the algorithm described in [29]. The algorithm returns a vector of sample values and a vector of the annotation type. It also returns a subtype, which contains more information of each wave, for example if a wave is biphasic, positive or negative. The subtype was not used.

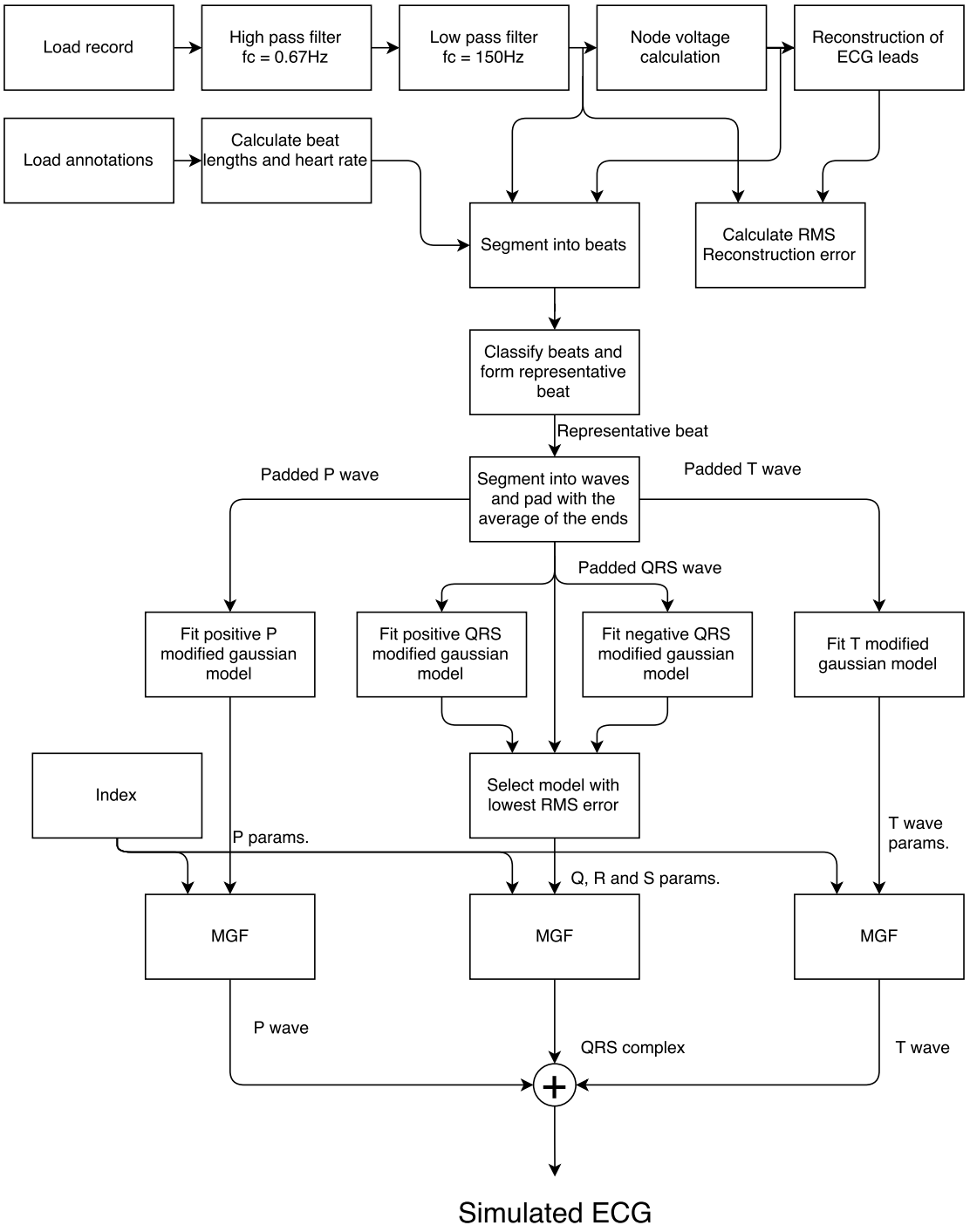

FIGURE 4.1: Overview of the method

The annotations are denoted as following, with i as the annotation index:

$$A_{samp}(i) = sample\ value \tag{4.1}$$

$$A_{type}(i) = sample\ type \tag{4.2}$$

The five available annotation types are:

A_{type}	meaning
'p'	p-wave
'N'	QRS-complex/R-peak
't'	t-wave
'('	onset of wave
')'	end of wave

4.1.2 Calculate RR or beat length

With the QRS annotations it is possible to calculate the RR-interval. The QRS annotations are found using the A_{type} and A_{samp}:

$$A_N = A_{samp}(A_{type} == 'N') \tag{4.3}$$

If f_S is the sample rate of the record. The QRS annotations in milliseconds can be found with:

$$A_{N,ms} = A_N \cdot \frac{1000}{f_s} \tag{4.4}$$

In all records in the PTB Diagnostic ECG database the sample rate is $f_s = 1000$ which means the annotations can be interpreted in milliseconds:

$$A_{N,ms} = A_N \tag{4.5}$$

The RR intervals can be found by taking the first order difference of A_N:

$$RR(i) = A_N(i+1) - A_N(i) \tag{4.6}$$

Although the RR interval lengths are more commonly used for analysis, it is more useful to know the beat length when segmenting an ECG into beats. This can be found by taking the second order centered difference:

$$L_{beat}(i) = \frac{A_N(n+1) - A_N(n-1)}{2}, \quad i \in 1...N-1 \tag{4.7}$$

The length of the first and last beat lengths can be estimated by the forward and backward first order difference:

$$L_{beat}(i) = A_N(n+1) - A_N(0), \qquad i = 0 \tag{4.8}$$

$$L_{beat}(i) = A_N(n) - A_N(n-1), \qquad i = N \tag{4.9}$$

The heart rate is then found as the median of the beat lengths:

$$L_b = median(L_{beat}) \tag{4.10}$$

$$hr = 1000 \cdot 60/median(L_b) \tag{4.11}$$

4.1.3 High pass and low pass filtering

To remove noise, frequency selective filtering is performed on the record. The filter parameters are found in appendix A.1. A second order high pass butterworth filter with $f_{c,hp} = 0.67$ Hz was designed to remove the baseline drift of the signal. A second order low pass butterworth filter with $f_{c,lp} = 150$ was designed to remove high frequency noise such as muscle artifacts. The filtering was performed using forward-backward filtering, resulting in a squared frequency response (described in chapter 2.8). The forward-backward filtering is defined as:

$$v(n) = x(n) * h(n), \tag{4.12}$$

$$w(n) = v(-n) * h(n), \tag{4.13}$$

$$y(n) = w(-n) \tag{4.14}$$

$$y(n) = filtfilt(x(n), h(n)) \tag{4.15}$$

If the signal is called x, the lead number is called l, and the filter response h we have:

$$v_l(n) = filtfilt(x_l, h_{hp}),\ l \in 0..14 \tag{4.16}$$

$$y_l(n) = filtfilt(v_l, h_{lp}),\ l \in 0..14 \tag{4.17}$$

The filtering is performed on all leads in the database.

4.1.4 Node voltage calculation

The idea of calculating the voltage potentials from the leads of a real ECG recording originates from "ECG Simulation"[3]. This approach is somewhat different as it considers all limb leads of the ECG when calculating the voltage potentials instead of just two of the limb leads.

By setting the equations in chapter 2.5 in matrix form we can find a simple way to calculate the ECG leads given the voltage potentials RA, LA and LL. L is defined as a columnn vector of the limb leads and augmented limb leads, M is the matrix describing the ECG relations, and N is a vector of the node voltages. From equations (2.1), (2.2),

(2.3), (2.9), (2.10) and (2.11) we have:

$$\underbrace{\begin{bmatrix} I \\ II \\ III \\ aVR \\ aVL \\ aVF \end{bmatrix}}_{L} = \underbrace{\begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}}_{M} \cdot \underbrace{\begin{bmatrix} RA \\ LA \\ LL \end{bmatrix}}_{N} \tag{4.18}$$

$$L = MN \tag{4.19}$$

For simulation we need to find the node voltage potentials N. M is not square and invertible, and has a rank of 2. This reflects the redundancy of the ECG. To find the node voltages N, we can find a solution to the system by minimizing the 2 norm.

$$e = ||MN - L||^2 \tag{4.20}$$

This can be found by finding the gradient of e and setting it to zero:

$$\nabla(e) = 0 \tag{4.21}$$

$$\nabla(e) = 2M^t(MN - L) = 0 \tag{4.22}$$

$$M^tMN = M^tL \tag{4.23}$$

$$N = (M^tM)^{-1}M^tL \tag{4.24}$$

This is also called the Moore Penrose pseudoinverse[48] denoted M^+.

$$N = M^+L \tag{4.25}$$

$$M^+ = (M^tM)^{-1}M^t \tag{4.26}$$

This can also be used to reconstruct the other leads if 2 or more leads are available. If C is the indexes in L where data is available, the leads can be calculated by:

$$M' = M(C, :) \tag{4.27}$$

$$N = (M'^tM')^{-1}M'^tL \tag{4.28}$$

$$L = MN \tag{4.29}$$

The precordial leads all have their own electrode referenced against the Wilson Central Terminal. Therefore the lead voltage can be used directly as voltage potentials with no

loss of information.

4.2 Segmentation and extraction of representative beat

This section explains how the filtered records are segmented into beats. Then these beats are used to find a representative beat and beat annotation.

4.2.1 Segment ECG into beats

The ECG is segmented into beats by using the annotations from ecgpuwave. To simplify the problem we only consider beats where the annotations include onset, peak and end of P, N and T waves. A normal beat with P, QRS, and T waves will appear as [(,p,),(,N,),(,t,)] in A_{type}, however many other sequences commonly appear in the annotations. Let $A_{N,i}$ be the annotation indexes of A_{type} where A_{type} is 'N'. The indexes for the beats are found with the following precedure:

1. Find where $A_{type} == 'N'$
2. Search for nearest 'p' annotation in the index range$(A_{N,i} - 6 \dots A_{N,i})$
3. Search for nearest 't' annotation in the index range$(A_{N,i} \dots A_{N,i} + 6)$
4. For each annotation find the start and end markers no longer than 2 steps from the annotation in the annotation index
5. Remove any annotation further away from the respective N annotation than L_b (median beat length)
6. Remove any annotation placed before 0 or after the last sample in the record

If any of the annotations are missing, it is deemed invalid and excluded from further analysis. The beats with which are not removed by this criteria are called the valid beats of the record.

To get wave annotations representative for the whole record, the 'N' annotation was subtracted from all the annotations of the valid beats. This results in the peak annotation of the QRS wave, the 'N' annotation, being 0. The onset of the QRS wave and the P wave annotations are negative, and the end of the QRS wave and the T wave annotations are positive. The annotations for the record are then found as the median of each annotation across every valid beat in the record.

The signal for each beat is found by extracting the signal in window with the length of the median beat length, with the peak of the QRS in the center.

The median and mean is then calculated along each sample index across all valid beats of the record. This results in a median and mean beat the same length as the median beat length, where each beat is aligned by the QRS annotation.

4.2.2 Classification of misaligned beats

In the database, there are several records with ectopic beats and noise. The QRS annotations are also not always placed at the same peak in the QRS complex. Therefore a method to find the beats which are aligned in time was required.

As a measure of the error for each beat, the mean squared difference from the median signal was used. If $y_{beat,l}(n)$ is the signal for a single beat and $y_{median,l}(n)$ is the median beat for all beats in the record with l being the lead number. L_{beat} is the length of the beat. The error measure for a beat can be found as:

$$J_e = \frac{1}{15}\sum_{l=0}^{14}\left(\frac{1}{L_{beat}-1}\sum_{n=0}^{L_{beat}}(y_{beat,l}(n) - y_{median,l}(n))^2\right) \tag{4.30}$$

4.2.3 K-means clustering

Based on experimentation, for most of the records the values of J_e had two or more well defined clusters. To find the beats with the lowest deviation, k-means clustering was used with k=2, and the beats in the clusters with the lowest mean were kept. Then the median of these beats along each sample index were calculated. The result of this is a representative beat, with a length equal to the median beat length, and where the index 0 is the QRS annotation.

4.3 Fit sum of modified Gaussians to representative beat

This section shows how the parameters for the modified gaussian functions are found using the representative beat and annotations. The beats are first segmented into a P, QRS and T wave and padded. Then a curve fitting method is used to find the parameters for each wave.

4.3.1 Pad wave with average of the endpoints

The median signal is sliced into a P, QRS and T wave by using the onset and end annotations of the record. The signal is then padded by the average of the endpoints of each wave. The goal of this padding is to make sure the curve fitting will converge to a useful solution. With the signal denoted x_{wave}, and the waves end points denoted A_{onset} and A_{end} the padded signal becomes:

$$x_{wave,pad}(n) = x_{wave},\ n \in A_{onset}..A_{end} \tag{4.31}$$

$$x_{wave,pad}(n) = \frac{x_{wave}(A_{onset}) + x_{wave}(A_{end})}{2}, n \in [A_{onset}, \ldots, A_{onset} + L_{pad}] \tag{4.32}$$

$$x_{wave,pad}(n) = \frac{x_{wave}(A_{onset}) + x_{wave}(A_{end})}{2}, n \in [A_{end} - L_{pad}, \ldots A_{end}] \tag{4.33}$$

$$\tag{4.34}$$

L_{pad} of 100 was used for all waves. Figure 4.2 shows lead I of the waves of patient107/s0199_re. The original signal is plotted in red and the padded signal is plotted in blue.

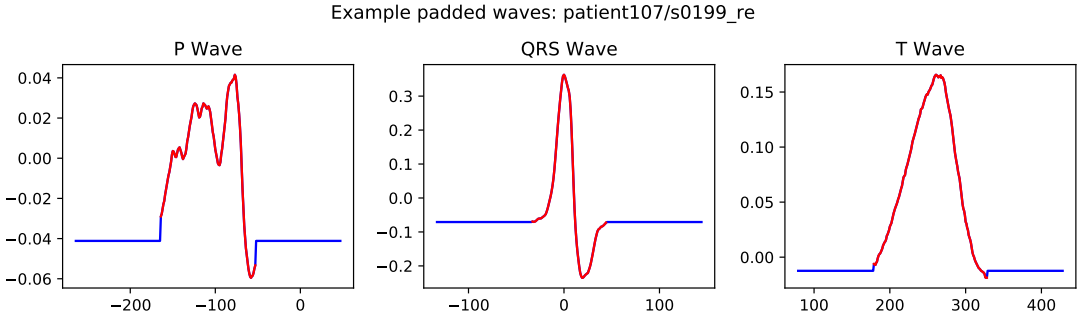

FIGURE 4.2: Example of wave padding with L_{pad} 100 of lead I. The original wave is plotted in red and the padded wave is plotted in blue.

4.3.2 Model fitting

The model selected for the P and T wave was a modified gaussian function. The model selected for the QRS complex was a sum of three modified gaussian functions. The parameters was found using non-linear least squares curve fitting with parameter guesses and bounds found by trial and error. The non-linear least squares problem was solved using the Trust Region Reflective algorithm (briefly explained in chapter 2.10). The initial parameters and bounds found for the P and T waves are found in tables 4.1 and 4.2.

For the QRS wave two different sets of bounds and initial parameters was used. One for positive R waves and negative Q and S waves, and the other one for negative R waves with positive Q and S waves. These parameters can be found in table 4.3 and 4.4. The model is fit using both of these and the model with the lowest mean squared error is selected.

TABLE 4.1: Parameter guesses and bounds for P-wave

	guess	min	max
a	1	-∞	∞
x0	$len(x_p)/2$	$A_{p,min}$	$A_{p,max}$
L_w	$len(x_p)$	0	$len(x_p)$
bias	0	-∞	∞

TABLE 4.2: Parameter guesses and bounds for T-wave

	guess	min	max
a	1	-∞	∞
x0	$len(x_t)/2$	$A_{t,min}$	$A_{t,max}$
L_w	$len(x_t)$	0	$len(x_t)$
bias	0	-∞	∞

TABLE 4.3: Parameter guesses and bounds for positive N-wave. $A_{n,1/3}$ and $A_{n,2/3}$ is the point 1/3 and 2/3 of the way between $A_{n,begin}$ and $A_{n,end}$ respectively.

	guess	min	max
a_0	-1	-∞	0
$x0_0$	$A_{N,min} + 10$	$A_{N,min}$	$A_{N,1/3}$
$L_{w,0}$	$len(x_N)$	0	$len(x_N)$
a_1	1	0	∞
$x0_1$	0	$A_{N,1/3}$	$A_{N,2/3}$
$L_{w,1}$	$len(x_N)$	0	$len(x_N)$
a_2	-1	-∞	0
$x0_2$	$A_{n,max} - 10$	$A_{N,2/3}$	$A_{N,max}$
$L_{w,2}$	$len(x_N)$	0	$len(x_N)$
bias	0	-∞	∞

TABLE 4.4: Parameter guesses and bounds for negative N-wave. $A_{n,1/3}$ and $A_{n,2/3}$ is the point 1/3 and 2/3 of the way between $A_{n,begin}$ and $A_{n,end}$ respectively.

	guess	min	max
a_0	1	0	∞
$x0_0$	$A_{N,min} + 10$	$A_{N,min}$	$A_{N,1/3}$
$L_{w,0}$	$len(x_N)$	0	$len(x_N)$
a_1	-1	-∞	0
$x0_1$	0	$A_{N,1/3}$	$A_{N,2/3}$
$L_{w,1}$	$len(x_N)$	0	$len(x_N)$
a_2	1	0	∞
$x0_2$	$A_{n,max} - 10$	$A_{N,2/3}$	$A_{N,max}$
$L_{w,2}$	$len(x_N)$	0	$len(x_N)$
bias	0	-∞	∞

4.4 ECG generation

In the previous section it was explained how the parameters of the sum of modified gaussians (eq. 2.39) representing each wave was found. This section explains how the model was used to generate new ECG waveforms.

4.4.1 ECG waveform generation

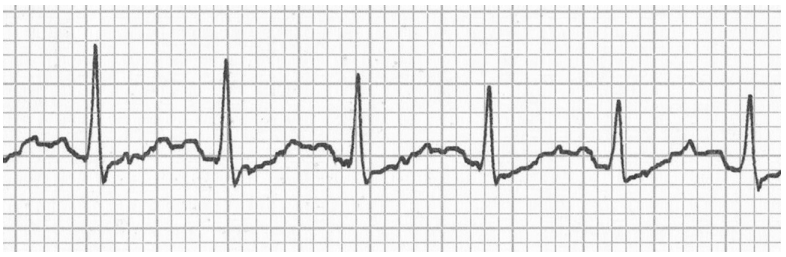

FIGURE 4.3: Tachycardia in an ECG showing "camel hump" appearance. The P wave is hidden in the T wave of the previous beat. (From `https://lifeinthefastlane.com/ecg-library/sinus-tachycardia/`

To generate the simulated ECG, each desired beat of the ECG was generated by evaluating the sum of MGF for twice the length of the median beat. This was then added to a buffer where high frequency noise or baseline drift could also be added to increase realism. To generate an ECG with the same heart rate as the record, a set of desired R wave positions was found using the QRS annotations. By generating the signal this way, it allows the signal from a T wave to be added to the P wave of the next beat. This is similar to the effect found in real electrocardiograms with tachycardia such as in figure 4.3.

4.4.2 QT and QTc correction

If the beat length in milliseconds is denoted L_b and the QT interval is called QT we have the corrected QT interval using Fredericia's formula (selected in appendix A.3):

$$QT_c = QT / \sqrt[3]{L_b/1000} \tag{4.35}$$

This is used as the original QTc. Whenever the heart rate is changed, a new QT interval is calculated using:

$$QT = QT_c \sqrt[3]{L_b/1000} \tag{4.36}$$

A comparison of the different methods of QTc interval calculations can be found in figure 4.4, where Fredericia's method is plotted in orange.

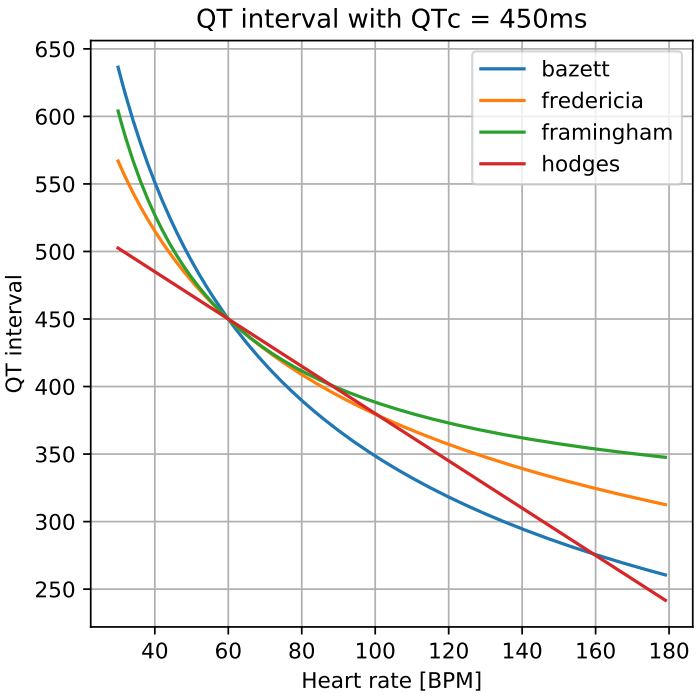

FIGURE 4.4: Comparison of the QT interval with different heart rates and formulas at a QTc of 450.

The new QT minus the old QT is then added to the x0 parameter in the MGF of the T wave. Thus the QT interval changes with heart rate as is to be expected. To simulate signs of different conditions, the length of the QTc interval can be changed. A prolonged QTc interval can be a sign of electrolyte imbalances, myocardial ischemia and is also a common side effect of many drugs.

It was also considered if the length of the PR segment should change based on the heart rate. Therefore a linear regression was performed between the detected PR segments and the heart rate of each record in the PTB Diagnostic database. This experiment is explained in appendix A.4. Based on this experiment it was decided not to change the PR interval based on heart rate.

4.4.3 ST elevation/depression

ST-elevation was added to the signal by adding a window function between the S and T locations of the model parameters. The window function used in the example was a tukey window with a alpha of 0.7, multiplied by the amount of ST-elevation desired. To round off the edges the window function was filtered with a 11 tap boxcar filter with zero-phase implementation.

4.4.4 Inverted waves, hyperacute T-waves, pathological Q waves, R-wave progression

Large T-waves or inverted or missing waves are a common symptom of myocardial ischemia. By modifying the amplitude and width parameters for the T wave, this can be simulated quite easily. A pathological Q wave can also be added to the simulation by changing the amplitude of the first MGF in the QRS complex. Bad R wave progression can be simulated by changing the amplitudes of the R and S waves.

5. Results

This chapter will begin with presenting the results of the voltage potential calculation described in chapter 4.1.4. It will then present the results from the QRS detection and wave delineation, followed by the formation of a representative beat and model fitting. Finally, examples of simulated 12 lead ECG as well as an example of a 12 lead ECG with added symptoms of myocardial infarction will be presented.

5.1 Node voltages and reconstruction of leads

To evaluate the voltage potential calculation only the limb leads and augmented limb leads were considered as these have redundancy. The idea behind the evaluation of the voltage potential calculation is that in a proper synchronized ECG, it should be possible to calculate the voltage potentials, and then use these to reconstruct the original ECG without large error.

For every record in the database, the voltage potentials RA, LA and LL was calculated as described in 4.1.4. The ECG was then reconstructed using the ECG relations and the difference between the original ECG and the reconstructed ECG was found. The RMS reconstruction error is then found as:

$$RMS\ reconstruction\ error = \sqrt{mean([Sig_{orig} - Sig_{reconst}]^2)} \tag{5.1}$$

5.1.1 Experiment 1: RMS Reconstruction error for original and filtered records

The RMS reconstruction error was calculated both for the records with filtering and without filtering for all records in the database. The filter parameters was found based on the experiment in appendix A.1. The mean, standard deviation, minimum, maximum and 25th, 50th and 75th percentiles was found.

	RMS error original	RMS error filtered
mean	0.001718	0.000817
std	0.020940	0.010381
min	0.000127	0.000064
25%	0.000176	0.000075
50%	0.000196	0.000080
75%	0.000222	0.000089
max	0.418093	0.201336

TABLE 5.1: RMS reconstruction error for original data and filtered data

Table 5.1 shows the RMS reconstruction error with and without filtering. The reconstruction error is lower after filtering which might be explained by the filter removing some of the high frequency noise adding to the reconstruction error. Based on this it is concluded that the filtered records are suitable for voltage potential calculation. In the following experiments only the filtered records will be considered.

5.1.2 Experiment 2: RMS Reconstruction error by reason of admission

The RMS reconstruction error of all the filtered records of the database was grouped by reason of admission and the mean, standard deviation, min and max was found for each group.

Table 5.2 shows the reconstruction error grouped by reason of admission. The highest reconstruction errors was found in the patients admitted for Heart failure (NYHA 4),

	N_records	mean	std	min	max
Bundle branch block	17	0.001095	0.004171	0.000066	0.017282
Cardiomyopathy	17	0.000082	0.000014	0.000068	0.000109
Dysrhythmia	16	0.000081	0.000011	0.000065	0.000097
Healthy control	80	0.000081	0.000010	0.000065	0.000103
Heart failure (NYHA 2)	1	0.000065		0.000065	0.000065
Heart failure (NYHA 3)	1	0.000083		0.000083	0.000083
Heart failure (NYHA 4)	1	0.041298		0.041298	0.041298
Hypertrophy	7	0.000080	0.000011	0.000066	0.000096
Myocardial infarction	368	0.000123	0.000725	0.000064	0.013936
Myocarditi	4	0.000071	0.000005	0.000067	0.000077
Palpitation	1	0.000080		0.000080	0.000080
Stable angina	2	0.000072	0.000008	0.000066	0.000078
Unstable angina	1	0.000065		0.000065	0.000065
Valvular heart disease	6	0.000082	0.000013	0.000068	0.000097
n/a	27	0.012315	0.045204	0.000064	0.201336

TABLE 5.2: RMS reconstruction error grouped by reason of admission

Myocardial infarction and the patients without any information about the reason of admission.

5.1.3 Experiment 3: Visualization of the voltage potential calculation

The records with the highest and lowest, as well as median RMS reconstruction error (after filtering), was selected for visual inspection. The original signal (limb leads and augmented limb leads), calculated voltage potentials, reconstructed signal and the difference between the original signal and reconstructed signal was plotted. If the original signal and reconstructed signal look identical, and the difference between them look like flat lines the result is good.

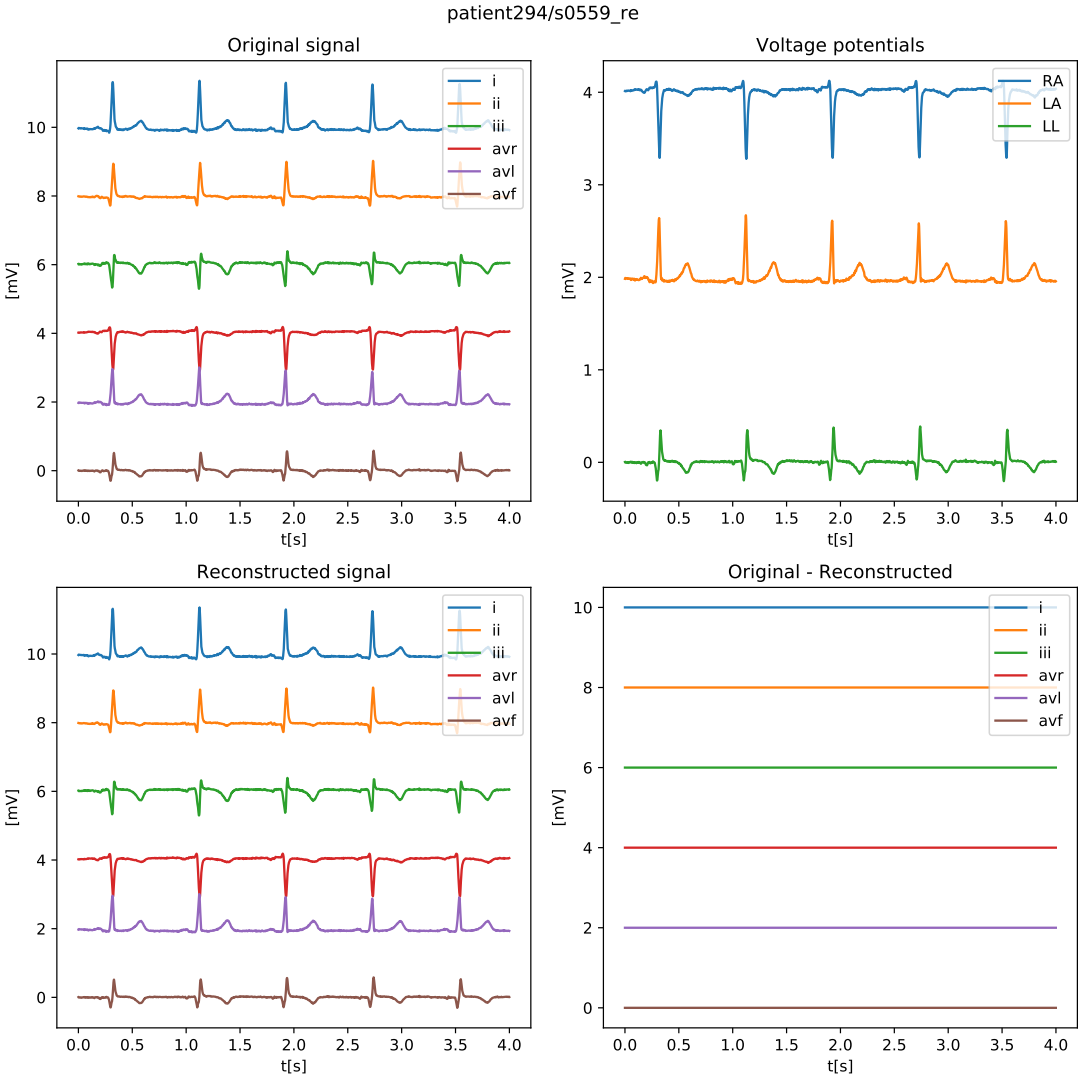

FIGURE 5.1: Example of voltage potentials, reconstructed leads and reconstruction error. Lowest error.

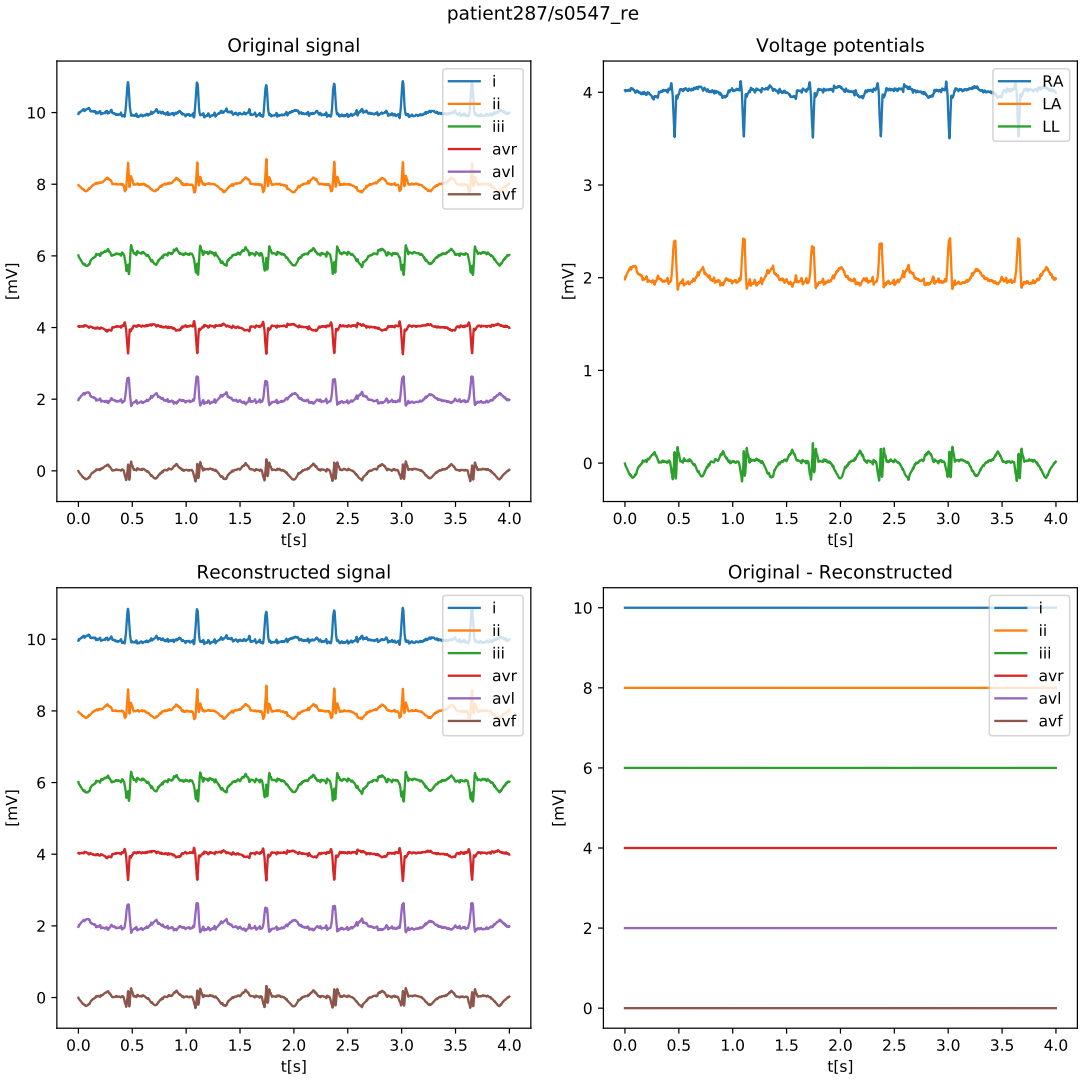

FIGURE 5.2: Example of voltage potentials, reconstructed leads and reconstruction error. Median error.

Figure 5.1 is the plot of the record with the lowest reconstruction error. The original signal and reconstructed signal look identical and the difference plot are flat lines. Thus it is concluded that the voltage potential calculation works well on this record. Similarly, figure 5.2 shows the record with the median reconstruction error, which also shows no visible errors in the reconstruction.

Figure 5.3 shows the record with the highest RMS reconstruction error. The records with a high reconstruction error seem to have high frequency content. The large error in this example originates from the spikes in the signal. The other parts of the ECG seems to reconstruct well.

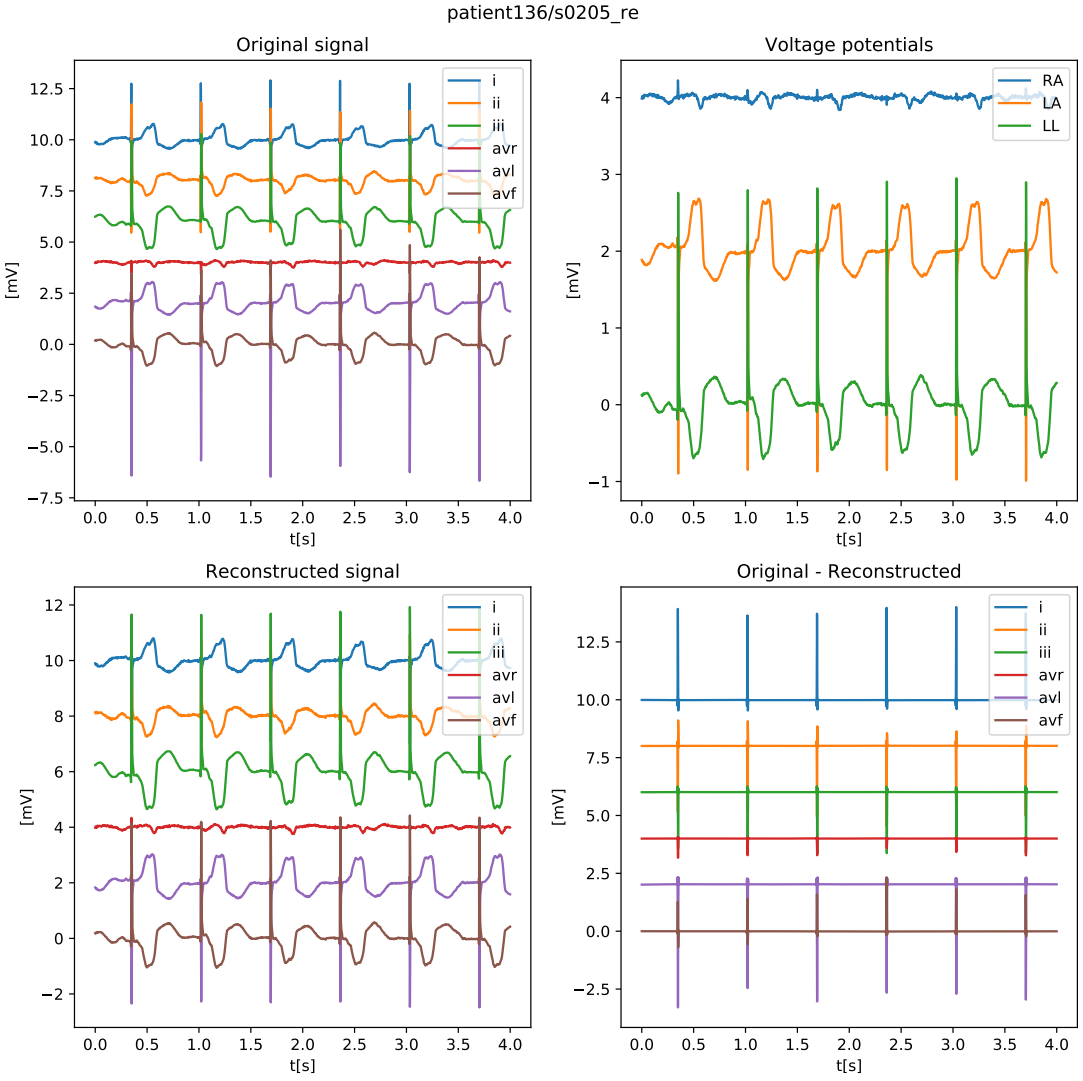

FIGURE 5.3: Example of voltage potentials, reconstructed leads and reconstruction error. High reconstruction error due to high frequency spikes.

5.2 Segmentation

To evaluate the segmentation of the records, the QRS detector described in chapter 2.6 was used on all records. The QRS detector was selected and verified in appendix A.2. These QRS annotations were used as a input to the ECG wave delineator ecgpuwave, which is described in chapter 2.7. The QRS annotations output from this algorithm was called N annotations to avoid confusion. Then the filtering criteria described in 4.2.1 was applied to each record, resulting in a set of valid beats for each record.

5.2.1 Experiment 4: Number of detections

For all records the number of QRS detections, N annotations, valid beats and the percentage of N annotations resulting in a valid beat was found. The mean, standard deviation, minimum, maximum as well as 25th, 50th and 75th percentiles was calculated. A histogram of the percentage of valid beats was also plotted, both for all records and for the healthy controls.

	QRS detections	N annotations	Valid beats	Beats kept
mean	135	134	111	83%
std	40	40	46	24%
min	32	31	0	0%
25%	119	117	81	71%
50%	138	137	120	97%
75%	159	158	142	100%
max	320	317	228	100%

TABLE 5.3: Table of QRS detections, N-annotatations, valid beats and percentage of beats kept by the annotation filtering

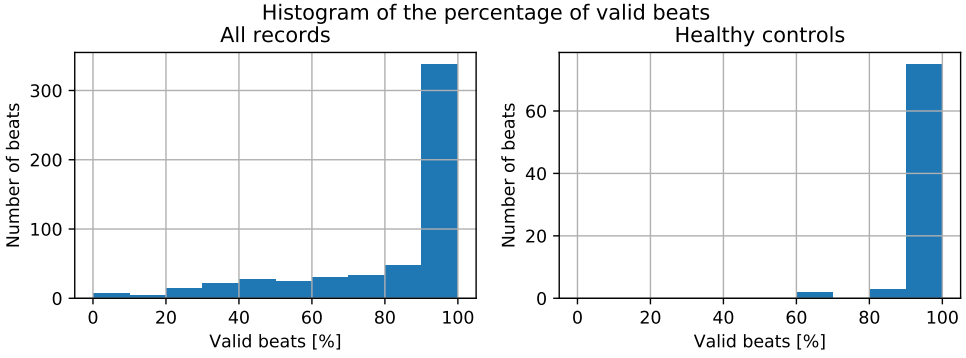

FIGURE 5.4: Percentage of valid beats. Left: All records, Right: Healthy controls

Table 5.3 shows the number of QRS detections in the first column. The next column shows how many N annotations this resulted in. It also shows how many beats were

kept after the annotations were filtered according to the criteria described in 4.2.1. This result shows that ecgpuwave typically removes one of the annotations. This is most likely a annotation too near the beginning or end of the signal. The mean ratio of beats kept is 83% which shows that a significant part of the beats gets removed. For 75% of the records 71% or more of the N annotations are deemed valid. 75% of the records also had more than 142 valid beats. 3 records had no valid beats.

The histogram in figure 5.4 shows the percentage of beats versus the number of records. This shows that for the majority of records the the amount of valid beats are high.

5.2.2 Experiment 5: Visualization of the detection

The three records with the highest percentage of valid beats, and the three records with the lowest percentage of valid beats were selected for manual inspection. The first four seconds of lead I in each record was plotted. If the filtering is reasonable it is expected that the records with the highest percentage of valid beats will show typical electrocardiograms with clearly defined P, QRS and T waves. For the records with the lowest percentage of kept beats, it is expected that the morphology of the ECG will be unusual, and some of the waves might be missing.

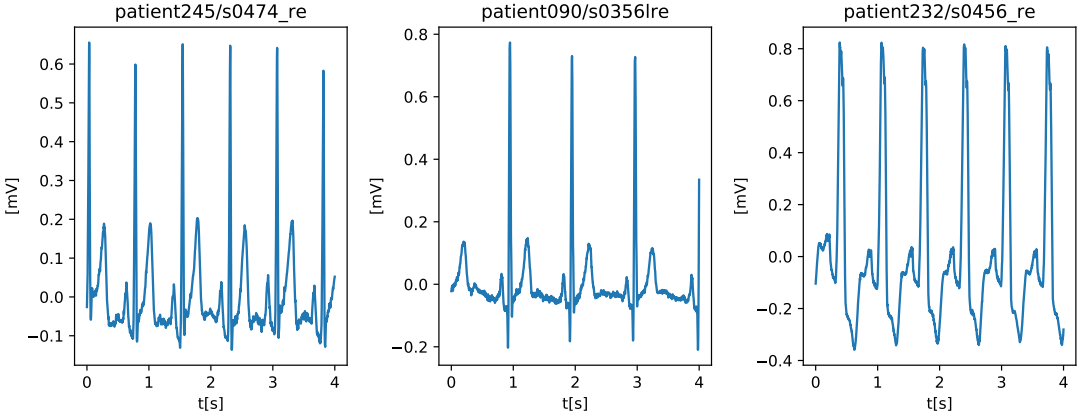

FIGURE 5.5: Lead I of the three randomly selected records out of 170 with 100% valid beats. All of these patients P, QRS and T waves, although "patient232/s0456_re" show ST-depression. Reason of admission from left to right: Myocardial infarction, Healthy control, Hypertrophy

Figure 5.5 shows 3 examples of the 170 records with 100% valid beats. The first two plots show very typical and clearly defined P, QRS and T waves. The last example show ST depression and inverted T waves, but all waves are clearly defined. Thus the records with the highest percentage of valid beats look like expected.

In figure 5.6 lead I of the three only records with no valid beats are plotted. It is clear from the plots that these electrocardiograms are very unusual. The first plot does not

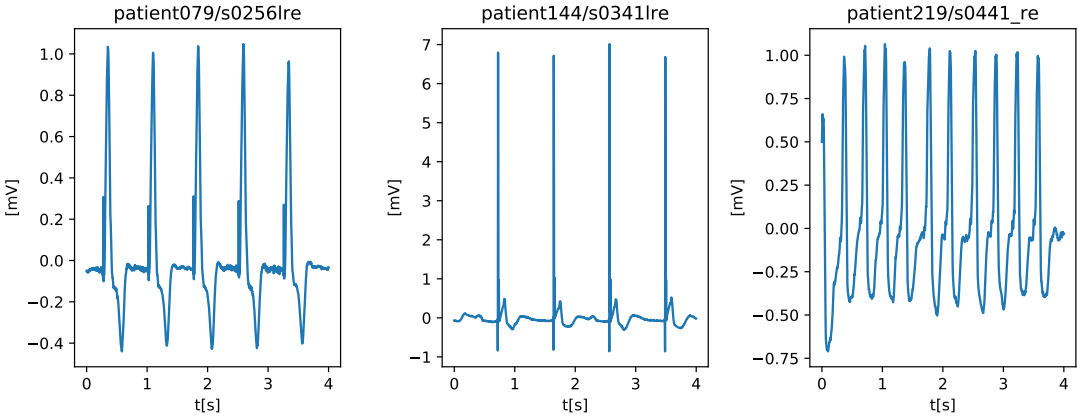

FIGURE 5.6: Lead I of the three records with no valid beats. All of these patients show highly unusual qrs and t morphologies. Reason of admission from left to right: Myocardial infarction, n/a, Bundle branch block

show any P waves. The middle plot shows an occasional P wave but the QRS and ST segments are highly unusual. The last example does not show any clear P or T waveform at all. This is as expected, and it is therefore concluded that the beat filtering criteria works well.

5.3 Representative beat

To find a representative beat of the record based on the valid beats segmented out in the previous chapter, the idea was to find a representative beat by taking the median of all beats in a record at each sample time, thus cancelling out random noise and variations in each beat. Since some of the QRS annotations are not aligned, the k-means algorithm with k=2 was used to remove the beats that was not aligned with the median beat, and the new representative beat was found as the median of all accepted beats at each sample time.

5.3.1 Experiment 6: Number of accepted beats

To evaluate the clustering performed during the representative beat formation, the representative beat for all records in the database with valid beats was found. The number of accepted beats and the percentage of beats accepted was counted for each record. Then the mean, standard deviation, minimum, maximum and 25th, 50th and 75th percentiles was found.

	N accepted	Accept ratio
mean	92	82%
std	41	13%
min	3	43%
25%	64	73%
50%	96	83%
75%	119	95%
max	213	100%

TABLE 5.4: Number and percentage of accepted beats after the k-means clustering, 3 records with no valid beats were not included

Table 5.4 shows the number and percentage of valid beats accepted by the clustering. Records with no valid beats were not included. A a minimum of 43% and a mean of 82% of the beats were kept by the algorithm.

5.3.2 Experiment 7: Representative beat visualization

To visualize the representative beat formation of a record, the steps of this algorithm was plotted as four figures. The record is loaded, filtered and segmented as described previously. The first figure shows lead I of an example valid beat from the record, as well as the mean and the median of lead I of the original beats. The second plot shows a histogram of the error measure used for the clustering. This error measure for one beat

is found by taking the mean squared difference between the beat and the median beat of the record.

The clustering will find two clusters in the error function. In the third plot all the leads from all the accepted beats (cluster with lowest error) are plotted on top of each other in green. All the rejected beats (cluster with highest error) are plotted in red. The median beat is plotted in black. If the clustering is successful the green plots and the red plots will be well separated, with the green plots being very similar to each other. The final figure is an example of a kept beat, a rejected beat and the representative beat.

These plots were reviewed manually for all records in the database with valid beats. The notes from the manual review can be found in appendix A.5. During the manual review 39 records was noted to have some errors, however only 3 records appeared to have a significantly distorted representative beat in lead I. The majority of the errors was due to the Q or S wave being detected as the R wave, resulting in a misalignment of the beat while the shape remained the same.

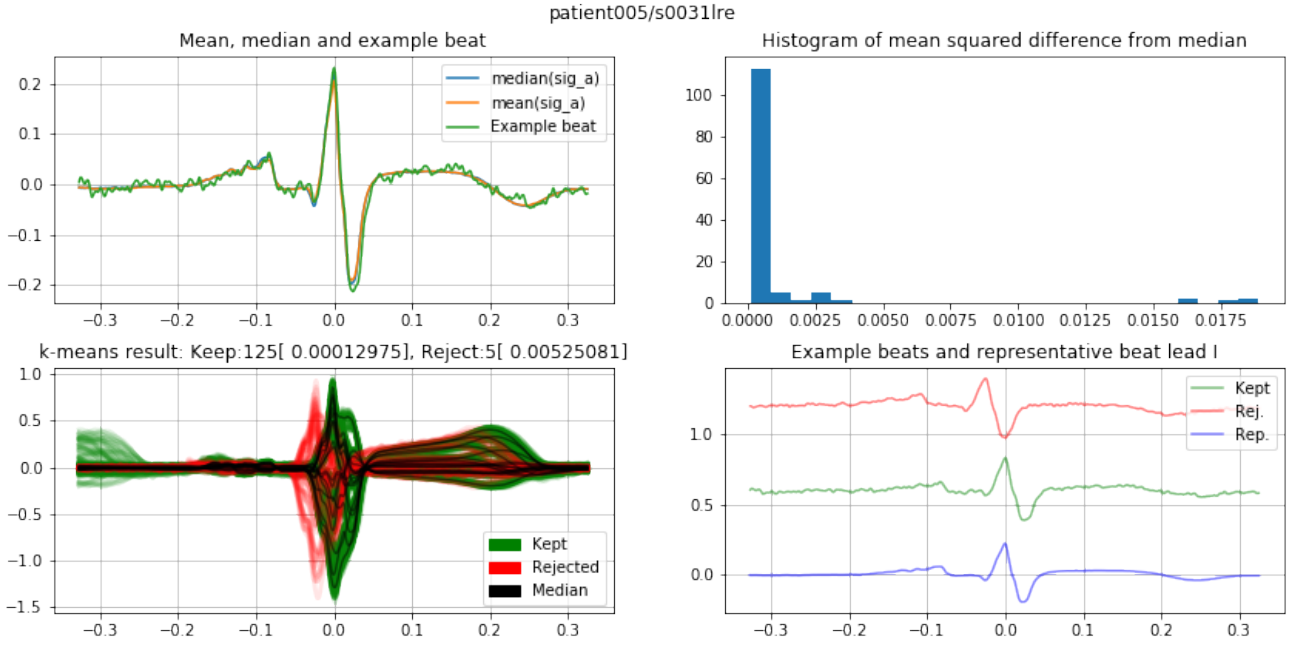

FIGURE 5.7: Representative beat

Figure 5.7 and 5.8 shows two examples of the formation of a representative beat. In figure 5.7 it is demonstrated how the algorithm rejects 5 beats which are clearly misaligned. The representative beat appears much more smooth than the original.

In figure 5.8 all beats seem to be aligned. Since k=2 the clustering will always result in one or more beats being rejected. In this case it seems to have rejected a beat which are significantly more noisy than the others. In general the representative beat appears much more smooth than the original.

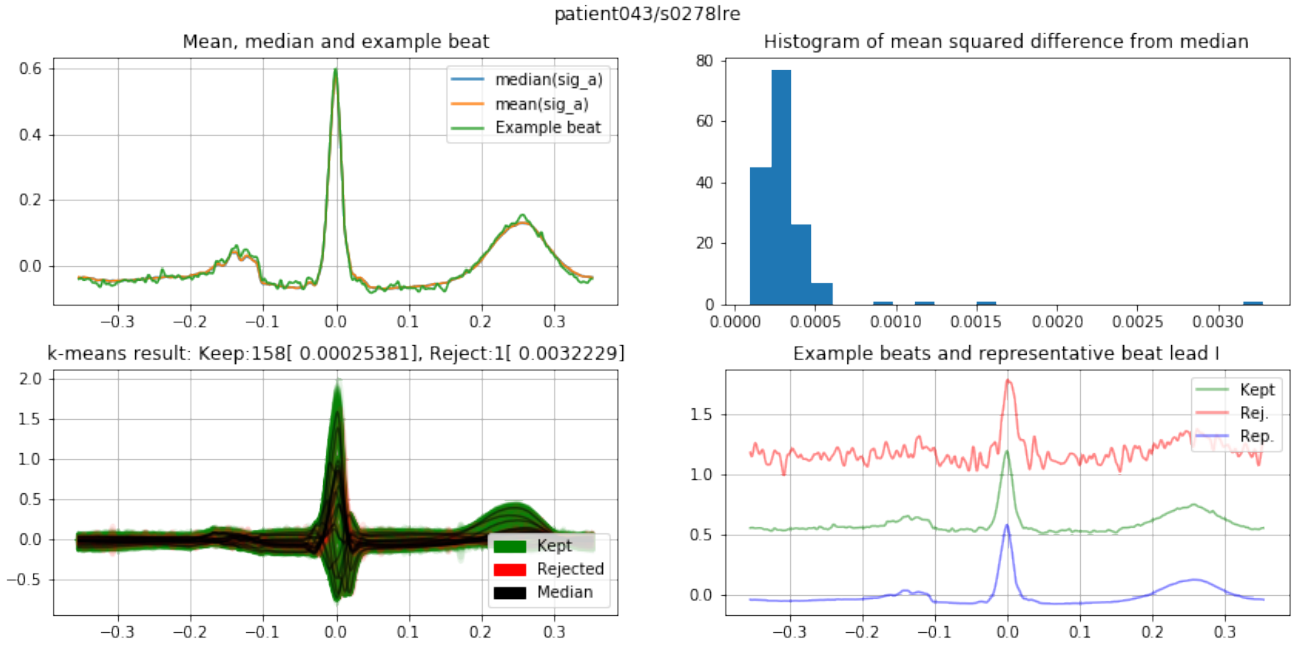

FIGURE 5.8: Representative beat

5.4 Curve fitting

A sum of gaussians was fitted to each lead of the representative beat in every record in the database using the the method described in chapter 4.3. The three records with no representative beat was not included.

5.4.1 Experiment 8: Convergence of the curve fitting

The curve fitting was first evaluated based on the number of record where a model for all leads and all waves (P, QRS T) was found. A record is said to not converge if the curve fitting failed for any of the waves on any of the leads of the record. The success rate was found by the percentage of records converging out of all records in the group.

	Converged records	Failed	Success rate	Total
All leads	439	107	80.4	546
Healthy controls	56	24	70.0	80

TABLE 5.5: Table of convergence rates.

Table 5.5 shows the rate of successfully modeled records and successfully converged records labelled healthy controls. The method found a model for all leads in 80.4% of all records. For the records admitted as healthy controls this number reduces to 70%.

5.4.2 Experiment 9: Convergence and simulation error by lead

To get an idea of which lead the method had most problem modeling, the amount of times the curve fitting failed to find a model was counted for each lead across all records with a representative beat.

The RMS simulation error for a record was found by the following steps:

1. Load record and find the representative beat
2. Generate beat from the model
3. For each lead calculate the RMS difference between representative beat and generated beat
4. Calculate mean RMS difference for each lead
5. Calculate the mean of the mean RMS difference for all leads in the record

The mean RMS simulation error was found for all the records. The leads with no model was not included. The mean of the RMS simulation error was calculated for each lead in all the records.

	Failed leads	Mean RMS error
I	18	0.042680
II	30	0.050056
III	18	0.044680
aVR	21	0.040961
aVL	10	0.034691
aVF	35	0.042452
V1	16	0.061803
V2	20	0.100620
V3	34	0.108767
V4	42	0.089778
V5	34	0.065981
V6	29	0.050751

TABLE 5.6: Mean RMS error by lead simulated ecg. Leads without convergence are excluded

Table 5.6 shows the number of times the method failed to find a model grouped by lead. It also shows the mean RMS difference between the simulated beat and the representative beat. The RMS error is highest in the precordial leads and lowest in aVL. This might be due to these leads typically having different amplitudes.

5.4.3 Experiment 10: Visualization of the simulation error

To visualize the location of the simulation error in each lead a plot of the mean simulation for all leads in all records was made. The plot was generated by the following steps:

1. Find representative beat for all records
2. Fit modified gaussian to every representative beat
3. Generate beat from the model
4. Find RMS difference between representative beat and simulated beat
5. For each lead, at every sample time with the QRS annotation centered at 0 calculate the mean

This is plotted for each lead. The plots represent the mean RMS difference between the representative beat and simulated beats for all leads across all records with valid beats

or all healthy control records. This means that for each index each index in the signals a mean is calculated. The signals are aligned by the QRS peak which has the index 0.

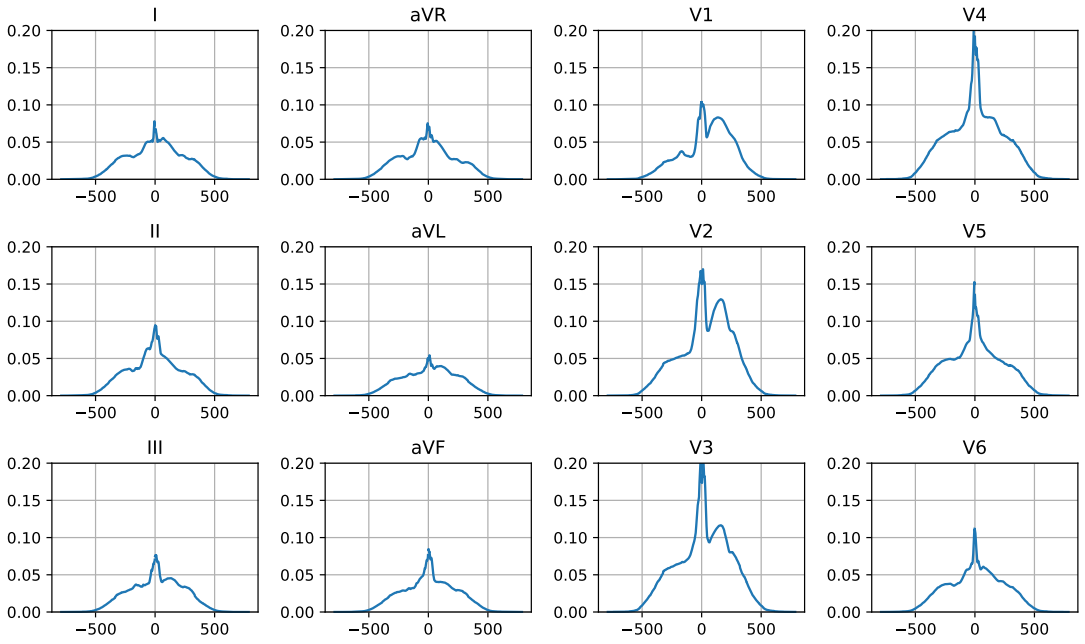

FIGURE 5.9: Mean RMS error along time axis for all converged leads.

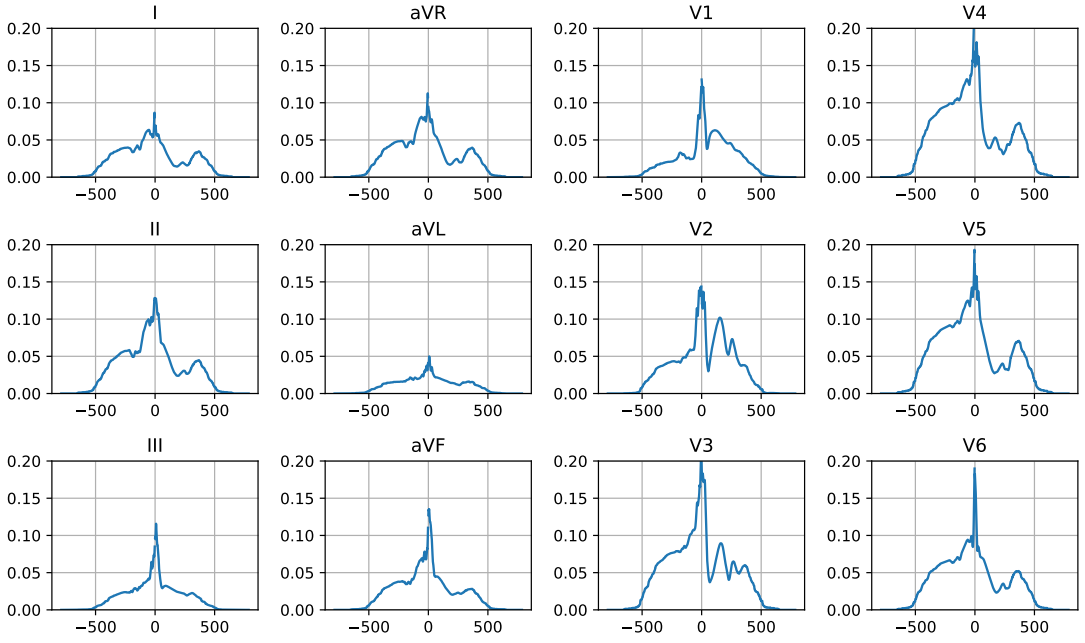

FIGURE 5.10: Mean RMS error along time axis for all converged leads in patients admitted as healthy controls.

Figure 5.9 and 5.10 shows the result of this calculation for all records and healthy control records respectively. The index 0 on the horizontal axis represents the QRS annotation. The figure show how the segment after the QRS complex has higher error, especially

in the precordial leads in all records compared to just the healthy controls. This might be because the symptoms of myocardial infarction often include ST elevation in those leads.

5.5 Examples of simulated ECG

The examples of simulated 12 lead ECG was compared to the representative beat by plotting each lead in the same figure. Three different records are displayed. The first one is a record where the simulation works quite well, and a model was found for all leads. The second example shows how the simulation works decently for all records except the lead aVR where no model was found. It is then showed how lead aVR can be reconstructed from the other simulated limb leads. An example where the simulation did not work so well is also presented. Finally it is showed how ECG can be simulated at different heart rates by changing the model, as well as how myocardial infarction can be simulated by changes in the model.

5.5.1 Experiment 11: Example simulated ECG

To display a simulated 12 Lead ECG each lead of the simulation was plotted in separate subplots along with the representative beat of the same record. The plot were made by the following steps:

1. Find representative beat of the record
2. Fit modified gaussian to representative beat
3. Generate each lead of the ECG for the same index as the representative beat, R-peak centered at 0
4. Plot the lead of the representative beat and generated ECG in the same figure

If the simulated ECG is able to represent the representative beat well the blue and orange plot will be similar. A lead where the model fitting failed will be plotted as a straight line.

Figure 5.11 shows an example of a simulated beat from the model in blue. This is compared to the representative beat for the record it was based on in orange. The difference is most pronounced where the waves of the representative beat are not symmetric as in lead V2 and V3 most likely due to ST elevation.

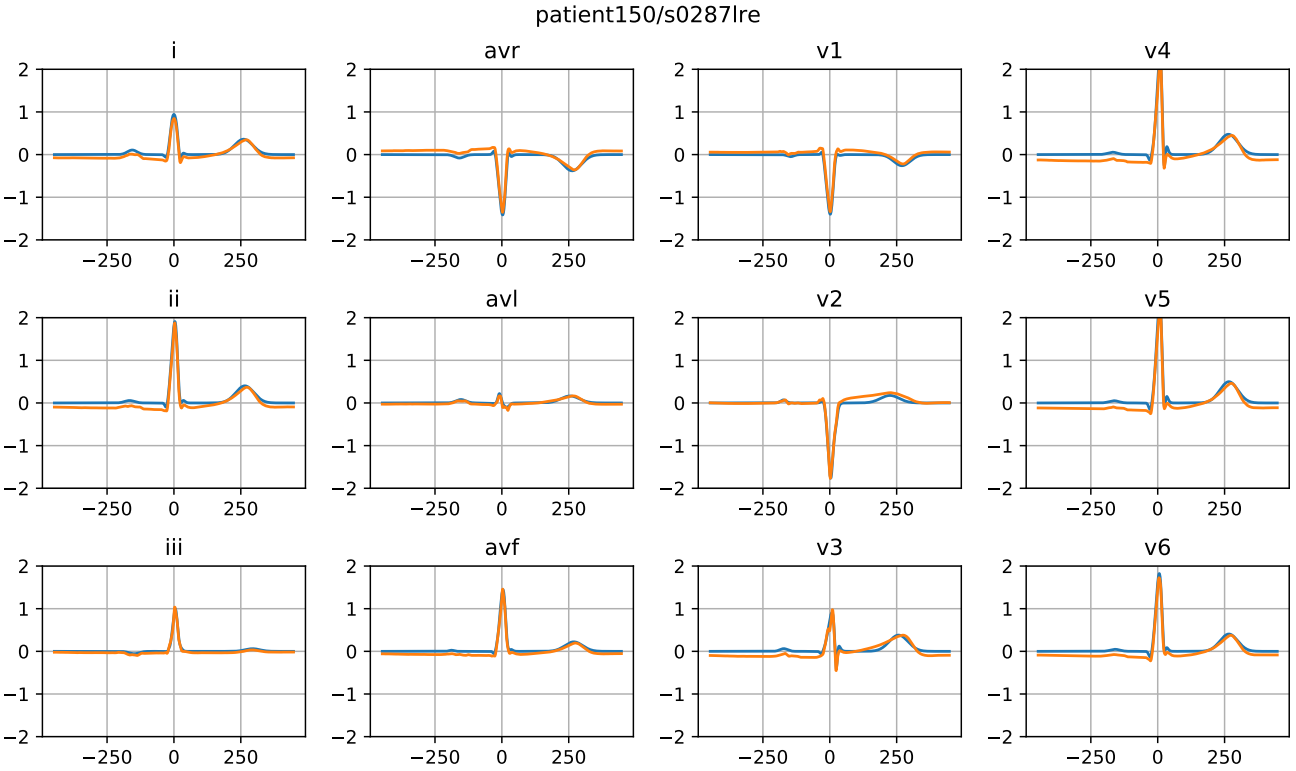

FIGURE 5.11: patient150/s0287lre: Simulated beat(blue) compared to representative beat (orange)

5.5.2 Experiment 12: Reconstruction of a limb lead from the other simulated limb leads

The plot of the record is constructed the same way as in Experiment 11. A record where the model fitting only failed on the aVL lead was selected. The leads with no model are plotted as a flat line. The lead aVL was then reconstructed from the other leads of the simulated ECG using the relations in chapter 4.1.4. The representative beat, simulated and reconstructed lead aVL was then plotted in the same figure.

Figure 5.12 shows an example record where the algorithm did not find a model for the QRS complex in lead aVR. As in the previous example the model fits the representative beat quite well for the other leads. With some errors in the ST and T segment in the precordial leads.

Figure 5.13 shows the reconstruction of lead aVR from the other simulated leads. The reconstruction seems to fit the original quite well. This suggests that the simulation works well enough to follow the ECG relations explained in chapter 4.1.4.

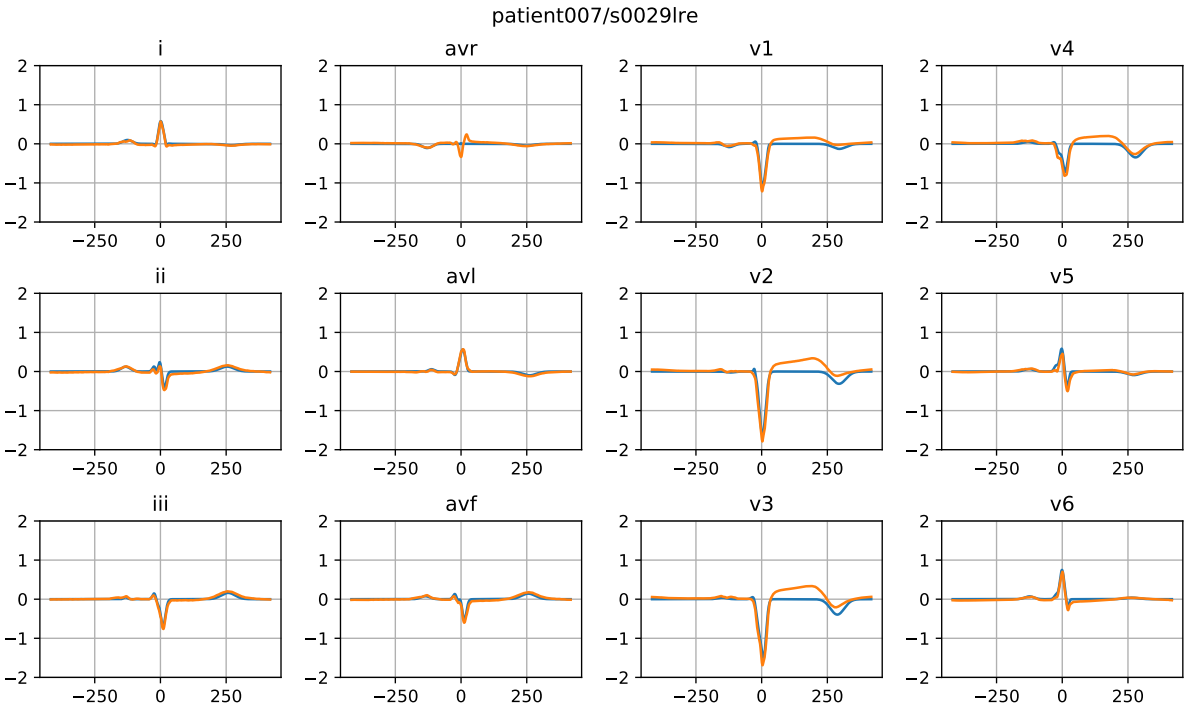

FIGURE 5.12: patient007/s0029lre: Simulated beat(blue) compared to representative beat(orange)

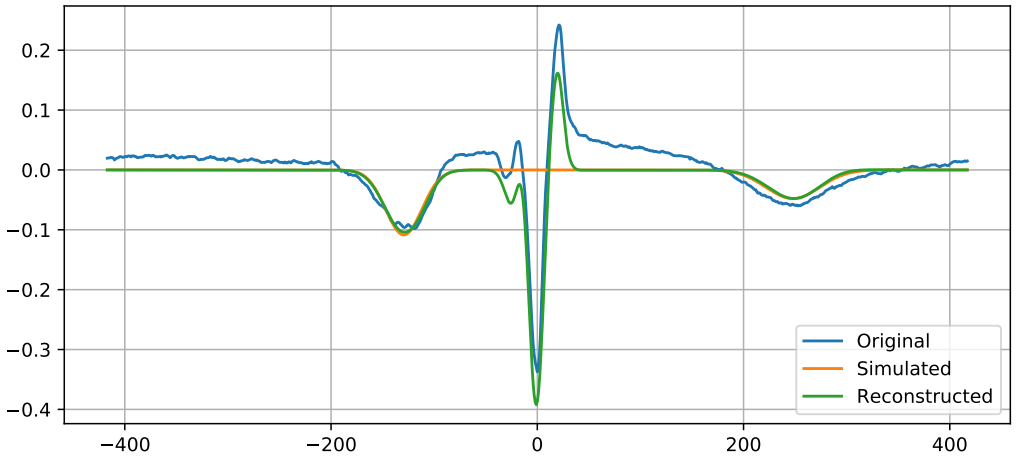

FIGURE 5.13: patient007/s0029lre: lead aVL reconstructed from the other simulated limb leads

5.5.3 Experiment 13: Example of simulation with high error

Figure 5.14 shows an example of a simulated record with high error. The model is plotted in blue and the representative beat of the record is plotted in orange in the same way as in experiment 11. The errors are highest in the ST segment and T wave, and

show how the model are unable to represent the more advanced morphologies present in this record.

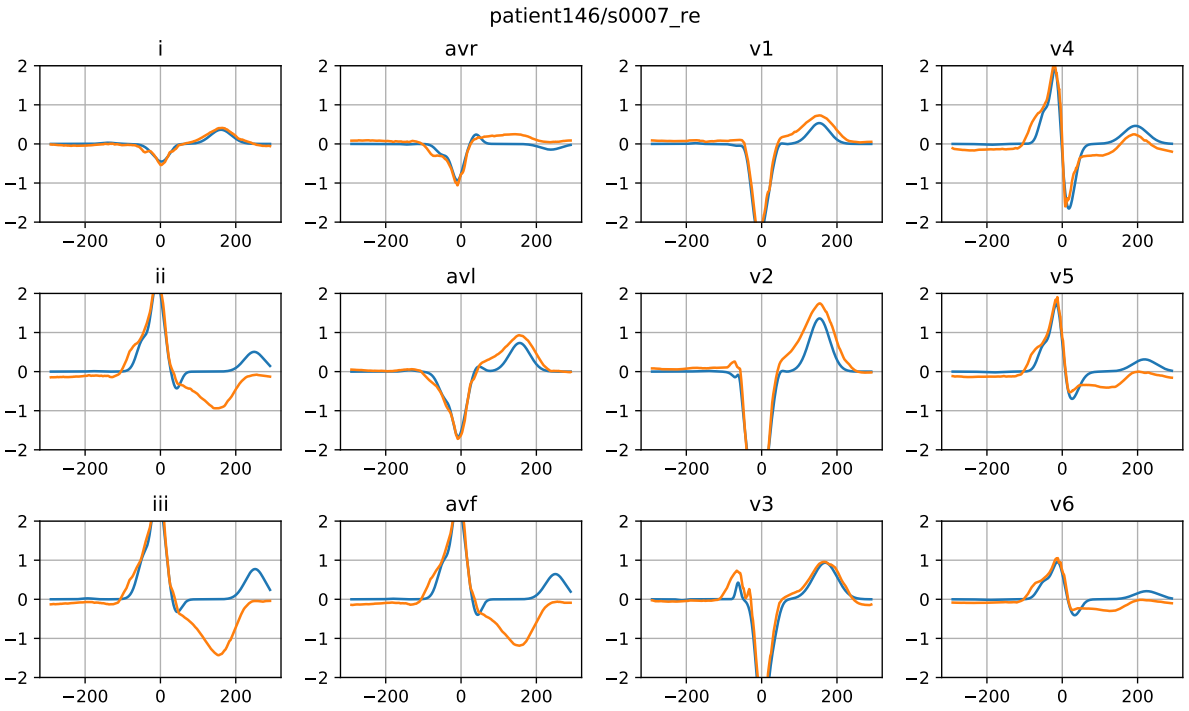

FIGURE 5.14: patient146/s0007_re: Simulated beat(blue) compared to representative beat(orange) for the record with highest simulation error.

5.5.4 Experiment 14: Rate change

To display how the simulated ECG changes with heart rate, the first 2.5 seconds of lead II and lead V4 in patient263/s0499_re was plotted. The heart rate of the record was found based on the median beat length of the record. The representative beat of the record was found and the model was used to generate 2.5 seconds of ECG at the record's heart rate. 2.5 seconds of simulated ECG was then generated at heart rates of 40, 70, 90 and 150 BPM. As explained in chapter 4.4.2 the QT interval of the model is changed according to the model. No change in the other parameters of the model was performed.

Figure 5.15 and 5.16 shows plots of the simulated lead II and lead V4 of the record at different heart rates. The original lead is plotted in the top left plot. The plot show that the QT interval will decrease with increasing heart rates as expected. For a heart rate of 150 beats per minute the P wave gets added to the T wave of the previous heart beat. In figure 5.15 this results in a "camel hump" appearance. In 5.16 the P wave disappears into the T wave. This is commonly observed in real electrocardiograms with tachycardia[49].

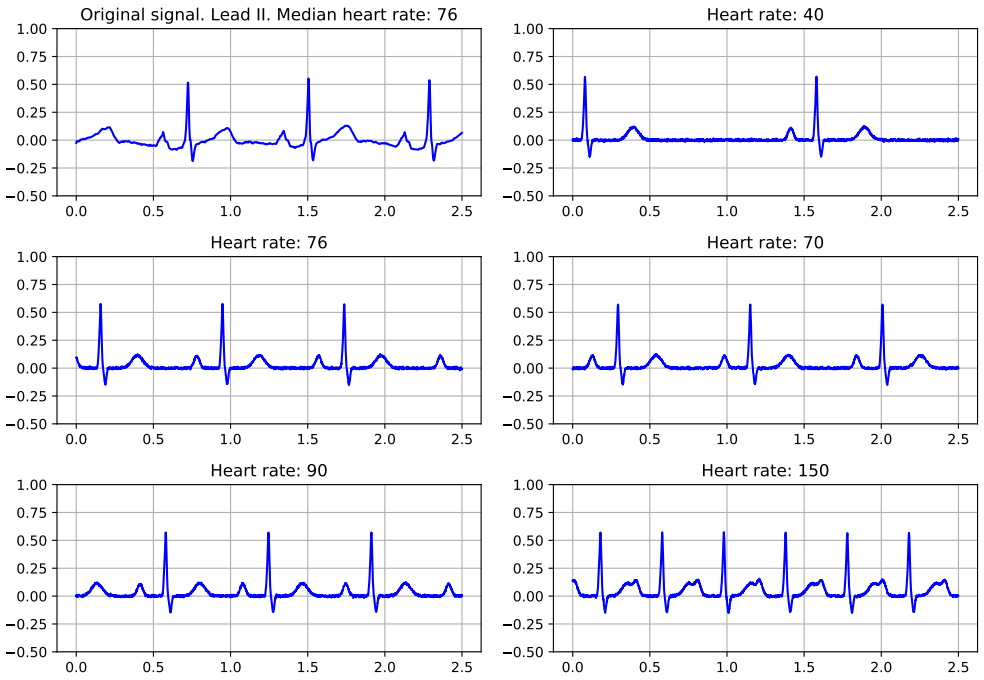

FIGURE 5.15: patient263/s0499_re: Lead II rate adjustment.

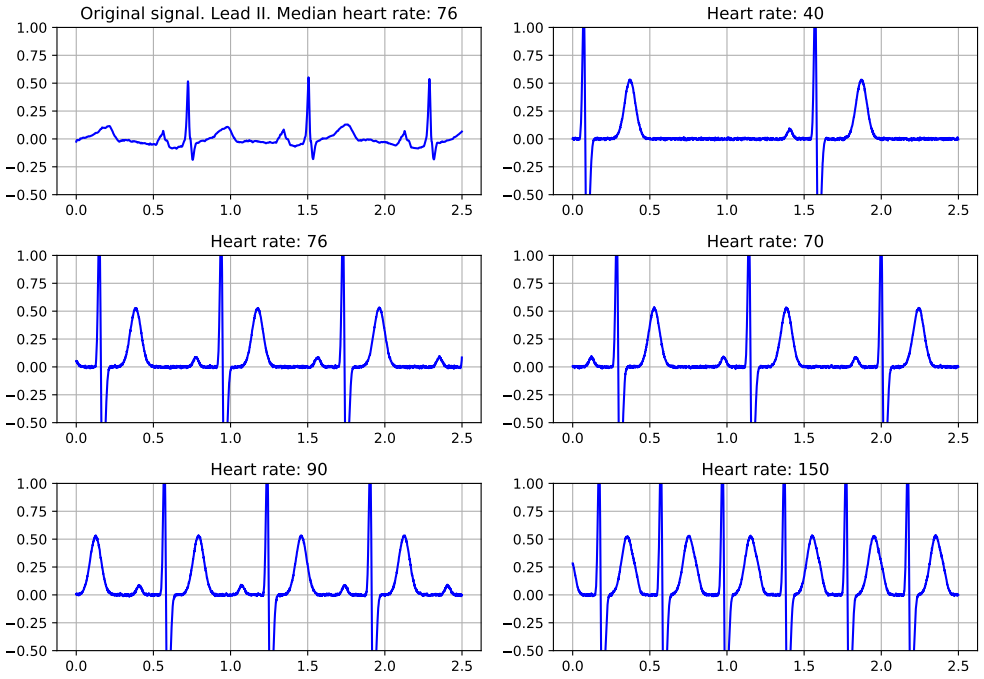

FIGURE 5.16: patient263/s0499_re: Lead V4 rate adjustment.

5.5.5 Experiment 15: Myocardial infarction simulation

To demonstrate how signs of myocardial infarction can be added to an ECG, an example record was chosen. Record patient184/s0363lre was selected as a suitable example as it was modeled well by the sum of gaussians. The representative beat of the record was found, and the model was fit to the representative beat.

To generate the ECG, the QRS annotations of the original ECG was used as R peak positions and the MGF model was evaluated for each annotation. ST elevation was added to lead I and aVL. ST depression was added to lead III. For the precordial leads ST elevation was added to lead V1-V5. The T wave amplitude and length for the precordial leads was scaled with the parameters in table 5.7.

	ST Elevation	T Wave amplitude scale	T Wave length scale
I	0.1	1.0	1.0
II	0.0	1.0	1.0
III	-0.1	1.0	1.0
aVR	0.0	1.0	1.0
aVL	0.1	1.0	1.0
aVF	0.0	1.0	1.0
V1	0.2	-23.4	2.0
V2	0.4	-37.7	2.0
V3	0.6	-46.3	2.5
V4	0.4	34.9	2.2
V5	0.1	22.5	1.7
V6	0.0	11.3	1.7

TABLE 5.7: Parameter scales used for the myocardial infarction simulation example on patient184/s0363lre

The parameter modifications are loosely based on the ECG changes caused by anterior STEMI infarction as described by numerous examples in [50].

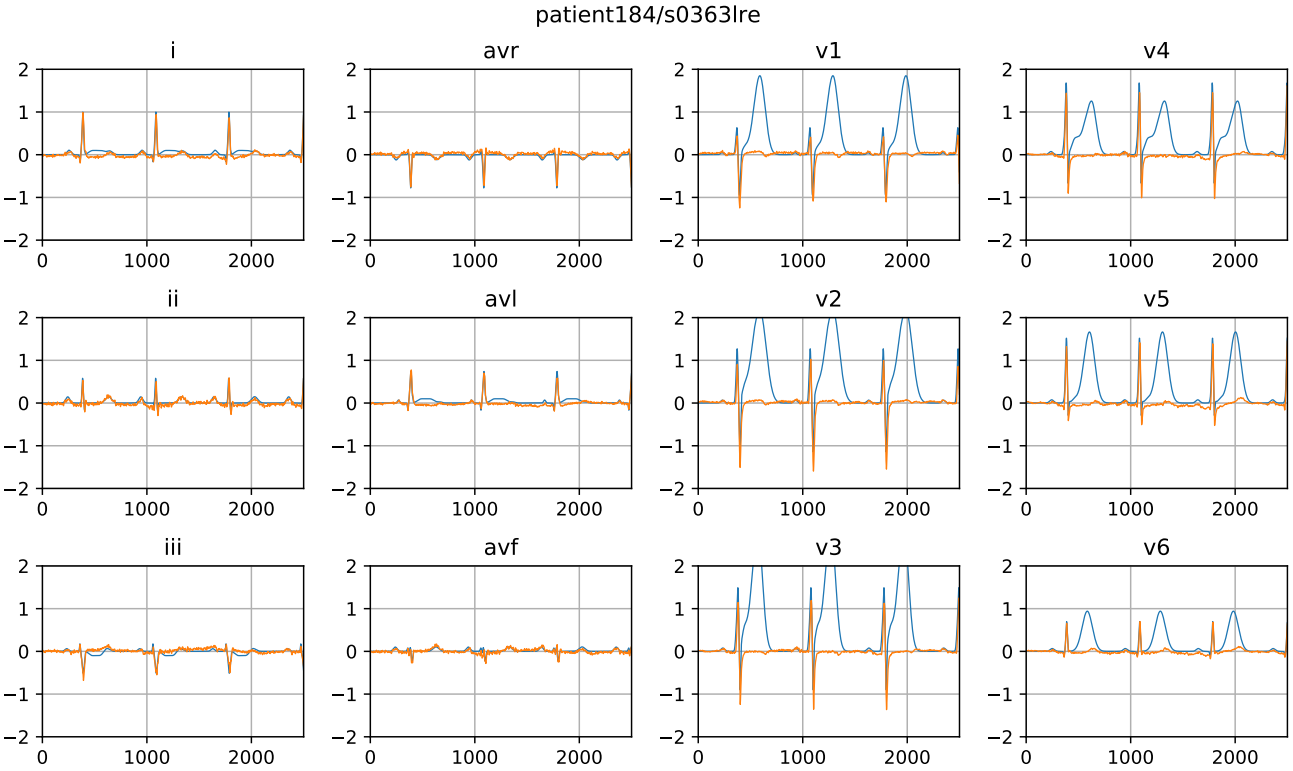

FIGURE 5.17: patient184/s0363lre: Simulated beat myocardial infarction (blue) compared to original record (orange). The simulation shows peaked T waves in the precordial leads and ST elevation in the precordial leads added to the precordial leads, as well as slight ST elevation in lead I, aVL as well as ST depression in lead III.

Figure 5.17 shows an example of myocardial infarction simulation. The original ECG is plotted in orange and the simulated ECG with added signs of myocardial infarction are plotted in blue. The plot shows visible ST elevation in the leads I, aVL, V1, V2, V3, V4 and V5, and ST depression in lead III. The peaked T waves are prominent in all precordial leads. Except for Q wave changes, the simulated ECG is similar to example 1 in [50] which was interpreted as a Hyperacute Anteroseptal STEMI.

6. Discussion

6.1 Analysis of the results

The results in chapter 5.1 show that the voltage potentials required for simulation can be calculated from the recorded electrocardiogram and played back with very little loss of information. With the exception of the recordings with very high frequency content it is assumed that the analog output stage will cause larger errors than that introduced by the calculations. This suggests that any electrocardiogram recording with two or more synchronized limb leads can be converted to the voltage potentials required to play back the record to a patient monitor or other ECG recorder. The filtered records show a lower reconstruction error than the raw data. The reason for this might be that the reconstruction error often occurs due to high frequency elements of the signal which is dampened by the low pass filter.

While the QRS detection did not always trigger on the exact same location in each beat, false negatives and false positives were rare for most records. The ECG wave delineation was less reliable, and had many false positives. By filtering away annotations based on some criteria, most of these were removed and it was possible to arrive to a representative signal for 546 out of 549 records. The sum of modified gaussians function was fit to this representative beat, using the set of annotations to segment the beat into waves. This fitting succeeded on 439 out of the 546 records with valid beats. The curve fitting was succeeded on 56 out of the 80 healthy control records.

The simulated ECG looks quite realistic for many records, and seem to capture the basic shape of the ECG quite well. Some of the records seem to have too complicated QRS complexes to be represented by a sum of three modified gaussians. It has been shown that it is possible to change the rate of the simulated ECG and adjust the QT interval accordingly. This looks realistic for both normal heart rates as well as bradycardia and tachycardia. With tachycardia the P wave and T wave might occur at the same time, resulting in the p waves becoming invisible or being added to the T wave. This effect is also seen in the simulated ECG.

It is also demonstrated how pathological signs can be added to the ECG by changing the parameters of the normal ECG. An example of this is myocardial ischemia or infarction, which can be added to the model as ST elevation, pathological Q waves, inverted or peaked T waves and other signs which might have diagnostic importance. This is meant as a proof of concept and require some more work to seem realistic, but it would be interesting to see how a patient monitor would analyze it.

6.2 Problems with the method

A problem with the method used is that the automatic annotations found by ecgpuwave was quite imprecise. The criteria to filter the annotations caused the beats without P waves to be ignored. In some of the records there were no visible P waves, however based on the criteria the false detections of the P waves would be sorted out and kept for further processing. In some records it was also observed that the annotations for a wave was too short, resulting in a wave of smaller amplitude in the model. The QTc interval also appears to be long for most records, suggesting that the annotations overestimate the length of the QT interval.

There were also some issues with the clustering algorithm used to select aligned beats. The error function typically had two well defined peaks, which was why k-means with a k of 2 was used. However in cases where all beats were aligned, this resulted in a single peak, thus unnecessarily removing many beats. In some cases, for example in records with misaligned beats and ectopic beats, the function might have three or more well defined clusters, sometimes resulting in the algorithm keeping some of the misaligned beats. Since the algorithm then uses the median of the kept beats, it is relatively robust to these errors, however some strange results was observed.

Despite the algorithm producing realistic looking ECG, there are some clear limitations to the model. The model is unable to represent asymmetric waves. This appears to be common in the T waves of the precordial leads. It is also unable to represent ST segment changes and U waves which are common features of the ECG, however in healthy patients the ST segment is normally isoelectric, and the U waves are not visible. Another issue is lead V1 where the P wave normally is biphasic. The model does not account for biphasic waves, so normally only half of the P wave is captured here.

6.3 Alternative solutions

A suggested alternative approach might be to use the representative beats directly, and based on the annotations modify the beats for rate and signs. Another approach might be to use a very large database of electrocardiograms where some features are extracted and search for a relevant ECG based on the features.

6.4 Further work

The first thing I will suggest as further work is to design and assemble the interface to a patient monitor. Using this it could be possible to check how the generated electrocardiograms are automatically interpreted by the patient monitor. Based on the feedback from the automatic interpretation the signs to add to the sinus rhythm could be developed. It would also be interesting to add things like ectopic beats and different rhythms, and to see if the patient monitor interpreted these correctly.

I would also suggest trying a more flexible model. For example adding the possibility to have asymmetric and biphasic waves and adding the ST-segment to the model. The drawback of increasing the model complexity might be increased difficulty to fit the parameters. The model might be easier to fit to the beats if the annotations are improved, or a manually annotated database is used instead. Alternatively a library of parameters for different beat types could be constructed.

In the method proposed here the misaligned beats are simply rejected. By using some kind of cross correlation algorithm it might be possible to align the beats and only reject the beats which are significantly different in shape, such as ectopic beats.

6.5 Conclusion

It has been shown that it is possible to calculate the voltage potentials required for simulation from ECG recordings. It has also been shown that the ECG leads can be reconstructed from these voltage potentials with very low errors on most records. The method presented also demonstrates how to extract a representative beat from an ECG recording and fit a simple parametric model to it. This model can be used to simulate the ECG at different heart rates as well as with extra signs of pathological conditions such as ST-elevation.

A. Experiments

This appendix explains several experiments performed on the PTB Diagnostic ECG database. The first chapter describes the determination of the filter type and parameters. The second chapter is a verification of the QRS detector. The third chapter explains the selection of the method for corrected QT interval calculation by the use of linear regression. Then a similar approach is used on the PR interval. Finally the notes of a manual review of the results from the representative beat calculation described in 5.3.2 are presented.

A.1 Determination of filter type and parameters

The selection of a suitable low pass and high pass filter was based both on the recommendations found in "Recommendations for the Standardization and Interpretation of the Electrocardiogram"[25], and by experimentation. In the recommendations it is stated that a cutoff frequency of 0.05 Hz should be used for conventional high pass filters, but that this could be relaxed to 0.67 Hz with the use of digital filters without phase distortion. This technique is explained in ch. 2.8. The low pass filter cutoff frequency recommended was 150 Hz. This filter was implemented as a second order Butterworth filter with a cutoff frequency of 150Hz with zero phase filtering. This appeared to reduce muscle artifacts without distorting the high frequency content of the QRS complexes significantly.

The high pass filter was determined by observing the effects of the filtering on individual beats in several records from the PTB Diagnostic ECG database. The record was loaded and filtered with a Butterworth filter with normal and zero phase implementation, as well as for the cutoff frequencies 0.05, 0.25, 0.5, 0.67, 1, 2, 5 and 10 Hz. Examples of this can be seen in figure A.1 and figure A.2. The conventional filtering show significantly higher distortion, especially in the ST segment. With the zero phase filtering a cutoff frequency of 0.67 Hz was found suitable. This is a somewhat stronger filtering compared to the recommendations in [25], due to the squared amplitude response.

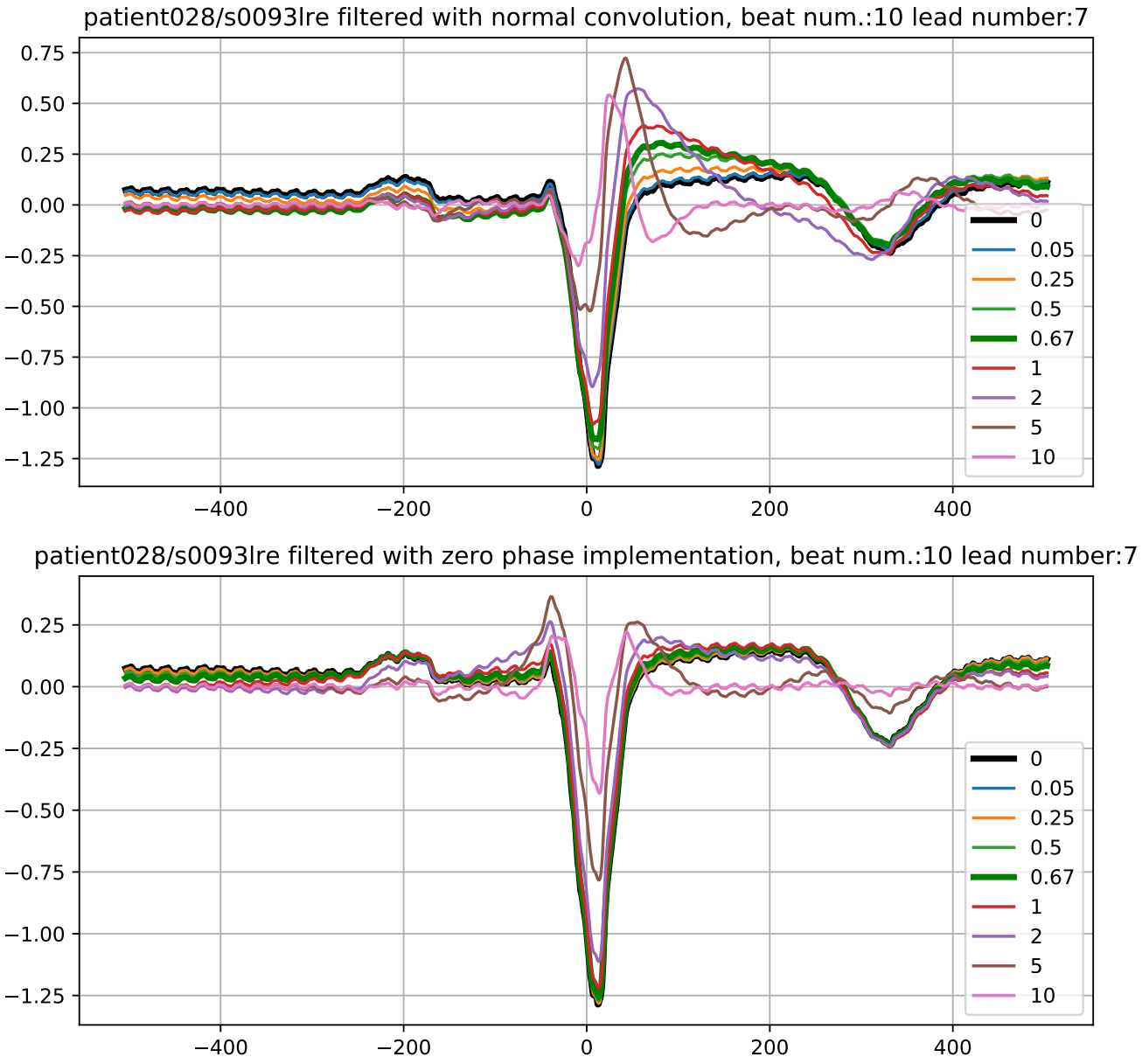

FIGURE A.1: Lead V2 filtered with a second order Butterworth filter different cutoff frequencies and method. With the zero phase implementation the equivalent amplitude response is squared.

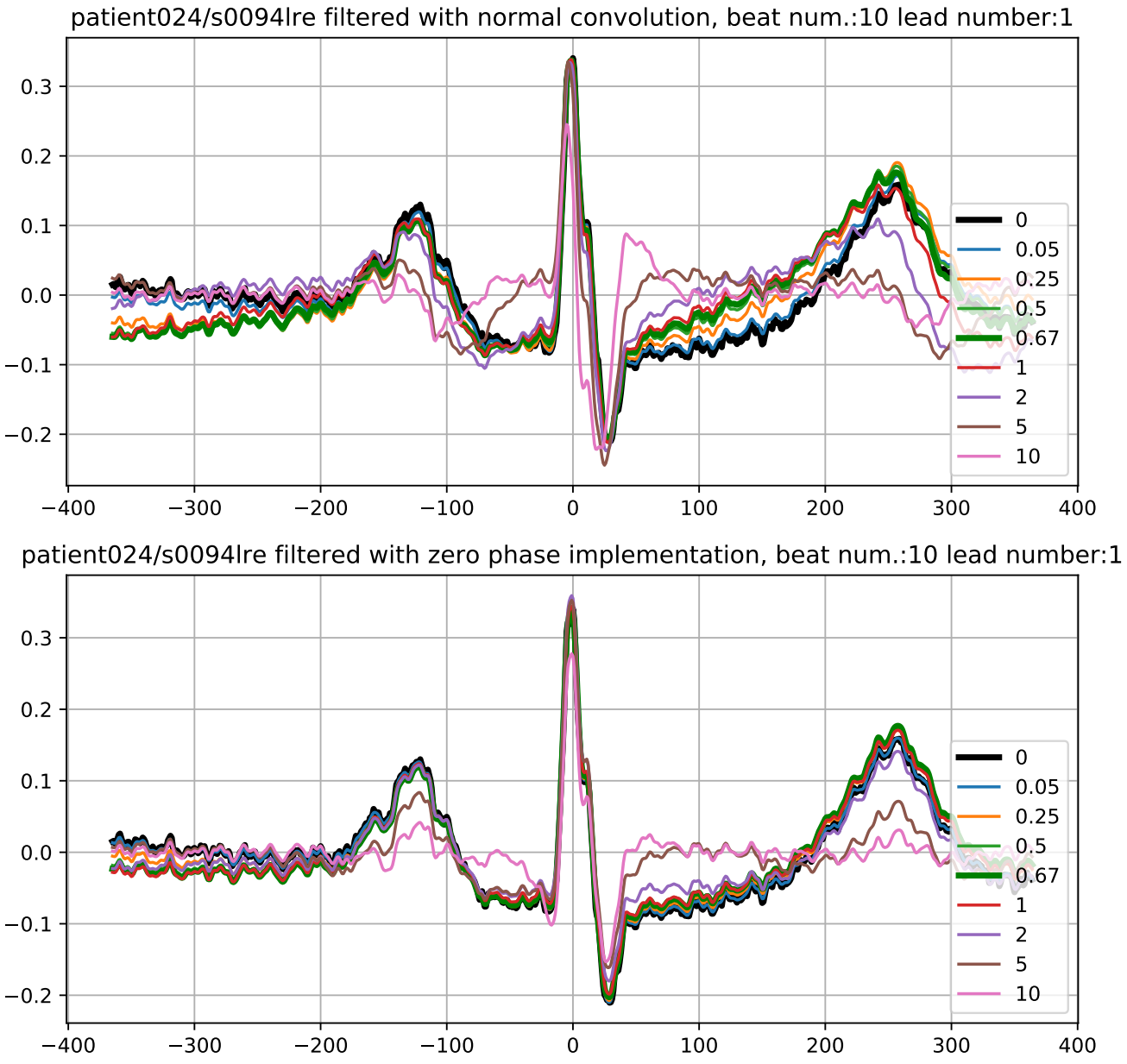

FIGURE A.2: Lead II filtered with a second order Butterworth filter different cutoff frequencies and method. With the zero phase implementation the equivalent amplitude response is squared.

A.2 Verification of the QRS annotations

This project required a reliable QRS detector. The algorithm in [26] was found to perform well in the benchmarks listed in the paper. Initial testing on the records of the PTB Diagnostic ECG database also showed promising results. This algorithm was to be tested against the built in QRS detector in ecgpuwave[29], based on the Pan Tompkins algorithm [51]. The standard parameters was used for the QRS detector (window size 5.0, low frequency cutoff 5.0Hz, high frequency cutoff 15.0Hz) for all records except "patient123/s0224_re". Here the window size had to be reduced to 2.0 because of large spikes in the signal.

The error measure used is based on the idea that a more consistent QRS detector will result in lower beat to beat variance in the detected beat lengths. A common way to study the beat to beat variance in electrocardiograms are called Poincaré plots. A Poincaré plot is a projection of a time series to a two dimensional space[52]. If we have the beat lengths denoted L(k), we can find the Poincaré plot by plotting L(k) against L(k+1). In healthy patients the dots will be close to the diagonal line, suggesting the heart rate varies slowly with respiration. If the patient has an unstable rhythm, ectopic beats or missed QRS detections, the points will deviate from the diagonal.

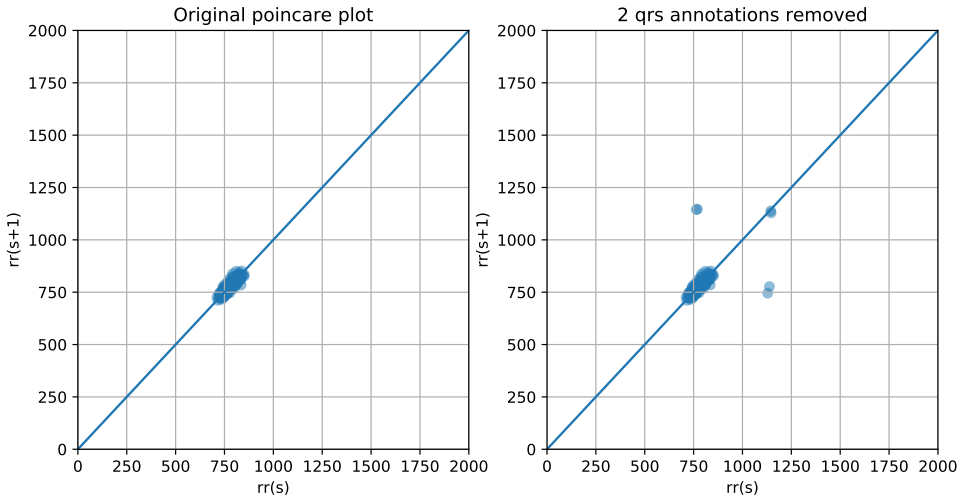

FIGURE A.3: Left: Poincaré plot of healthy control, Right: Poincaré plot of healthy control with one qrs annotation removed.

In figure A.3 two examples of Poincaré plots are shown. In the first plot the Poincaré plot of an example record is shown. The dots are close to the diagonal line plotted in blue. In the right plot, the same plot was made for the same record with two of the QRS annotations removed. The plot shows a similar group of dots, however the two removed

	Pan Tompkins	Kathirvel et. al
count	549	549
mean	19.154545	15.840260
std	85.164704	23.529007
min	0.534348	0.366648
25%	4.462629	3.854872
50%	8.034501	7.570757
75%	15.331871	15.824562
max	1945.038623	186.917073

TABLE A.1: RMS deviation of the Poincaré plot for each QRS detection method

QRS annotations have caused 4 dots to move up along the diagonal. It also caused two dots to move above the diagonal and two dots to move below the diagonal line.

A measure for this deviation was found by rotating the plot -45 degrees by multiplying each value pair with the rotation matrix. With L(k) being the beat lengths, and $p_{rot,0}$ and $p_{rot,1}$ as a the x and y coordinates of the rotated plot, we have:

$$\begin{bmatrix} p_{rot,0} \\ p_{rot,1} \end{bmatrix} = \begin{bmatrix} 0.70710678 & -0.70710678 \\ 0.70710678 & 0.70710678 \end{bmatrix} \begin{bmatrix} L(k) & L(k+1) & ... & L(N-1) \\ L(k+1) & L(k+2) & ... & L(N) \end{bmatrix} \tag{A.1}$$

The error measure is found by taking the root mean square of $P_{rot,1}$ (the vertical axis of the rotated poincare):

$$RMS_{dev} = \frac{1}{N-1} \sum_{k=0}^{N-1} \sqrt{|P_{rot,1}|^2} \tag{A.2}$$

The QRS detection from [26] was run on all records and used as an input to ecgpuwave. This was to be compared to the results without a input QRS annotation (built in QRS detector in ecgpuwave). The length of each beat $L_{beat}(i)$ was estimated by the second order difference as explained in chapter 4.1.1. The RMS_{dev} was then calculated as above for both algorithms. Finally the mean, standard deviation, median, minimum, maximum and 25th, 50th and 75th percentiles was calculated for RMS_{dev} of both the modified Pan Tompkins algorithm and the Kathirvel algorithm.

Table A.1 shows the RMS deviation of the Poincaré plot for each QRS detection method. Based on this error criterion the Kathirvel et al. method performed better than the built in QRS detector, and is therefore selected as the QRS detection method for this project.

A.3 Selection of method for corrected QT interval calculation

There are several methods for computing the QTc interval. To determine which method to use, the detected QT intervals in the database was plotted against the detected heart rate for all records and for the healthy controls for each method. The heart rate was found based on the median beat length of each record as explained in chapter 4.1.2. The QT intervals was found using the median annotations described in chapter 4.2.1. The QT interval was found as the onset annotation of the QRS complex minus the end annotation of the T wave. These plots can be seen in figure A.4 and A.5. The first plot shows the QT interval versus the heart rate. The trend here is that the QT interval shortens with increasing heart rate, which corresponds with the theory.

It is assumed that the ideal QTc calculation will minimize the linear relationship between the heart rate and the QTc interval. Therefore a linear least-squares regression was performed between the heart rate and the QT intervals, as well as between the heart rate and the QTc intervals for each method. The results of the linear regression is presented in table A.2 and A.3. The linear relationship is also plotted as the red line in figure A.4 and A.5. All the methods show improvement over no correction at all.

Despite being a commonly used formula for calculating QTc, Bazett's formula had the worst performance on all records with a slope of 0.632. On the healthy controls it performed better, with a slope of 0.397. This corresponds to correcting the intervals too much. The best performing method for all records was Hodges method which had a slope of -0.316. The best performing method on the healthy controls was Bazett's formula with a slope of 0.397. The slopes for Fredericia and Hodges was quite similar in all records. Since it was also suggested to replace Bazett's formula in a recent retrospective study[53], Fredericias method is therefore selected to calculate the QTc intervals in this project.

	intercept	p value	r value	slope	std err
QT interval	554.816	0.000	-0.598	-2.016	0.116
QTc Bazett	398.289	0.000	0.210	0.632	0.126
QTc Fredericia	454.457	0.010	-0.111	-0.316	0.122
QTc Framingham	458.648	0.001	-0.143	-0.388	0.115
QTc Hodges	449.816	0.022	-0.098	-0.266	0.116

TABLE A.2: Linear regression parameters between heart rate and QT interval as well as QTc calculated with different methods for all records

	intercept	p value	r value	slope	std err
QT interval	560.619	0.000	-0.734	-2.442	0.256
QTc Bazett	389.199	0.131	0.170	0.397	0.260
QTc Fredericia	448.676	0.024	-0.252	-0.590	0.257
QTc Framingham	441.674	0.057	-0.214	-0.481	0.249
QTc Hodges	455.619	0.008	-0.293	-0.692	0.256

TABLE A.3: Linear regression parameters between heart rate and QT interval as well as QTc calculated with different methods for healthy controls

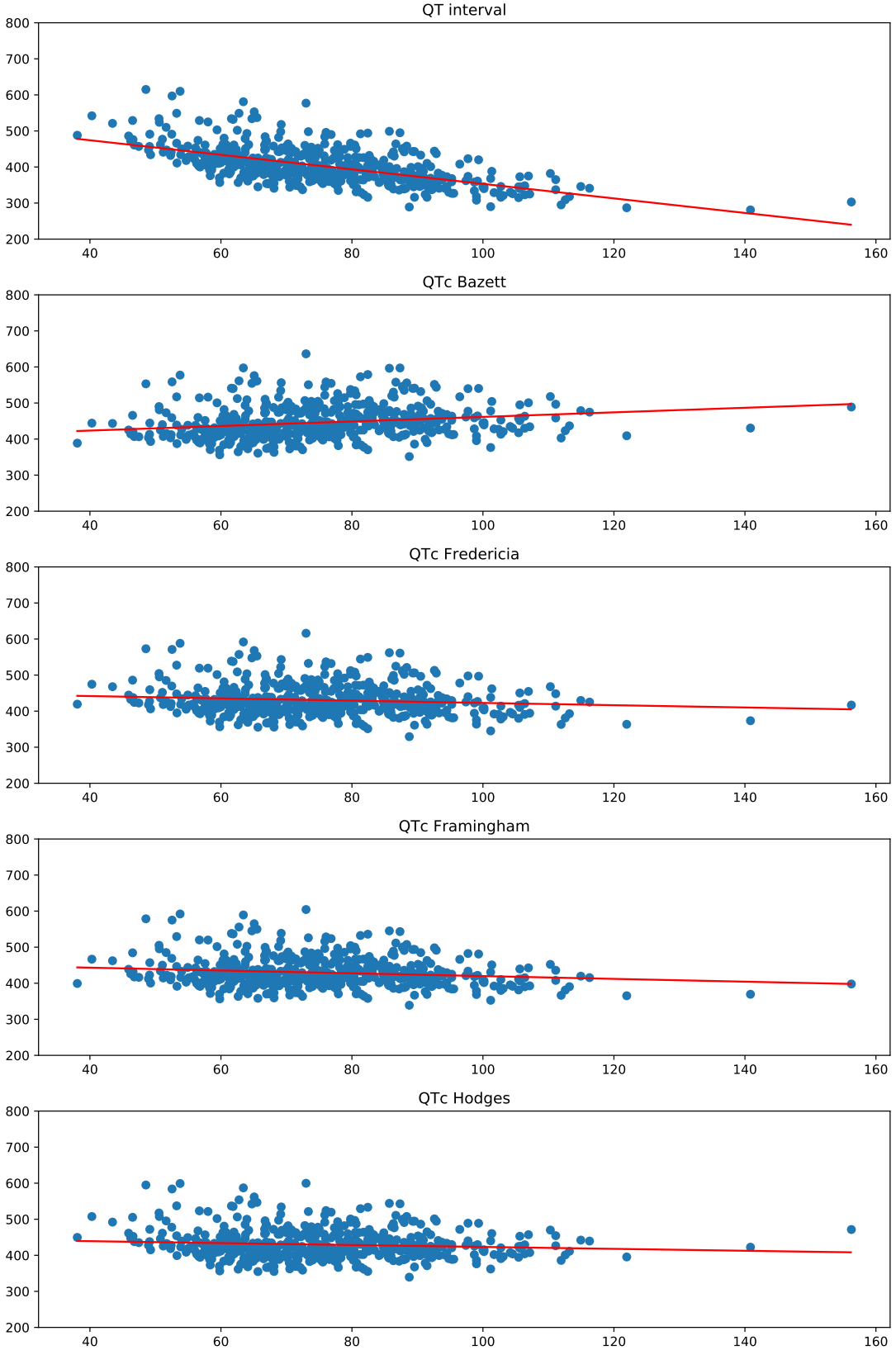

FIGURE A.4: Comparison of the calculation of QTc using different methods for all records

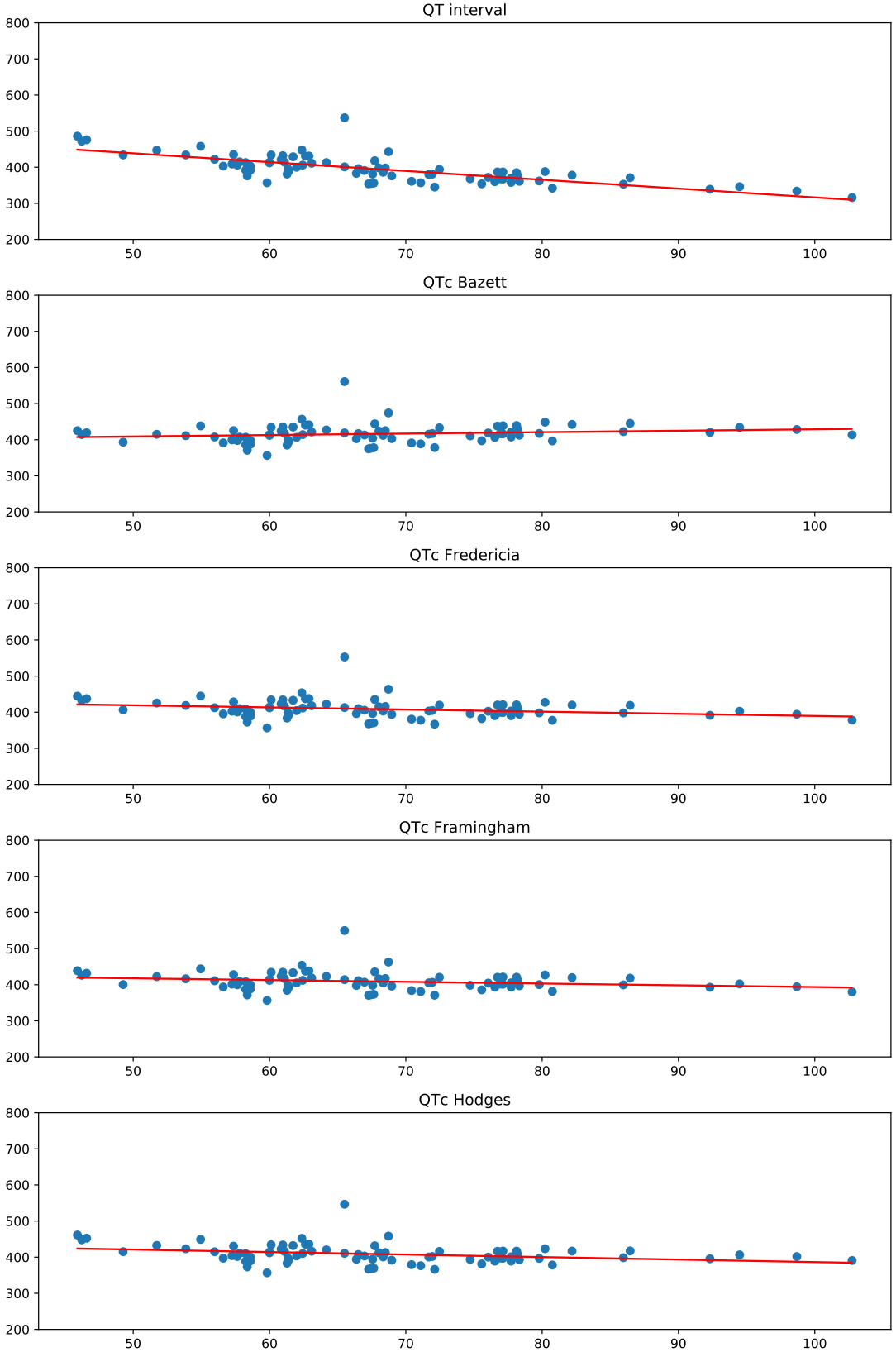

FIGURE A.5: Comparison of the calculation of QTc using different methods for healthy controls

A.4 PR interval

To study if there was a linear relationship between the PR interval and the heart rate, the PR intervals was found using the median annotations described in chapter 4.2.1. The PR interval was found as the onset annotation of the QRS complex minus the onset annotation of the P wave. The PR interval length was plotted against the heart rate and a linear regression was performed for all records in the PTB Diagnostic ECG Database (figure A.6). The linear regression showed a very weak linear relationship between the heart rate and the PR interval. Therefore it was decided not to correct for PR interval during simulation.

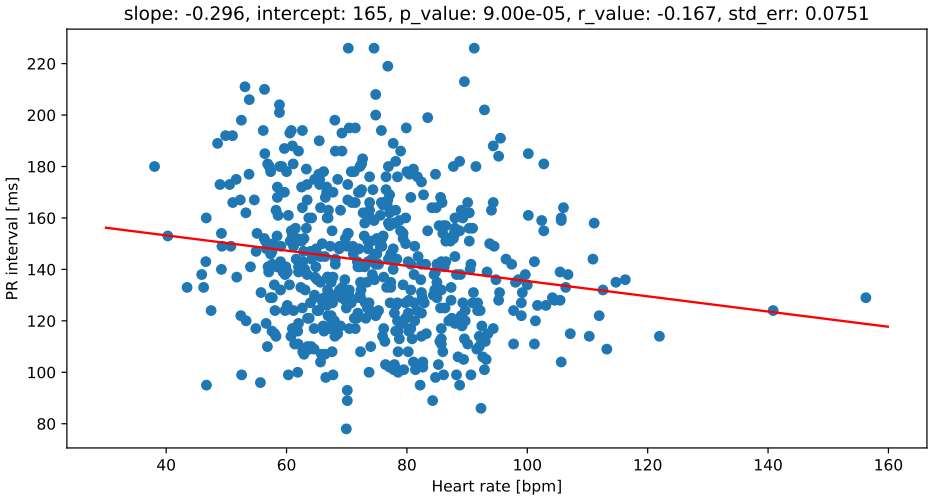

FIGURE A.6: Linear regression between hr and PR interval

A.5 Representative beat

To evaluate the performance of the representative beat formation explained in 4.2, the plots described in chapter 5.3.2 was reviewed for the whole database. The plots were manually inspected and the criteria for a failed record is the following:

- The algorithm kept beats that were clearly misaligned
- The algorithm kept beats with different morphology
- The peak of the R wave of the accepted beats are not located near 0

The following errors was noted in the manual inspection (zero indexed record number in parenthesis):

patient010/s0036lre (23): S peak detected as R peak
patient010/s0042lre (24): S peak detected as R peak
patient012/s0050lre (31): S peak detected as R peak
patient021/s0065lre (57): Algorithm kept misaligned beats, result OK
patient025/s0091lre (72): Algorithm kept misaligned beats, result OK
patient031/s0104lre (90): Algorithm kept misaligned beats, Q peak detected as R peak, distorted representative beat
patient038/s0128lre (115): S peak detected as R peak
patient053/s0128lre (167): S peak detected as R peak
patient060/s0209lre (177): Q peak detected as R peak
patient078/s0262lre (225): S peak detected as R peak
patient078/s0317lre (226): S peak detected as R peak
patient082/s0271lre (240): Q peak detected as R peak
patient088/s0339lre (259): S peak detected as R peak
patient091/s0361lre (273): Algorithm kept misaligned beat, result OK
patient091/s0408lre (274): Algorithm kept misaligned beat, result OK
patient097/s0382lre (297): Algorithm kept different beat, 3 peaks in error, result OK
patient099/s0388lre (305): S peak detected as R peak
patient099/s0397lre (306): S peak detected as R peak
patient102/s0416lre (314): Algorithm kept different beat, 3 peaks in error, result OK
patient156/s0299lre (369): S peak detected as R peak
patient174/s0324lre (392): S peak detected as R peak
patient174/s0325lre (393): S peak detected as R peak
patient181/s0416lre (406): Algorithm kept different beat, 3 peaks in error, result OK
patient183/s0175_re (409): T peak detected as R peak
patient185/s0336lre (411): S peak detected as R peak
patient198/s0415lre (427): Algorithm kept misaligned beats distorted representative beat
patient199/s0404lre (428): S peak detected as R peak
patient202/s0422_re (433): S peak detected as R peak
patient206/s0427_re (437): S peak detected as R peak
patient208/s0429_re (439): Algorithm kept misaligned beat, result OK
patient214/s0436_re (446): S peak detected as R peak
patient233/s0457_re (467): S peak detected as R peak
patient233/s0483_re (471): S peak detected as R peak
patient242/s0471_re (483): S peak detected as R peak
patient278/s0528_re (522): S peak detected as R peak
patient278/s0529_re (523): S peak detected as R peak
patient278/s0530_re (524): S peak detected as R peak

patient282/s0539_re (531): S peak detected as R peak
patient284/s0551_re (534): S peak detected as R peak
patient285/s0544_re (536): Algorithm failed, result is flat
patient287/s0547_re (538): Algorithm kept misaligned beat, result OK
patient290/s0553_re (542): Algorithm kept misaligned beat, result OK
patient292/s0555_re (544): Algorithm kept misaligned beats, unstable rhythm?, distorted representative beat
patient293/s0557_re (546): Algorithm kept misaligned beat, result OK

The errors are summed up in table A.4.

Error type	Count
S peak detected as R peak	26
Q peak detected as R peak	2
Algorithm kept misaligned beat, result OK	8
Algorithm kept different beat, 3 peaks in error, result OK	3
Visible distortion of the representative beat	3
Total	39

TABLE A.4: Number of errors found during manual inspection of the representative beats

The three records with visible distortion of the representative beats were:

- patient031/s0104lre (90): Algorithm kept misaligned beats, Q peak detected as R peak, distorted representative beat
- patient198/s0415lre (427): Algorithm kept misaligned beats distorted representative beat
- patient292/s0555_re (544): Algorithm kept misaligned beats, unstable rhythm?, distorted representative beat

B. Software installation and overview

The first two sections explains how to install the software requirements for running the scripts used in this project for Windows and Linux respectively. The last chapter contains a brief explanation of the most important files and which figures or tables the file is used to generate.

B.1 Windows install using Anaconda

The following steps were used to install the required software on Windows 10. "$" means run command in command line.

1. install wget `https://www.gnu.org/software/wget/`
2. add wget to path
3. install make `https://www.gnu.org/software/make/`
4. add make to path
5. download and install anaconda (python 3.6 version) `https://www.continuum.io/downloads`
6. pip install wfdb==0.1.2
7. install wfdb software pack following instructions at `https://www.physionet.org/physiotools/wfdb.shtml`. Use cygwin32. When installing cygwin32 install gfortran compiler as well as all the packages mentioned in the linked website
8. Assuming the install directory of cygwin is `C:\cygwin\` add the following folders to path: `C:\cygwin\bin`, `C:\cygwin\home\fedda\wfdb-10.5.24\app` and `C:\cygwin\home\fedda\wfdb-10.5.24\lib`

To install ecgpuwave:

1. download `https://physionet.org/physiotools/ecgpuwave/src/ecgpuwave-1.3.3.tar.gz`
2. extract ecgpuwave-1.3.3.tar.gz in a suitable location
3. create a folder for ecgpuwave and copy contents of ecgpuwave-1.3.3 there
4. open in atom or similar and do find and replace ”100000” with ”400000” for all files in the folder (ctrl+shift+f in atom) (209 times in 10 files)
5. navigate to the ecgpuwave folder
6. $ make install
7. $ make check
8. If tests 1 and 2 are passed the install was successful

open in atom or similar and do find and replace ”100000” with ”400000” for all files in the folder enter this folder from the command prompt (might need root) (on windows run in cygwin) run ”make” run ”make install” run ”make check” - if tests 1 and 2 are passed the install was succesfull

B.2 Linux install using Miniconda

The installation of the software was tested in Ubuntu 16.04.2 desktop i386 running on Oracle VM VirtualBox.

First install Miniconda and run the commands after the $ in the terminal:

1. Install miniconda python 3.6 version from `https://conda.io/miniconda.html`
2. $ conda install numpy
3. $ conda install scipy
4. $ pip install wfdb==0.1.2
5. $ conda install scikit-learn
6. $ conda install pandas

For Anaconda the packages except

Install the WFDB Software following the instructions in `https://physionet.org/physiotools/wfdb-linux-quick-start.shtml`
If missing packages, try:
$ sudo apt-get install gcc libcurl4-openssl-dev libexpat1-dev

ecgpuwave can be installed by following the instructions in `https://physionet.org/physiotools/ecgpuwave/src/INSTALL`, however some modifications must be done to the code before compiling to allow for signals longer than 100000 samples. The following explains the required steps:

1. $ sudo apt-get install gfortran
2. Download: `https://physionet.org/physiotools/ecgpuwave/src/ecgpuwave-1.3.3.tar.gz`
3. Extract to suitable location (using tar: $ tar -xf ecgpuwave-1.3.3.tar.gz)
4. Navigate to the ecgpuwave folder ($ cd ecgpuwave-1.3.3)
5. Replace the number 100000 with 400000 in all files: $ find ./ -type f -exec sed -i -e 's/100000/400000/g' ;
6. $ sudo make install
7. $ sudo make check

B.3 Overview of the software

This section explains which figures and tables each script generates as well as a description of some of the scripts.

The scripts use ecg3.py for loading records, segmentation and extraction of representative beats. gauss_fitting_script.py is used for fitting the modified gaussian functions to the representative beat. player.py is used to generate ECG from the models and ecg_tools.py is a set of common functions used in many other files.

B.3.1 00-qrs_detection.py

Runs QRS detection on all records on the database. The detections are written to a temporary text file and wrann from the WFDB package is used to write the QRS annotations to one PhysioNet annotation file per record.

B.3.2 01-ecgpuwave.py

This script runs ecgpuwave on all records in ptbdb using the annotations found in the previous script. To use the built in QRS annotations, the line "ptbdb.ecgpuwave()" can be changed to "ptbdb.ecgpuwave(input_ann=False)". This script usually takes around 1-2 hours to finish.

B.3.3 02-plot_example_record.py

This script plots the first 4 seconds of an example record and marks the annotations found by ecgpuwave as vertical blue, red and green lines for the QRS annotation, P annotation and T annotation respectively.

B.3.4 03-calculate_rms_error.py

This script calculates the RMS Reconstruction error of the records in the list "record_numbers".

B.3.5 04-calculate_rms_error_all_records.py

This script calculates the RMS reconstruction error for all records, both with filtering and without filtering. It also performs all the steps until finding the representative beat, and stores the median beat annotations, number of QRS complexes detected, number of N annotations from ecgpuwave, number of valid beats, reason of admission, median beat length, heart rate, number of accepted and rejected beats by the clustering and the cluster centroids. For each record this is added to a pandas dataframe, and when all records are processed the dataframe is saved to `pickle/node_voltage_table.p` using pickle.

B.3.6 05-table_reconstruction_error.py

Generates table 5.1 and 5.2.

B.3.7 06-segmentation_table.py

Generates table 5.3.

B.3.8 07-segmentation_plot_best_and_worst.py

Plots examples of the records with lowest and highest percentage valid beats. Generates figure 5.5 and 5.6.

B.3.9 08-representative_beat_table.py

Generates table 5.4.

B.3.10 09-qt_correction_regression.py

Generates figure A.4, A.5 and A.6, and table A.2 and A.3.

B.3.11 10-node_voltage_figure.py

Generates figure 5.1, 5.2 and 5.3.

B.3.12 11-representative_beat_figure.py

Generates figure 5.7 and 5.8.

B.3.13 12-fit_record.py

Finds the representative beat of an example record and fits a modified gaussian model to it.

B.3.14 13-fit_all_records.py

Fits modified gaussian model to all the records of the database with valid beats and calculates simulation errors. The data is stored in `pickle/sq_sim_err.p`, `pickle/fit_all_leads.p` and `pickle/fit_all_leads_arrays.p` and used by the following scripts.

B.3.15 14-sim_tables_and_err_figs.py

Generates table 5.5 and 5.6. Generates figure 5.9 and 5.10

B.3.16 15-sim_figures.py

Generates figure 5.11, 5.12,5.13 and 5.14

B.3.17 16-sim_rate_change.py

Generates figure 5.15 and 5.16.

B.3.18 17-sim_stemi.py

Generates figure 5.17.

Bibliography

[1] Life in the fastlane. Ecg library - qt interval. https://lifeinthefastlane.com/ecg-library/basics/qt_interval/, 2017. [Online; accessed 2-June-2017].

[2] Fatimah Lateef et al. Simulation-based learning: Just like the real thing. *Journal of emergencies, trauma, and shock*, 3(4):348, 2010.

[3] Helge Fossan. Ecg simulation, 2016.

[4] Patrick E McSharry, Gari D Clifford, Lionel Tarassenko, and Leonard A Smith. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 50(3):289–294, 2003.

[5] Fluke Biomedical. Prosim 8 vital signs patient simulator. http://www.flukebiomedical.com/biomedical/usen/Patient-Simulators/ProSim-8-vital-signs-patient-simulator.htm?PID=72624, 2017. [Online; accessed 30-May-2017].

[6] Gaumard. Hal®s1020 emergency care simulator. http://www.gaumard.com/s1020, 2017. [Online; accessed 30-May-2017].

[7] Wikipedia. Atrium (heart) — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Atrium_(heart)&oldid=776740693, 2017. [Online; accessed 8-May-2017].

[8] J.E. Hall and A.C. Guyton. *Guyton and Hall Textbook of Medical Physiology*. ClinicalKey 2012. Saunders/Elsevier, 2011. ISBN 9781416045748. URL https://books.google.no/books?id=di5PtQAACAAJ.

[9] Dictionary.com unabridged - cardiac cycle. http://www.dictionary.com/browse/cardiac-cycle, May 2017.

[10] Wikipedia. Cardiac action potential — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Cardiac_action_potential&oldid=778437057, 2017. [Online; accessed 8-May-2017].

[11] Wikipedia. Cardiac pacemaker — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Cardiac_pacemaker&oldid=771801823, 2017. [Online; accessed 9-May-2017].

[12] Arnold M Katz. *Physiology of the Heart*. Lippincott Williams & Wilkins, 2010.

[13] Life in the fastlane. Ecg library - lead positioning. https://lifeinthefastlane.com/ecg-library/basics/lead-positioning/, 2017. [Online; accessed 9-May-2017].

[14] Life in the fastlane. Ecg library - p-wave. https://lifeinthefastlane.com/ecg-library/basics/p-wave/, 2017. [Online; accessed 28-May-2017].

[15] Wikipedia. Electrocardiography — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Electrocardiography&oldid=779412049, 2017. [Online; accessed 21-May-2017].

[16] Life in the fastlane. Ecg library - abnormalities of the r wave. https://lifeinthefastlane.com/ecg-library/basics/r-wave/, 2017. [Online; accessed 28-May-2017].

[17] Life in the fastlane. Ecg library - q wave. https://lifeinthefastlane.com/ecg-library/basics/q-wave/, 2017. [Online; accessed 28-May-2017].

[18] Life in the fastlane. Ecg library - t wave. https://lifeinthefastlane.com/ecg-library/basics/t-wave/, 2017. [Online; accessed 28-May-2017].

[19] Life in the fastlane Gerard Fennesy. Ecg "rule of fours". https://lifeinthefastlane.com/ecg-rule-of-fours/, 2017. [Online; accessed 2-June-2017].

[20] ACLS Medical Training. Guide to understanding ecg artifact. https://www.aclsmedicaltraining.com/blog/guide-to-understanding-ecg-artifact/, 2017. [Online; accessed 9-June-2017].

[21] Wikipedia. Driven right leg circuit — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Driven_right_leg_circuit&oldid=754249294, 2017. [Online; accessed 21-May-2017].

[22] Wikipedia. Kirchhoff's circuit laws — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Kirchhoff%27s_circuit_laws&oldid=780847805, 2017. [Online; accessed 21-May-2017].

[23] Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health, Seventh Edition.. Einthoven's law. (n.d.). http://medical-dictionary.

thefreedictionary.com/Einthoven%27s+law, 2003. [Online; accessed 21-May-2017].

[24] Jaakko Malmivuo and Robert Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.

[25] Paul Kligfield, Leonard S. Gettes, James J. Bailey, Rory Childers, Barbara J. Deal, E. William Hancock, Gerard van Herpen, Jan A. Kors, Peter Macfarlane, David M. Mirvis, Olle Pahlm, Pentti Rautaharju, and Galen S. Wagner. Recommendations for the standardization and interpretation of the electrocardiogram. *Circulation*, 115(10):1306–1324, 2007. ISSN 0009-7322. doi: 10.1161/CIRCULATIONAHA.106.180200. URL http://circ.ahajournals.org/content/115/10/1306.

[26] P. Kathirvel, M. Sabarimalai Manikandan, S. R. M. Prasanna, and K. P. Soman. An efficient r-peak detection based on new nonlinear transformation and first-order gaussian differentiator. *Cardiovascular Engineering and Technology*, 2(4):408–425, 2011. ISSN 1869-4098. doi: 10.1007/s13239-011-0065-3. URL http://dx.doi.org/10.1007/s13239-011-0065-3.

[27] Pablo Laguna, Raimon Jané, Eudald Bogatell, and David Vigo Anglada . Physionet - qrs detection and waveform boundary recognition using ecgpuwave. https://physionet.org/physiotools/ecgpuwave/, 2017. [Online; accessed 6-June-2017].

[28] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[29] Pablo Laguna, Raimon Jané, and Pere Caminal. Automatic detection of wave boundaries in multilead ecg signals: Validation with the cse database. *Computers and biomedical research*, 27(1):45–60, 1994.

[30] R Jané, A Blasi, J García, and P Laguna. Evaluation of an automatic threshold based detector of waveform limits in holter ecg with the qt database. In *Computers in Cardiology 1997*, pages 295–298. IEEE, 1997.

[31] Wikipedia. Linear phase — wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Linear_phase&oldid=726136814, 2016. [Online; accessed 23-May-2017].

[32] Ivan Selesnick. Four types of linear-phase fir filters. `http://cnx.org/contents/MuhmqM-3@3/Four-Types-of-Linear-Phase-FIR`, 2005. [Online; accessed 23-May-2017].

[33] JG Proakis and DG Manolakis. Digital signal processing, 4th-ed, 2007.

[34] Mathworks. Matlab documentation - filtfilt (zero-phase digital filtering). `https://se.mathworks.com/help/signal/ref/filtfilt.html`. [Online; accessed 24-May-2017].

[35] The Scipy community. Scipy docs - scipy.signal.filtfilt. `https://docs.scipy.org/doc/scipy-0.19.0/reference/generated/scipy.signal.filtfilt.html`. [Online; accessed 24-May-2017].

[36] Stackexchange Signal Processing Matt L. and aralar. What is the advantage of matlab's filtfilt, stackexchange signal processing. `https://dsp.stackexchange.com/questions/9467/what-is-the-advantage-of-matlabs-filtfilt`, 2016. [Online; accessed 30-May-2017].

[37] Julius Orion Smith. *Introduction to digital filters: with audio applications*, volume 2. Julius Smith, 2008.

[38] WolframMathWorld. Gaussian function. `http://mathworld.wolfram.com/GaussianFunction.html`, 2017. [Online; accessed 28-May-2017].

[39] Wikipedia. Non-linear least squares — wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Non-linear_least_squares&oldid=782457642`, 2017. [Online; accessed 12-June-2017].

[40] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[41] Mary Ann Branch, Thomas F Coleman, and Yuying Li. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing*, 21(1):1–23, 1999.

[42] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python. `http://www.scipy.org/`, 2001–. [Online; accessed 12-June-2017].

[43] scikit learn. K-means - scikit-learn user guide. `http://scikit-learn.org/stable/modules/clustering.html#k-means`, 2017. [Online; accessed 28-May-2017].

[44] Life in the fastlane. Ecg library. `https://lifeinthefastlane.com/ecg-library/`, 2017. [Online; accessed 8-June-2017].

[45] R Bousseljot, D Kreiseler, and A Schnabel. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. *Biomedizinische Technik/Biomedical Engineering*, 40(s1):317–318, 1995.

[46] Life in the fastlane. Life in the fastlane. `https://lifeinthefastlane.com/`, 2017. [Online; accessed 8-June-2017].

[47] Chen Xie, Tom Pollard, Alistair Johnson, Kevin Rose. wfdb-python - native python wfdb package. `https://github.com/MIT-LCP/wfdb-python`, 2017. [Online; accessed 8-June-2017].

[48] Wikipedia. Moore–penrose pseudoinverse — wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Moore%E2%80%93Penrose_pseudoinverse&oldid=778808374`, 2017. [Online; accessed 26-May-2017].

[49] Life in the fastlane. Ecg library - sinus tachycardia. `https://lifeinthefastlane.com/ecg-library/sinus-tachycardia/`, 2017. [Online; accessed 9-June-2017].

[50] Life in the fastlane. Ecg library - anterior myocardial infarction. `https://lifeinthefastlane.com/ecg-library/anterior-stemi/`, 2017. [Online; accessed 13-June-2017].

[51] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236, 1985.

[52] Wikipedia. Poincaré plot — wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Poincar%C3%A9_plot&oldid=779492268`, 2017. [Online; accessed 4-June-2017].

[53] Bert Vandenberk, Eline Vandael, Tomas Robyns, Joris Vandenberghe, Christophe Garweg, Veerle Foulon, Joris Ector, and Rik Willems. Which qt correction formulae to use for qt monitoring? *Journal of the American Heart Association*, 5(6), 2016. ISSN 2047-9980. doi: 10.1161/JAHA.116.003264. URL `http://jaha.ahajournals.org/content/5/6/e003264`.