



FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

Peetu Kujala

**EXTRACTING HEART RATE DEPENDENT
ELECTROCARDIOGRAM TEMPLATES FOR A
BODY EMULATOR ENVIRONMENT**

Master's Thesis
Degree Programme in Biomedical Engineering
September 2023

Kujala P. (2023) Extracting Heart Rate Dependent Electrocardiogram Templates for a Body Emulator Environment. University of Oulu, Degree Programme in Biomedical Engineering, 92 p.

ABSTRACT

Medical device and analysis method developments often include tests on humans, which are expensive, time consuming, and sometimes even dangerous. In order to perform human tests, special safety conditions and ethical and legal requirements must be taken into account. Emulators that can emulate the physiological functions of the human body could solve these difficulties. In this study, the heart rate depended electrocardiogram templates for this kind of an emulator were extracted. The real-life electrocardiogram preprocessing included a high-pass filter and a Savitzky-Golay filter. A beat detection algorithm was developed to detect QRS complexes in the signals and classify beat artefacts based on the RR interval sequences and two adaptive thresholds. Heart rate levels were detected using the K-means clustering technique. Vectorcardiogram signals were converted from the electrocardiogram signals using the inverse Dower's transformation matrix, and vectorcardiogram templates were extracted to the respective heart rate levels. Finally, a graphical user interface was created for the mentioned methods. The developed beat detection algorithm was tested with the MIT-BIH Arrhythmia Database and the comparison was made with the state-of-the-art algorithms. The beat detection algorithm resulted the sensitivity of 99.77 %, precision of 99.65 %, and detection error rate of 0.58 %. Based on the results, the proposed methods and extracted vectorcardiogram templates were successful.

Keywords: QRS detection algorithm, vectorcardiogram, beat artefact, RR interval, clustering

Kujala P. (2023) Syykkeestä riippuvien elektrokardiogrammimallien poiminta kehoemulaattoriympäristöön. Oulun yliopisto, Lääketieteen tekniikan tutkinto-ohjelma, 92 s.

TIIVISTELMÄ

Lääketieteellisten laitteiden ja analyysimenetelmien kehitystyö sisältää usein ihmisille suoritettavia testejä, jotka ovat kalliita, aikaa vieviä ja joskus jopa vaarallisia. Ihmiskokeiden toteuttamiseksi on otettava huomioon erityisiä turvallisuusehtoja, sekä eettisiä ja laillisia vaatimuksia. Emulaattorit, jotka pystyvät jäljittelemään ihmiskehon fysiologisia toimintoja, voivat olla ratkaisu näihin ongelmiin. Tässä tutkimuksessa sykkeestä riippuvia elektrokardiogrammimalleja poimittiin tämän tyyppiselle emulaattorille. Tosielämän elektrokardiogrammisignaalien esikäsittely sisälsi ylipäästösuodattimen ja Savitzky-Golay suodattimen. Sydämen lyöntien tunnistuss algoritmi kehitettiin tunnistamaan QRS-komplekseja signaaleista ja luokittelemaan lyöntiartefakteja RR-intervallisekvenssien ja kahden adaptiivisen kynnsarvon perusteella. Syketasot tunnistettiin käyttämällä K-means klusterointitekniikkaa. Vektorikardiogrammisignaalit muunnettiin elektrokardiogrammisignaaleista käyttämällä käänteistä Dowerin muunnosmatriisia ja vektorikardiogrammimallit poimittiin vastaaville syketasoille. Lopuksi luotiin graafinen käyttöliittymä mainituille menetelmille. Kehitetty lyöntien tunnistus algoritmi testattiin MIT-BIH Arrhythmia Database-tietokannalla ja vertailu suoritettiin vastaavien algoritmien kanssa. Algoritmi suoriutui 99,77 % herkkyydellä, 99,65 % spesifisyydellä ja 0,58 % virheprosentilla. Tulosten perusteella ehdotetut menetelmät ja poimitut vektorikardiogrammimallit olivat onnistuneita.

Avainsanat: QRS tunnistus algoritmi, vektorikardiogrammi, lyöntiartefakti, RR-intervalli, klusterointi

TABLE OF CONTENTS

ABSTRACT

TIIVISTELMÄ

TABLE OF CONTENTS

FOREWORD

LIST OF ABBREVIATIONS AND SYMBOLS

1. INTRODUCTION.....	8
2. ELECTROCARDIOGRAPHY.....	10
2.1. Anatomy of the Human Heart	10
2.2. Electrical Activity of the Heart	11
2.3. ECG Measurement.....	14
2.3.1. Conventional 12-Lead ECG	14
2.3.2. ECG Signal.....	16
2.3.3. Artefacts	20
2.3.4. Vectorcardiography	23
3. ECG ANALYSIS AND PROCESSING	29
3.1. Heart Rate and Its Variability	29
3.1.1. Heart Rate Variability Analysis Methods.....	30
3.1.2. Morphological Variations.....	33
3.2. Signal Preprocessing	39
3.2.1. Sampling Rate Selection	39
3.2.2. Filter Selection	39
3.3. QRS Detection	41
3.4. Clustering.....	43
4. METHODS	47
4.1. Preprocessing Methods	47
4.2. Beat Detection Algorithm.....	48
4.2.1. R-Peak Detection	49
4.2.2. Artefact Classification	50
4.2.3. Artefact Correction.....	53
4.3. Template Extraction	54
4.3.1. Heart Rate Level-Based Clustering	54
4.3.2. VCG and Inverse Dower's Matrix.....	55
4.3.3. Template Verification.....	56
4.4. Graphical User Interface.....	57
5. RESULTS.....	60
5.1. Data Used in Tests	60
5.2. Outcome	61
6. DISCUSSION	69
7. SUMMARY	72
8. REFERENCES	73
9. APPENDICES	86

FOREWORD

This paper is the culmination of my studies in Biomedical Engineering. I completed my Bachelor's degree in Electrical Engineering, and decided to study my Master's degree in Biomedical Engineering in Signal and Image Processing. I would like to thank my parents, who taught me the perseverance and self-belief I needed to accomplish this achievement. Without your support, I would not be in this situation.

Professor Tapio Seppänen offered me the opportunity to join his research group in the Center of Machine Vision and Signal Analysis research center, which I decided to accept. I would like to express my honestest gratitude to my supervisor, Professor Tapio Seppänen, who guided me through this thesis and supported me throughout my Master's degree programme. I am grateful to have you as my supervisor and personal contact professor.

In addition, special thanks to my technical supervisor, Kai Noponen, who offered me support from the start to the end line. I received numerous good perspectives on the implementation of the work from you. Your feedbacks and guidance were really valuable throughout the work, and our discussions gave me motivation and helped me to achieve the best possible outcome for this paper.

Oulu, September 6th, 2023

Peetu Kujala

LIST OF ABBREVIATIONS AND SYMBOLS

ECG	electrocardiogram
VCG	vectorcardiogram
HR	heart rate
GUI	graphical user interface
Na ⁺	sodium ion
K ⁺	potassium ion
Ca ²⁺	calcium ion
SA node	sinoatrial node
AV node	atrioventricular node
bpm	beats per minute
V _m	transmembrane voltage
aV _L , aV _R , aV _F	augmented limb leads
V ₁ -V ₆	precordial leads
AHA	American Heart Association
RA	right arm
LA	left arm
RL	right leg
LL	left leg
IEC	International Electrotechnical Commission
MIT-BIH	Massachusetts Institute of Technology-Beth Israel Hospital
EHV	electric heart vector
Hz	hertz
EMI	electromagnetic interference
EMG	electromyography
BCOM	body center of the mass
2D	2-dimensional
3D	3-dimensional
1D	1-dimensional
SVEC	stereovectorelectrocardiography
ANS	autonomic nervous system
SNS	sympathetic nervous system
PNS	parasympathetic nervous system
RRI	R-R interval
HRV	heart rate variability
HF	high-frequency
RSA	respiratory sinus arrhythmia
LF	low-frequency
VLF	very-low-frequency
ULF	ultra-low-frequency
PSD	power spectral density
FFT	fast Fourier transformation
AM	autoregressive model
ARMA	autoregressive moving average model
MA	moving average model

PVC	premature ventricular contraction
A-V block	atrioventricular block
RBBB	right bundle-branch block
LBBB	left bundle-branch block
A/D	analog-to-digital conversion
LPF	low-pass filter
HPF	high-pass filter
dB	decibel
LFF	line-frequency filter
IIR	infinite impulse response
FIR	finite impulse response
QD	quartile deviation
NaN	Not a Number
SEN	sensitivity
PRE	precision
DER	detection error rate
TP	true positive
FN	false negative
FP	false positive
<i>cm</i>	centimetre
<i>mV</i>	millivolt
<i>ms</i>	millisecond
<i>s</i>	second
P_X, P_Y, P_Z	potentials of the individual electrode clips
f_s	sampling frequency
f_c	cut-off frequency
$H(z)$	transfer function
$D\%$	difference percentage
V	VCG lead matrix
M	transformation matrix
E	ECG lead matrix
D	inverse Dower's matrix
Φ	electric potential
α	scaling factor

1. INTRODUCTION

Medical device development focuses on developing different innovations (such as instruments, machines and software) to enhance the health of humans and the quality of medical treatments. Often, in the development of such innovations, the best source of data would be people. However, interacting with the human body is expensive, time consuming, and even dangerous in certain situations. Tests with human subjects involves special safety conditions, and ethical and legal requirements. Hence, the demand of emulators that can emulate the physiological functions of the human body is increasing.

This study is part of the project, which aims to develop a phantom, an artificial human torso that can be used for testing medical devices and analysis methods. Phantom emulates human physiological properties and the motion of a human torso. Physiological signals induced in the emulator can be used on testing the properties for medical devices. The motion of the phantom is produced by using a hexapod platform.

In the earlier developments in the project, the base structure of the electrocardiogram (ECG) signal synthesis for the human torso emulator has been implemented. This includes some preprocessing methods for the real-life ECG signal, RR-interval generation model, and vectorcardiogram (VCG) template extraction to the different heart rate levels. However, the limitation of the earlier developments is that they are designed to work only for a specific ECG data. In addition, none verification methods exists for the aforementioned processes. Since the goal of the project is to be able to generate several ECG synthesis templates for different situations, it is important to be able to work with various ECG data. Furthermore, the verification of the beat detection is an important factor for the quality of the generated VCG templates. Hence, the overall goal of this research is to further develop the existing methods, their accuracy, and ability to work with several ECG data.

The aim of this study is to develop heart rate related ECG templates for the aforementioned emulator environment. ECG signal template generation aims to develop realistic heart beat intervals and workload related heart rate (HR) models. These developed models can be used as source signals to generate realistic modeling cardiac signals in the phantom utilizing three dipole sources placed within the thorax. Inside the phantom, the current is carried by ions mixed to the gel. Potential differences across the surface of the phantom are measured with electrodes placed on the torso. In addition, this study aims to create a graphical user interface (GUI) for the template extraction process. This way the user can adjust the parameters used in the process. The developed GUI is presented in Section 4.4.

This research does not include other physiological signal implementations. These other signal libraries are required to build a multimodal physiological signal library, which is the ultimate goal of the project. This paper also does not cover the mechanical or electrical properties of the phantom, except to the extent of this chapter. In addition, there are various methods for the ECG synthesis. However, this study mainly focuses on the development of the beat detection algorithm and other internal functions inside the ECG synthesis process. Therefore, this research only focuses on the method selected for this study and only refers to the other methods with an extend of acknowledgment.

This thesis is structured in the following manner: Chapter 2 describes the anatomical and electrical properties of the heart, ECG measurement, ECG signal properties and artefacts, and vectorcardiography. Chapter 3 includes heart rate and its variability, ECG signal preprocessing methods, beat detection and related researches, and clustering properties. Chapter 4 introduces the methods and tools used in this research, including signal preprocessing methods, beat detection algorithm functions, VCG template generation methods, and graphical user interface implementation. Chapter 5 presents the results. Chapter 6 and Chapter 7 represents the discussion and summary parts, respectively.

2. ELECTROCARDIOGRAPHY

Electrocardiography is a noninvasive method to measure the electrical activity of the heart. The signal produced is called an electrocardiogram (ECG). It records potential differences between given sites on the body surface that vary during the cardiac cycle [1]. More specifically, it reflects the differences in transmembrane voltages in cardiac muscle cells that appear during depolarization and repolarization within each cardiac cycle. One ECG beat corresponds to one cycle of the electrical activity of the heart. The traditional ECG measurement and the ECG signal are further explained in Section 2.3.1 and Section 2.3.2, respectively.

The first human ECG was recorded in the year 1887 by Augustus Waller (1856-1922) using a capillary electrometer [2, 3, 4]. However, the ECG did not become clinically relevant until the year 1901, when Willem Einthoven (1860-1927) developed his string galvanometer for this purpose [3]. Einthoven was awarded the Nobel Prize in the year 1924 in the field of physiology or medicine, "for the discovery of the mechanism of the electrocardiogram"¹. Nowadays, the ECG has become the most widely used noninvasive tool for cardiac investigation [5].

2.1. Anatomy of the Human Heart

The heart is about the size of a fist and weighs around 250-300 grams. It is located in the chest between the lungs behind the sternum and above the diaphragm, about 1.5 centimetres (cm) to the left of the midsagittal plane. The orientation of the heart is tilted slightly in a way that the right side of the heart covers the anterior aspect and the left side the posterior aspect, respectively. The heart is surrounded by a fibroserous sac called the pericardium [6]. The walls of the heart are consisted of the cardiac muscle, called myocardium. [7 p.182-183]

The heart contains four chambers: two atria (the right and left) and two ventricles (the right and left). Connected to the heart are the great vessels: the superior and inferior vena cava, the pulmonary artery and vein, as well as the aorta. The heart has four valves that control the heart's blood flow. Between the right atrium and ventricle lies the tricuspid valve, which controls the blood flow from the right atrium to the right ventricle. The pulmonary valve is located at the border of the right ventricle and the pulmonary artery. The mitral valve stands between the left atrium and ventricle, and the aortic valve controls the blood flow from the left ventricle to the aorta. [7 p.182-184]

The anatomy of the heart is shown in Figure 1. The colors in the figure demonstrate the blood flow in the heart. Blue color indicates deoxygenated blood and red oxygenated blood, respectively. The deoxygenated blood returns from the systemic circulation to the right atrium via the superior and inferior venae cavae. The superior vena cava drains blood from the upper half of the body, and the inferior vena cava returns blood from the lower half of the body [6]. From the right atrium, the blood continues through the tricuspid valve to the right ventricle. The right ventricle pumps the blood into the pulmonary artery via the pulmonary valve. The pulmonary artery

¹<https://www.nobelprize.org/prizes/medicine/1924/summary/>

divides into the right and left pulmonary arteries, which carry the blood to the right and left lungs, respectively.

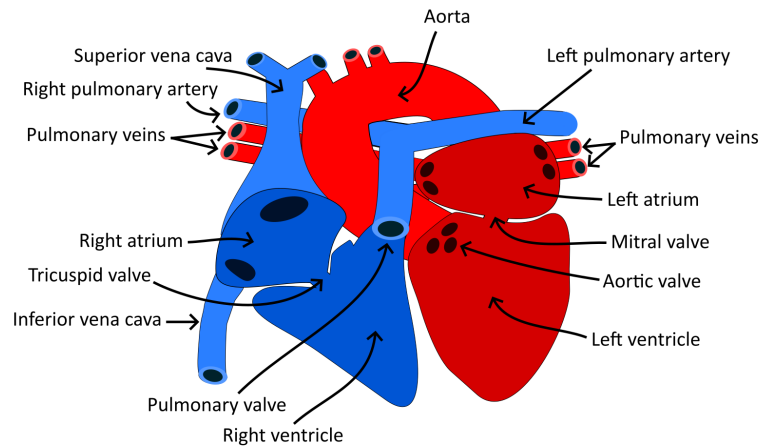


Figure 1. The anatomy of the heart.

In the lungs, the blood is oxidized and the carbon dioxide produced by cellular respiration is removed from the blood. Oxygenated blood returns from the lungs via the pulmonary veins to the left atrium. From there, the blood travels to the left ventricle through the mitral valve. Finally, the blood is pumped into the aorta from the left ventricle through the aortic valve. Since the left ventricle pumps the blood to the systemic circulation, its pressure is much higher than in the right ventricle, which pumps the blood to the pulmonary circulation [7 p.183]. The cardiac cycle is composed of systole (ventricular contraction) and diastole (ventricular relaxation). During systole, blood pressure is at its highest (systolic pressure) and during diastole, blood pressure reaches its lowest level (diastolic pressure) [8].

Oxygen is not delivered to the heart from the blood circulating inside it, but to transport the blood, the heart needs its own blood vessels. Coronary arteries are arteries running on the surface of the heart that carry oxygen and nutrients to the heart muscle itself [9]. The left and right coronary arteries are the earliest branches of the aorta, and divide into smaller branches along the heart. Mainly, the right coronary artery supplies the right side of the heart and the left coronary artery the left side of the heart, respectively. Typically, the right coronary artery is bigger than the left.

2.2. Electrical Activity of the Heart

The simplified ionic flow of a cardiac muscle cell (cardiomyocyte) during an action impulse is illustrated in Figure 2. Cardiac muscle cells and their environment contain different concentrations of ions, mainly sodium (Na^+), potassium (K^+) and calcium (Ca^{2+}). Potassium occurs at a higher concentration inside the cells, while sodium and calcium concentrations are higher in the extracellular space [10]. In cardiac muscle cells, an action potential of about 100 millivolts (mV), propagates due to sequential activation and inactivation of ion channels. Electric activation takes place from the inflow of sodium and calcium ions across the cell membrane [10, 11]. This phase is

called depolarization. The cells are connected together by gap junctions, which have a low resistance, and so the activity in one cell is easily spread to neighboring cells [12, 13, 14].

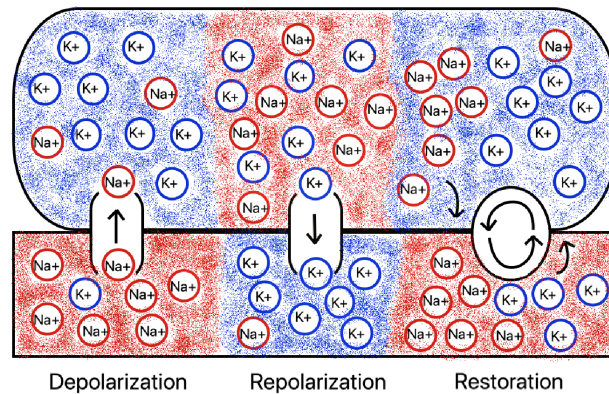


Figure 2. The ionic flow of the cardiomyocyte.

The conduction system of the heart is presented in Figure 3. The action potential of the heart is generated in the sinus node (sinoatrial or SA node), located in the right atrium at the superior vena cava [14]. The cells of the SA node are self-excitatory, so called pacemaker cells. The depolarization propagates from the SA node throughout the atria into the atrioventricular node (AV or A-V node), which is located at the borderline of the atria and ventricles [7 p.188]. The atria and the ventricles are isolated electrically, and the electrical impulses can only pass from the atria to the ventricles via the AV node.

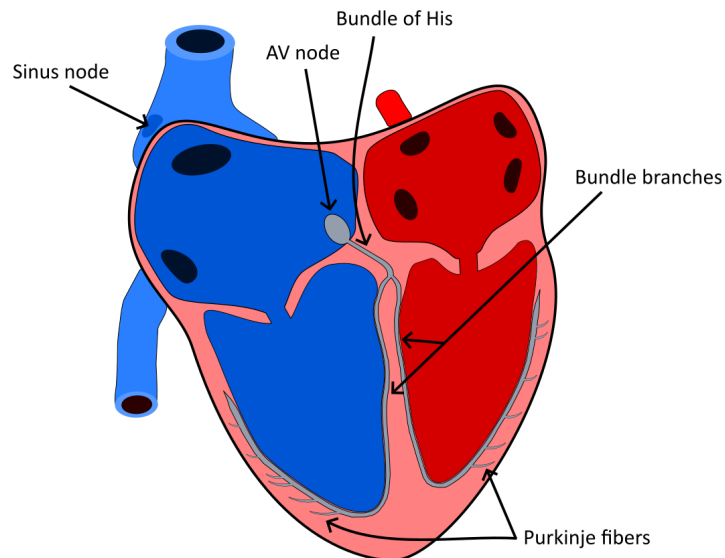


Figure 3. The conduction system of the heart.

The AV node, like the SA node, is a pacemaker. The propagation system of the AV node is proximally consist of a common bundle, called the bundle of His [7 p.188]. More distally, this common bundle separates into two bundle branches, which spread along each side of the septum, forming the right and left bundle branches. These bundles ramify into Purkinje fibers, which diverge to the inner sides of the ventricular walls. The propagation through the AV node is characterized by a conduction delay that ensures that the ventricular contraction occurs after the atria empty its blood into the ventricles [7 p.392]. If the AV node does not receive a stimulus from the atria, it will generate an electrical impulse at a rate of 40-60 beats per minute (bpm).

After cardiac depolarization, a plateau phase takes place. In this phase, the membrane potential (or transmembrane voltage, V_m) of the cells remains almost constant [7 p.185]. Repolarization follows the plateau phase and is a result of the outflow of potassium ions [10]. Unlike depolarization, repolarization is not a propagating phenomenon. Repolarization occurs because the action pulse has only a certain duration, not because the repolarization of a neighboring cell [7 p.195]. At the outer side of the cardiac muscle (epicardium), the duration of the action impulse is much shorter than at the inner side of the cardiac muscle (endocardium) [7 p.188]. This creates an illusion, where the termination of the activity appears as if it were propagating from the epicardium towards the endocardium.

Figure 4 shows the action impulse of the cardiac muscle cell. The duration of the cardiomyocyte's action impulse is about 300 milliseconds (ms). The time to cardiac muscle cell to complete the activation is relatively short, roughly 1 ms, where the recovery of the cell takes approximately 100 ms. The resting region of the cell is around -80 mV. [7 p.185-196]

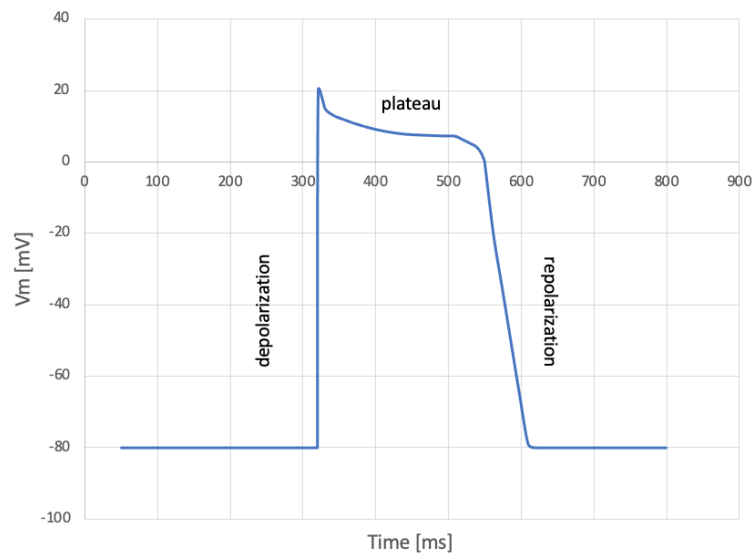


Figure 4. The action impulse of the cardiomyocyte.

2.3. ECG Measurement

2.3.1. Conventional 12-Lead ECG

The methods for measuring ECG have been developed since Waller made the first human ECG recordings. Waller used five electrodes, one on each of the four extremities and the mouth, with 10 leads derived from the different combinations [15]. Willem Einthoven reduced the number of electrodes to three, including only both arms and the left leg. Remaining electrodes resulted three leads, which formed the so called Einthoven's triangle. Frank Norman Wilson (1890-1952) improved Einthoven's system by adding a central terminal in the center of Einthoven's triangle [16]. This increased the accuracy of electrocardiographic diagnosis. In the year 1942, Emanuel Goldberger presented augmented leads, which produced 50 % larger signals than the Wilson central terminal. Precordial leads (chest leads) were presented by Wilson in the year 1944. [7 p.387-401]

Today, the standard 12-lead ECG consists of 3 bipolar limb leads (I, II and III), 3 augmented unipolar limb leads (aV_L , aV_R and aV_F), and 6 precordial leads (V_1 - V_6). The limb leads view the heart in a frontal and sagittal plane, where the precordial leads record the heart's electrical activity in a transverse plane. The American College of Cardiology, together with the American Heart Association (AHA) and the Heart Rhythm Society, has formulated guidelines and set technical standards for ECG equipment [1]. Limb lead electrodes are placed on the arms and legs distal to the shoulders and hips, and six electrodes are placed on the chest. In addition, two general color codes for leads by the International Electrotechnical Commission (IEC) and the AHA have been made to avoid misplacement. The color code used depends on the location of use. More detailed electrode placements and proper color codes for the standard 12-lead ECG are shown in Table 1. Electrode placements are further demonstrated in Figure 5. [1]

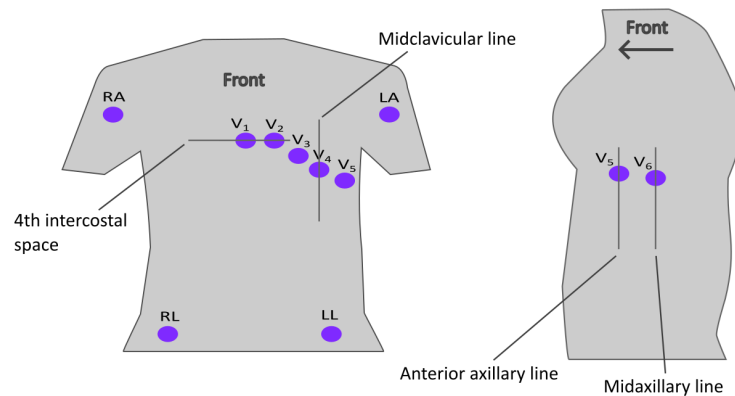


Figure 5. Electrode placements for the 12-lead ECG.

Skin preparation and electrode placement have important effects on the ECG. It has been shown, that cleaning and gentle abrasion before electrode placement can reduce noise and thus improve the quality of the recorded ECG [17]. If possible, the patient should be placed in the supine position, generally with a pillow under the head. Patient

Table 1. Electrode placements and lead color codes

Electrode placement	<i>IEC</i>¹		<i>AHA</i>²	
	Electrode	Color code	Electrode	Color code
Right arm	R	Red	RA	White
Left arm	L	Yellow	LA	Black
Right leg	N	Black	RL	Green
Left leg	F	Green	LL	Red
Fourth intercostal space right sternal border	C1	Red	V1	Red
Fourth intercostal space left sternal border	C2	Yellow	V2	Yellow
Midway between V2 and V4	C3	Green	V3	Green
Fifth intercostal space in the midclavicular line	C4	Brown	V4	Blue
In the horizontal plane of V4 at the anterior axillary line	C5	Black	V5	Orange
In the horizontal plane of V4 at the midaxillary line	C6	Purple	V6	Purple

¹⁾ International Electrotechnical Commission.

²⁾ American Heart Association.

positional change, such as elevation or rotation, can affect the recorded amplitudes and axes. [1]

Although the 12-lead system has become the standard method for ECG measurement, various alternative lead applications have been developed [18]. Placing the limb leads on the torso can reduce the noise caused by the motion of the arms and legs during ambulatory or exercise ECG, as suggested by Mason and Likar [19] in the year 1966. The Mason-Likar modification is the most important modification of the 12-lead system used in exercise ECG. Motion artefacts of the limbs are also a particular problem in routine recordings of young children, and torso leads are sometimes utilized in those situations. [1]

Nowadays, most of the electrocardiograms are recorded by digital machines, which can generate automatically preliminary findings based on the morphology criteria. These modern ECG machines are equipped with software that measures ECG intervals and amplitudes, provides a virtually instantaneous interpretation, and compares the tracing to earlier measurements. Machine learning methods, like deep learning, are often utilized in these innovations [20, 21]. [1]

Currently, large amounts of data are available and the computational power to handle them have increased significantly. The World's four international standard ECG databases (Massachusetts Institute of Technology-Beth Israel Hospital (MIT-

BIH) Arrhythmia Database², AHA ECG Arrhythmia Database³, European Society of Cardiology Database⁴, and European ST-T ECG Database⁵) offers large quantities of ECG data. Besides this, wide volumes of ECG data has been recorded in hospitals, research institutions, and homes. [26]

2.3.2. ECG Signal

Formation

Cardiac cells are coupled together by gap junctions. This helps the activation to propagate to the neighboring cells. In the rest, the transmembrane voltage is negative. When the cell is depolarized, the transmembrane voltage is in the plateau phase, and hence positive. While the activation front propagates, it produces voltage differences between the activated cells and the cells that are still at rest. This phenomenon causes dipole moments along the propagation front, and the resulting source can be approximated as a double layer. If the observation point is much greater distance than the radius of the source, the field appears to originate from a single dipole, whose strength and orientation are the vector sum of the source components. The resulting vector is called the electric heart vector (EHV). [7 p.265-271]

Electrodes record the electric potential occurring beneath them. The potential at the electrode E is the scalar product of the source dipole vector \mathbf{p} and the lead vector \mathbf{c} . This is demonstrated in Equation (1):

$$\Phi_E = \mathbf{c} \bullet \mathbf{p} = |\mathbf{c}| |\mathbf{p}| \cos \theta \quad (1)$$

where Φ_E is the electric potential at the electrode, \mathbf{p} is the sum of the potentials of each dipole component, and θ is the angle between the vectors \mathbf{c} and \mathbf{p} . The lead vector \mathbf{c} is a three-dimensional transfer coefficient which describes how the dipole source influences the potential at a given electrode relative to the potential at the reference location. The value of the lead vector depends on the locations of the dipole source, the electrode, the reference point, and the distribution of resistivity of the materials surrounding the dipole source. [7 p.272-275]

The lead voltage measurement is performed as a potential difference between two electrodes (bipolar) or between an electrode and a reference point (unipolar). The reference point can be, for example, the Wilson central terminal [27, 28], which is the average of the limb potentials. The potential difference between any two points P_i and P_j is defined as in Equation (2):

$$V_{ij} = \Phi_i - \Phi_j \quad (2)$$

where V means the lead voltage and Φ is the electric potential at a given point [7 p.274]. In the case of unipolar electrode, the reference point is usually approximated as a zero-potential point.

²MIT-BIH Arrhythmia Database. PhysioNet URL: <https://doi.org/10.13026/C2F305> [22, 23]

³Not available for public use. PhysioNet URL: <https://doi.org/10.13026/C27P4P> [24, 23]

⁴European Society of Cardiology. Website URL: <https://www.escardio.org>. Accessed 10.5.2023

⁵European ST-T ECG Database. PhysioNet URL: <https://doi.org/10.13026/C2D59Z> [25, 23]

ECG machines convert the lead voltages into waves. If the dipole sum vector is pointing away from the positive unipolar electrode, it is shown as a negative wave (downward). If the dipole sum vector is pointing towards the unipolar electrode, it is introduced as a positive wave (upwards). When the unipolar electrode is somewhere between the above scenarios, the resulting wave is a combination of the positive deflection for the energy pointing towards the electrode and the negative deflection for the energy pointing away from it. This is visualized in Figure 6. If the polarity of the electrode is changed to negative, these rules are reversed. Depolarization produces the EHV pointing in the direction of the propagation. Conversely, repolarization generates the EHV pointing at the opposite direction of the propagation. Although repolarization is not a propagation phenomenon, it can be considered as such for this purpose. [7 p.389-396]

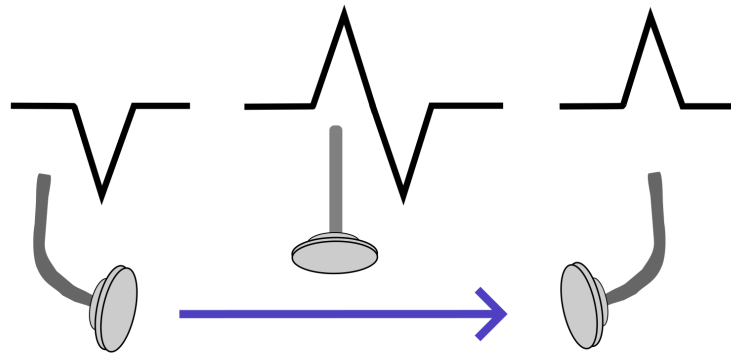


Figure 6. ECG wave formation due to different unipolar electrode placement. The arrow indicates the direction of the dipole sum vector.

The ECG electrodes are placed in the way that they capture a 3-dimensional recording of the heart's electrical axes. The placements of the electrodes were introduced in Section 2.3.1. The leads are formed between the electrodes as shown in Figure 7. The formed leads can be moved to pass the center of the heart, because vectors are equal if they are parallel and of the same intensity [29 p.26].

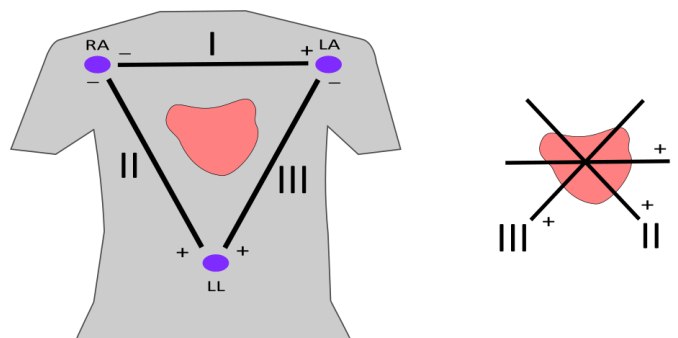


Figure 7. An example of the leads formed between the electrodes. The leads can be moved to pass the center of the heart.

The limb leads (I, II and III) together with the augmented leads (aV_L , aV_R , aV_F) cover the frontal plane of the heart, and the precordial leads (V_1 - V_6) the transverse plane, respectively [7 p.402-403]. The sagittal plane of the heart is recorded by one precordial lead and one augmented limb lead. The projections of the lead vectors are presented in Figure 8. Sometimes the polarity of the lead aV_R is changed to $-aV_R$.

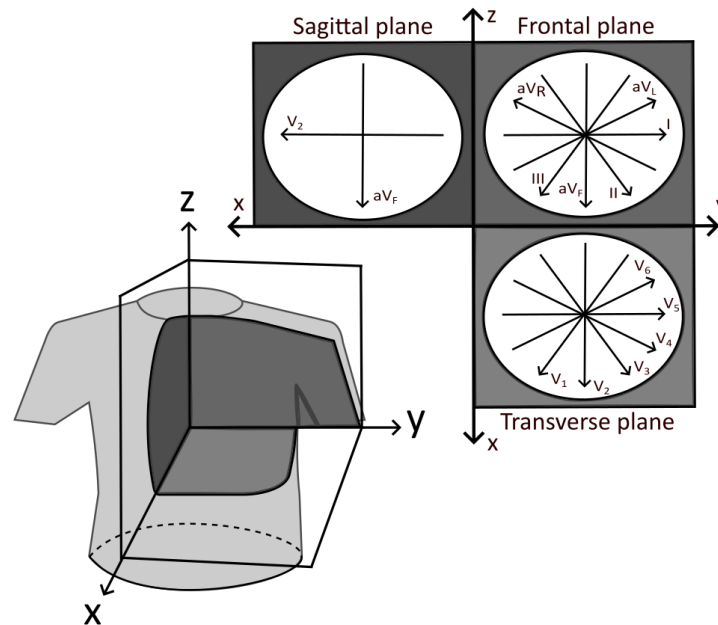


Figure 8. The projections of the lead vectors. The positive pole of the lead is indicated by an arrow and name.

Morphology

A 10 second (s) 12-lead ECG recording is showed in Figure 9. ECG machines convert the electrical activity recorded by the electrodes into waves and draw a trace on a moving electrocardiograph paper or to a digital screen. The paper moves at a speed of 25 millimetres per second (mm/s). Time is plotted on the x-axis and voltage on the y-axis. The x-axis is divided into large and small squares, which represents 0.2 s and 0.04 s, respectively. If the ECG machine is calibrated correctly, an increase of voltage by 1 mV should move the stylus by 1 cm. [30]

ECG signal morphology is demonstrated in Figure 10. The deflections in the signal are marked in an alphabetical order starting with the letter P, which represents atrial depolarization. The right atrium gets depolarized before the left atrium because the SA node is located in the right atrium [30]. This causes the first half of the P-wave to be formed by the depolarization of the right atrium, and the second half by the depolarization of the left atrium. Ventricular depolarization produces the QRS complex, and ventricular repolarization results the T-wave. Atrial repolarization occurs during the QRS complex, but its amplitude is so low that it cannot be seen on a normal ECG. U-wave is a small, flat wave sometimes seen after the T-wave and before the next

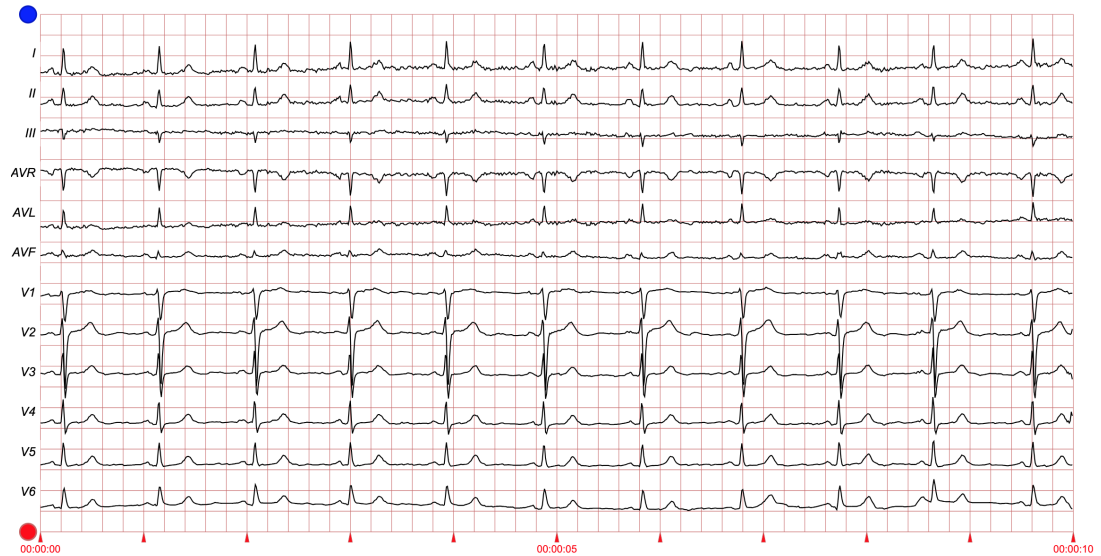


Figure 9. The 12-lead ECG. The lead signals are on their own rows. This figure has been captured from PhysioNet's PTB-XL dataset⁷.

P-wave. Its origin is still not established and it is rarely used for diagnosis in clinical practise [31]. [7 p.392-397]

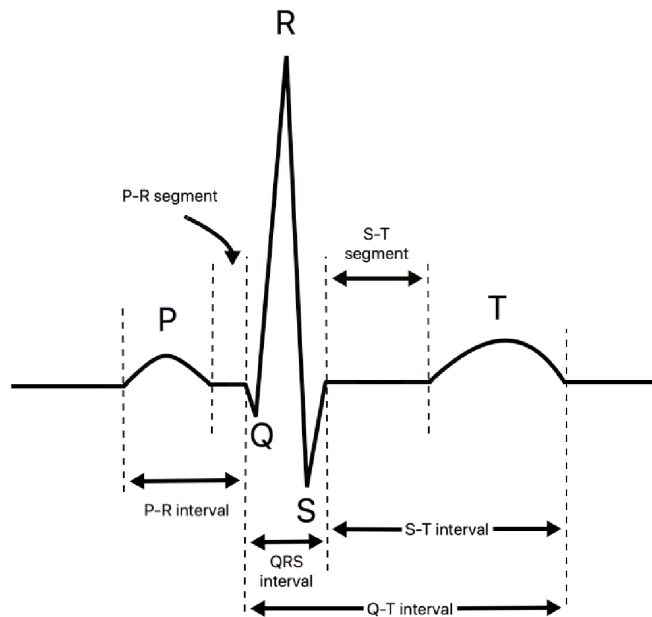


Figure 10. Morphology of the electrocardiogram signal.

A segment is a specific portion of the ECG. For instance, the P-R segment indicates the portion between the end of the P-wave and the beginning of the Q-wave, and represents the phase of the cardiac cycle when the impulse travels from the AV node

⁷[23, 32, 33]

to the ventricles. The S-T segment represents an electrically neutral time for the heart, where the ventricles are between depolarization and repolarization. Physically, it refers to the time that the cardiac muscle maintains contraction in order to push the blood out of the ventricles. [29 p.41-46]

An interval is the distance between two cardiac events, measured as time. For example, the time interval between the beginning of the P-wave and the beginning of the QRS complex is called the P-R interval. It corresponds to all of the events from the generation of an electrical impulse in the SA node to the moment of ventricular depolarization. The Q-T interval represents all of the events of ventricular systolic phase, including the beginning of ventricular depolarization to the end of repolarization. The time between the end of ventricular depolarization and the end of ventricular repolarization is marked as the S-T interval. [29 p.41-49]

2.3.3. *Artefacts*

There are several factors that can cause distortions and affect the quality of the ECG signal. These distortions are also called artefacts. They may be subtle with no interference with correct interpretation, or they can be substantial and cause major deformations in the signal. Artefacts can also be misinterpreted and expose patients to needless or risky treatments [34]. [35]

ECG artefacts are originated from the external causes or from the patient's body [36]. The first can be induced by a dysfunction of the electrodes or cables (misplacement, damaged or detached equipment, improper filters) or by an electromagnetic field (electrical devices, light fixtures). Artefacts originated from the patient's body are caused by muscles, skin, respiration and inhomogeneities of the body. Voluntary and involuntary movements cause muscle contractions that are often the cause of artefacts. Skin can induce artefacts through a poor contact with electrodes, perspiration, or stretch of the epidermis, which produces an electric potential. Electrocardiographic artefacts can closely mimic the native ECG waves and deviated S-T segment [34]. Generally, common ECG artefacts can be identified by looking their clinical context, morphology, and relation to the normal cardiac rhythm. [35]

Electrode misplacement is a common cause of distortions in the ECG signal. The effects of different electrode placements on the ECG signal formation were illustrated in Figure 6. In the year 2007, Rudiger with his colleagues found that electrode misplacement was occurred frequently in outpatient clinics (0.4 %) and even more often in intensive care units (4 %) [37]. Another study, done by Hedén and his colleagues with a total of 11,432 ECG recordings, showed that misplacement was done in nearly 2 % of the instances [38]. Electrode misplacement can lead to changes in the ECG morphology and lead to misdirected therapeutic decisions. For instance, a normal ECG without signs of ischemia in patients with ischemic heart disease have been resulted by electrode misplacement [39]. On the other hand, electrode misplacement have been reported to cause erroneous diagnoses for healthy patients [39, 40].

Baseline wandering is a phenomenon where ECG signal's baseline, so called isoelectric line is distorted. An illustration of the baseline wandering can be seen in Figure 11. The baseline seems to go up and down ('wander') rather than been straight. This kind of distortion is usually caused by the electrode-skin impedance

mismatch, motion of the patient, or respiration. Baseline wandering occurs with the low-frequency bandwidth below 1 hertz (Hz). It has also been noted that cell phones near to an ECG machine can produce artefacts that occur as sharp, fast, and low-amplitude spikes [36]. This kind of disturbance is called electromagnetic interference (EMI). [41, 42]

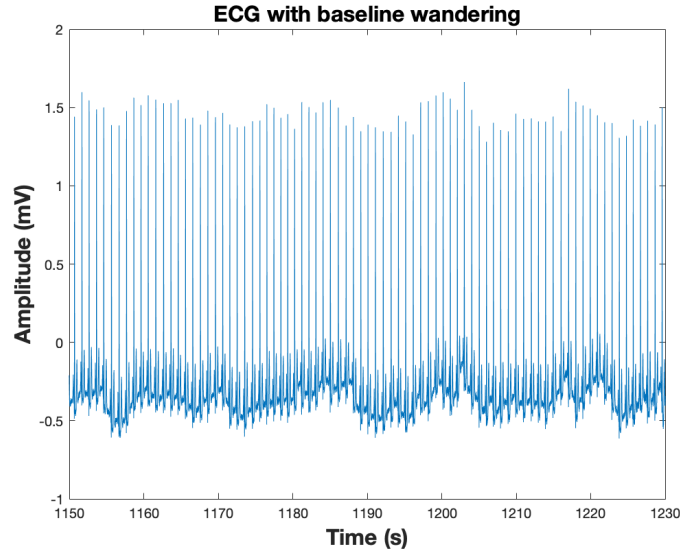


Figure 11. Illustration of the baseline wandering. This figure has been captured from the signal collected from the Physionet's MIT-BIH Arrhythmia Database⁹.

Respiration affects the ECG signal due to the movement of the thorax, the so-called mechanical interaction [43]. This is caused by two factors. Firstly, the volume changes in the lungs that alter the electrical impedance. The distribution of resistivity of the materials surrounding the dipole source was one factor that affected the lead vector described in Section 2.3.2. Secondly, the position of the electrodes with respect to the heart is changing due to the movement of the thorax. The importance of the electrode placement with respect to the dipole source was discussed also in Section 2.3.2. Both of these factors affect the morphology of the ECG signal. In addition, respiration modulates the heart rate. Respiration sinus arrhythmia (RSA) is a well-known phenomenon in the heart rate variability. This is further explained in Section 3.1, where the heart rate variability is introduced.

Movement produces artifacts in the ECG signal. Even small movements of the patient during the measurement phase creates skin movements that can influence the ECG signal. In addition, any muscle activation creates electromyographic (EMG) noise with the degree depended on the activation level. During walking or running, the body center of the mass (BCOM) is constantly moving on a path [44]. When walking, both legs touch the ground simultaneously and the BCOM reaches its highest point at the middle of the stance phase [45]. During running, the ground contact is much shorter and the BCOM reaches its lowest point at the middle of the stance phase. It has been showed that body position can change the ECG signal and heart rate [46]. When the heart rate is higher in the standing stage than in supine, the resulting signal have

⁹[22, 23]

a decreased amplitude in the R-wave. Also, different postures affect the conducting media adjacent to the heart in different ways and therefore the resultant ECG signals are different.

High-frequency noise is often the consequence of powerline interference (50-60 Hz) and EMG noise (100-500 Hz) [47]. The noise wave from powerline interference is typically sinusoidal with random phase but constant frequency [48]. Its origin is, as its name suggests, the power line to recording machines. Powerline interference is always present, even if special precautions are taken. A demonstration of the powerline interference is shown in Figure 12.

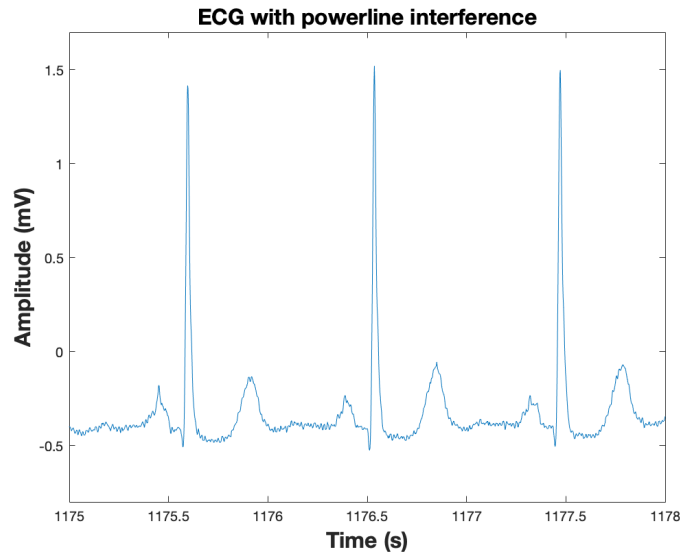


Figure 12. Demonstration of the powerline interference. The signal is noisy between R-peaks. This signal is from the Physionet's MIT-BIH Arrhythmia Database¹¹.

Beat artefacts

In this paper, the 'beat artefact' term is used for general definition of various abnormal RRI values in the ECG signal. These artefacts are divided into three categories: ectopic beats, missed beats, and extra beats. Ectopic beats are caused by physiological factors, and they can be seen as morphological abnormalities in the ECG signal. Morphological variations of the ECG signal and heart rate are further discussed in Section 3.1.2.

In this study, beat detection errors can also be referred as beat artefacts. When the initial beat detection is performed, the software makes detection errors by detecting invalid points in the signal, or ignoring correct R-peaks. Incorrect detection points are called extra beats and missed R-peaks are referred as missed beats. The beat detection algorithm developed in this research is based on the detection of morphological beat artefacts as well as the notice of R-peak detection errors. The functions and architecture of the algorithm are explained in more detail in Section 4.2.

¹¹[22, 23]

2.3.4. Vectorcardiography

In Section 2.3.2, the vector behavior of signals generated by the heart was introduced. The measurement and display of the electric heart vector (EHV) is called vectorcardiography (VCG). The first VCG was introduced by Hubert Mann in the year 1920 [49]. He constructed a monocardigram from the Einthoven's limb leads, which was the projection of the vector loop in the frontal plane. Therefore, it only consist two-dimensional data that did not include back-to-front information from the sagittal and transverse planes. [7 p. 404-406]

Nowadays, the VCG gives a spatial representation of the electric forces generated during cardiac activity, analyzed in three spatial planes. The frontal plane is located between the X and Y leads, the transverse plane between the X and Z leads, and the sagittal plane between the Y and Z leads [50]. In VCG, electrical activity of the heart is described by three loops. The first loop represents the P-wave, the second one, which is the largest loop, corresponds to the QRS complex, and the third loop represents the T-wave.

The method that follows the trajectory of the end point of the vector projected on to principal planes is called spatial vectorcardiography. This kind of a VCG is usually presented in a 2-dimensional (2D) or 3-dimensional (3D) coordinate system, which axes can be either the body axes or the cardiac axes. Figure 13 and Figure 14 illustrate the 2D and 3D VCG, respectively. In the 2D and 3D images, the dependence between individual leads is displayed on the mV scale. The temporal information can be displayed by modulating the intensity of the beam so that the trace is periodically interrupted. Each segment has a teardrop shape, which indicates the direction of the trajectory. [7 p. 404]

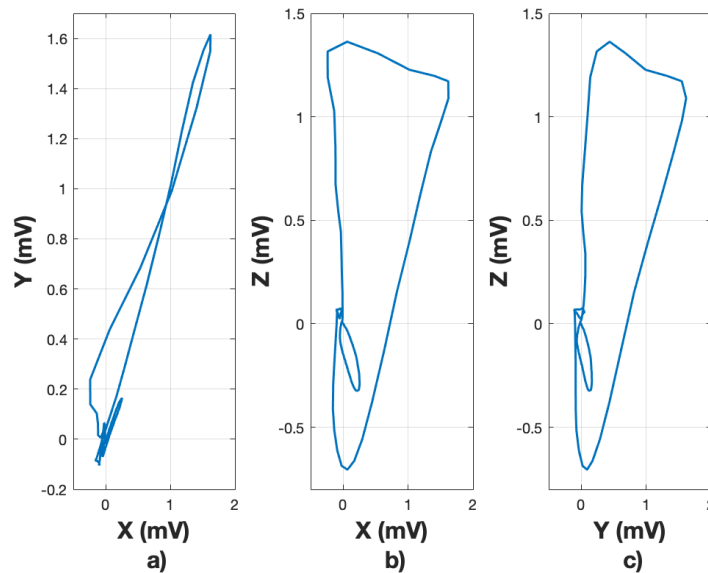


Figure 13. 2D spatial vectorcardiography. a) XY-plane, b) XZ-plane, c) YZ-plane.

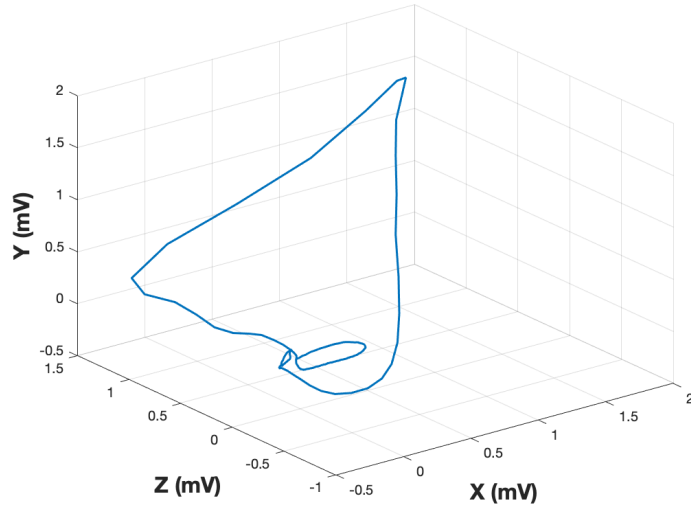


Figure 14. 3D spatial vectorcardiography.

The VCG signal can also be displayed in a 1-dimensional (1D) image as three vector components (X, Y and Z) as functions of time like shown in Figure 15. This is called scalar vectorcardiography.

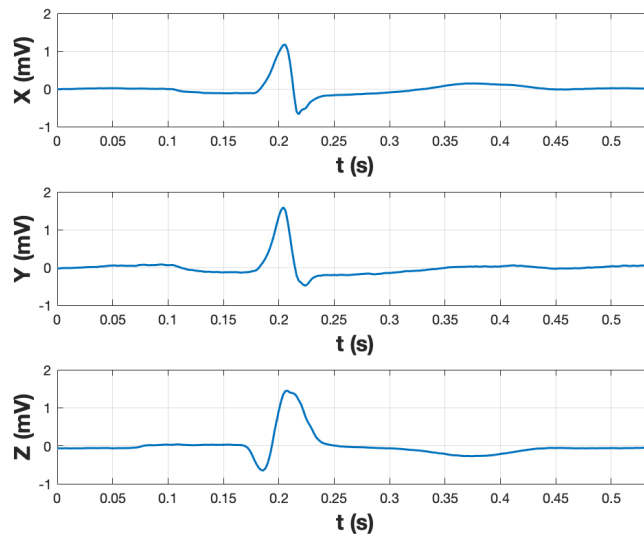


Figure 15. Scalar vectorcardiography. Vector components are displayed as functions of time.

The benefits of the VCG have been recognized for decades [51]. The spatial orientation and the magnitude of the vectors give a better idea of the magnitude and direction of the electrical forces at any moment. Since these properties, the VCG provides a better tool for the diagnosis for some cardiac dysfunctions [51, 52, 53, 54]. Some studies have also shown the combined use of the ECG and VCG to be beneficial [55, 56]. Despite all the benefits of the VCG, ECG is increasingly being used.

Frank lead system

In 1956 Ernest Frank published his orthogonal VCG lead system [57], which nowadays is the most used VCG system throughout the world [58]. Frank's orthogonal lead system consist 7 electrodes as demonstrated in Figure 16. The points of the electrodes are indicated by letters A, C, E, F, H, I and M [7 p. 412]. Electrodes A and I are in the left and right central axillary line, respectively. Electrode E is in the middle of the chest, and electrode M is on the back, opposite to electrode E. The placement of the electrode C is between the electrodes A and E. Electrode F is located on the left leg and electrode H on the neck.

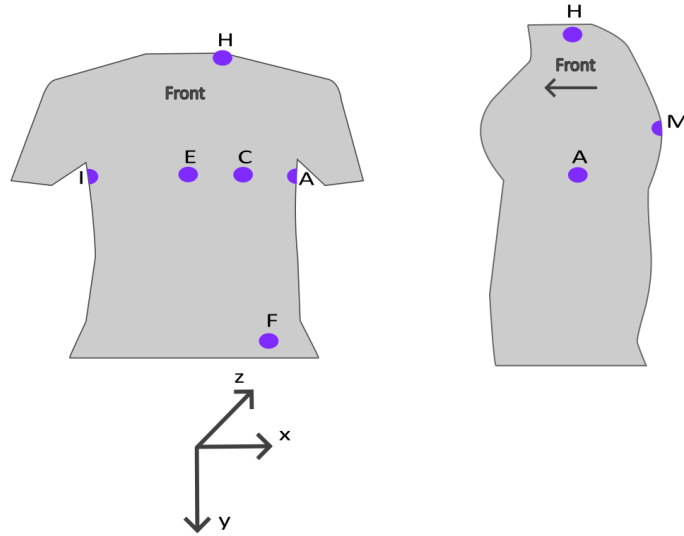


Figure 16. Frank VCG system electrodes.

The right-to-left component is represented as the x-axis, the y-axis is the head-to-foot component, and the z-axis is the front-to-back component [59]. The individual electrodes are routed to the resistor network, which is illustrated in Figure 17 [57]. Mathematical Equation (3), Equation (4), and Equation (5) are used to derive the signals from the leads:

$$P_X = 0,610A + 0,171C - 0,781I \quad (3)$$

$$P_Y = 0,655F + 0,345M - 1,000H \quad (4)$$

$$P_Z = 0,133A + 0,736M - 0,264I - 0,374E - 0,231C \quad (5)$$

where P illustrates the potentials on the individual electrode clips and other letters represent the individual electrodes.

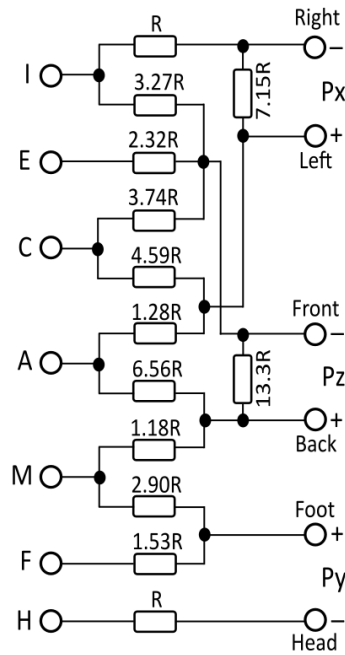


Figure 17. Frank's lead system resistor network.

Other lead systems

In addition to the Frank lead system, various other VCG lead systems have been developed. Schmitt and Simonson developed many versions of VCG lead systems, calling them stereovectorelectrocardiography (SVEC). The third version, SVEC III, was published in 1955 [60]. It requires a total of 14 electrodes and creates a lead field in the thorax which is symmetric in relation to the sagittal plane.

In year 1961, McFee and Parungao published a simple VCG system called the axial system [61], based on a lead field theoretic approach and which consist of three uniform lead fields. They made discover that the closer to the heart the electrodes are placed the more electrodes are needed to achieve a homogeneous lead field within the heart's area.

In his book [7 p.419-420], Malmivuo presents Nelson lead system, which was developed in the year 1971 by Nelson and his colleagues. This lead system, suitable for clinical use, based on the Gabor-Nelson theorem. The lead system includes electrodes placed on three levels of the thorax with eight on each level, one electrode on the head, and one on the left leg. According to Malmivuo, Nelson and his collaborators claimed that on the basis of their measurements, their VCG lead system is much more accurate than the McFee-Parungao or Frank lead systems. In addition, the Nelson lead system should be very insensitive to electrode misplacement.

VCG from ECG

In clinical practice, measurement of the VCG is not usually used because it requires additional electrodes placed on the patient's body. Instead, mathematical

transformations can be used for deriving the VCG from 12-lead ECG and vice versa. The first consideration of the transformation of individual lead systems was introduced by Burger and his colleagues in 1952 [62]. After that, various transformation methods have been developed [59]. In order to convert the ECG leads to VCG, a mathematical transformation can be implemented as a multiplication of two matrices, as shown in Equation (6),

$$\mathbf{V} = \mathbf{M}\mathbf{E} \quad (6)$$

where \mathbf{M} is the transformation matrix, \mathbf{E} is a matrix which rows are formed by the independent ECG leads and \mathbf{V} is a matrix which rows correspond to the derived VCG signals. [50]

Quasi-orthogonal transformation is based on the transformation coefficients used on certain leads in order to achieve the orthogonal leads [50]. Several different quasi-orthogonal transformation methods with different transformation coefficients have been presented. In the quasi-orthogonal, transformation matrix's all input leads have coefficients equal to zero, except for the selected three ECG leads.

In many model-based methods, each ECG lead contributes to a certain extend to a specific orthogonal lead [50]. The individual weight coefficients then form the resulting transformation matrix. The difference between quasi-orthogonal systems and this kind of an approach is that individual VCG leads in the quasi-orthogonal system correspond to exactly one ECG lead, while in this approach the VCG leads are formed by the weight coefficients of the individual ECG leads.

The transformation matrix derived by a regression technique was presented by Kors and his research team in 1990 [63]. The transformation coefficients were derived by minimizing the sum of the squared differences between the measured VCG and the transformed VCG. Guillem et al. [64] introduced a transformation matrix derived using the regression approach, which is optimized for P wave. They named the transformation matrix as PLSV. In addition, they also developed a matrix optimized only for the QRS complex, called QLSV. Both of these matrices were derived from a total of 124 patients. Regression model for each patient were found by the least square value method, and the resulting matrix was given as the mean value of the transformation matrices for all patients.

In 1980, Dower and his colleagues introduced transformation coefficients for deriving the 12-lead ECG from three leads measured by the Frank lead system [65]. Their transformation coefficients are shown in Table 2. Eight years later, Edenbrandt and Pahlm derived a pseudo-inverse matrix that can be used for transformation from 8 independent ECG leads to VCG [66].

In year 1990, Kors et al. [63] compared the Kors quasi-orthogonal transform, the inverse Dower transform, the Kors regression model, and the Frank orthogonal lead system. They used a total 90 recordings. Based on the measurement of the mean absolute deviation and the evaluation of the cardiologists, Kors concluded that the regression and the inverse Dower transformations achieved better results than the quasi-orthogonal transformation. The differences between the regression transformation and the inverse Dower transformation were negligible.

Edenbrandt and his colleagues [67] compared the inverse Dower transformation with the directly measured Frank lead VCG and two quasi-orthogonal methods. The

Table 2. Dower's transformation matrix

Lead	X	Y	Z
I	0.632	-0.235	0.059
II	0.235	1.066	-0.132
III	-0.397	1.301	-0.191
aV_R	-0.434	-0.415	0.037
aV_L	0.515	-0.768	0.125
aV_F	-0.081	1.184	-0.162
V₁	-0.515	0.157	-0.917
V₂	0.044	0.164	-1.387
V₃	0.882	0.098	-1.277
V₄	1.213	0.127	-0.601
V₅	1.125	0.127	-0.086
V₆	0.831	0.076	0.230

research was performed on 80 people with different types of heart attacks. The best result was achieved by the inverse Dower transformation and its recording was the most similar to the Frank lead system.

Guillem et al. [64] tested their PLSV and QLSV methods. They compared their models with the inverse Dower transformation method during the QRS complex and P wave detection on 247 recordings. In the P-wave transformation, the PLSV yielded the best results and the QSLV achieved the best results for the QRS complex detection. Both methods gave better results than the inverse Dower transformation.

Jaros et al. [59] tested different transformation methods in their research including 549 recordings with mixed contribution of healthy patients and patients with pathological heart diseases. They included the Kors regression transformation, the Kors quasi-orthogonal transformation, the inverse Dower transformation, the PLSV, and the QLSV. They made conclusions based on the correlation coefficients and mean squared error in relation to the reference Frank lead system. The Kors regression method yielded the best results for the X and Y leads. In case of the Z lead, none of the methods were statistically significant compared to others. The smallest accuracy was performed by the Kors quasi-orthogonal method.

3. ECG ANALYSIS AND PROCESSING

In this chapter, the ECG analysis and signal processing concepts relevant to this study are reviewed. Heart rate (HR) and heart rate variability (HRV) can be calculated from the ECG signal. HRV has been shown to be a useful tool for analyzing the physiological status of the subject and is widely used for different ECG analysis methods. HRV is affected by different morphological variations of the ECG signal due to artefacts or physiological factors. Therefore, it is important to understand the effects and backgrounds of these factors. HRV and morphological variations of the ECG signal are discussed in Section 3.1.

Artificial ECG signal development process starts by real-life ECG signal processing. Signal acquisition can be done by implementing own ECG measurement process or utilizing already acquired signals. The data used in this research is introduced in Section 5.1. ECG signals contain noise due to the measurement situation and other physiological signals, and therefore the ECG signals need to be preprocessed before they can be used by any manner. Preprocessing reduces artefacts in the signal and makes it clearer. Signal processing methods are generally discussed in Section 3.2 and presented further in the aspect of this research in Section 4.1.

Typically, some kind of beat detection algorithm is utilized to detect QRS-complexes in the signal and calculate signal properties, such as heart rates and beat intervals (RR intervals). This study comprises the development of a beat detection algorithm, that can be used to detect normal heartbeats from several ECG signals. The algorithm calculates heart rates and beat intervals from the signals. In addition, the algorithm locates and classifies artefact beats (missed, extra, and ectopic beats), and makes the right corrections for them. State-of-the-art beat detection algorithms are presented in Section 3.3 and the developed beat detection algorithm is introduced in Section 4.2.

Artificial ECG signal generation includes extracting VCG templates from the ECG signal to the correct heart rate levels using mathematical models. In this study, these heart rate levels are detected using clustering method. Clustering is introduced generally and in the aspect of this research in Section 3.4 and Section 4.3.1, respectively.

As mentioned in Chapter 1, there are several different ECG synthesis methods developed [68, 69, 70, 71, 72]. However, the main focus of this study is on the technical aspects of the selected method. Therefore, other ECG synthesis methods are excluded from this research.

3.1. Heart Rate and Its Variability

Heart rate (HR) is the number of heartbeats per minute. Normal HR is 60 to 100 beats per minute (bpm). If the rate is lower, it is referred to as bradycardia, and if higher, it is called tachycardia [30]. The time interval between two successive R-peaks is called RR interval (RRI) (also NN interval) or beat interval. Heart rate variability (HRV) is the variation of RR intervals over the time period.

HR is regulated by the autonomic nervous system (ANS), which consist of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The parasympathetic nerve is often indicated as the vagal nerve since it is the primary

parasympathetic nerve [73]. Increased activity of the SNS or decreased activity of the PNS causes acceleration of the HR [74]. Conversely, increased activity of the PNS or decreased activity of the SNS induces deceleration of the HR.

The SNS activation boost is occurring in response to stress and exercise. It increases the action impulse generation rate of pacemaker cells in the SA node [73]. The PNS regulates the function of circulatory system, glands, digestion system and urinary organs. Its activation increases due to the function of internal organs (for instance, increased activity of the digestive system after eating), trauma, or allergic reactions. Increased activation of the PNS decreases the firing rate of the SA node. The differential effects of the ANS on the SA node, and thus the HR, are due to the differential effects of the neurotransmitters for the SNS (norepinephrine) and PNS (acetylcholine) [75]. Together, these nervous systems provide a regulatory balance of physiological autonomic functions.

HRV has been associated with several physiological factors and health status. In general, increased HRV is usually seen as associated with a good health condition, whereas lowered HRV might imply pathological changes [76]. One of the main clinical implementations of the HRV is the risk prediction of sudden cardiac death after acute myocardial infraction [74, 77]. In addition, HRV is noticed to provide early warning signs of diabetic neuropathy [74]. Studies show that HRV decreases with depression, and therefore depression has a negative impact on survival after myocardial infraction [78, 79]. It has been studied that HRV also depends on the gender and age. The variation of the HR for healthy subjects with the age between 20 and 70 were studied by Bonnemeier et al. [80]. They found that HRV decreased with age and the variation was greater in women than in men. Nagy and her colleagues studied HR differences in human neonates [81]. Their research showed that HRV is lower in boys than in girls. Kiviniemi and his research group suggests based on their research, that HRV may serve as an objective tool for tailoring an aerobic training program in men and women [82]. Besides these, HRV has been studied with relation to renal failure, blood pressure, drugs, smoking, alcohol, and sleep [74].

3.1.1. Heart Rate Variability Analysis Methods

Since HR is regulated by the ANS, HRV is a useful tool for analyzing the status of the ANS noninvasively [74]. The degree of variability in the HR gives information about the functioning of the nervous control on the HR and the heart's ability to adapt. Since the SNS increases the HR and thus reduces the intervals between the beats, it decreases the HRV. Consequently, when the PNS increases, it raises the intervals between the beats and increases the HRV. Figure 18 shows HR and a tachogram of beat intervals during incremental exercise of a healthy subject. As the intensity of the exercise increases, the effect of increased activity of the SNS causes a decrease in the HRV. This can be seen as a reduced amplitude of the signals.

In the year 1996, an international Task Force published standards for the HRV analytical methods and definitions of the HRV measures [83]. HRV can be assessed with various analytical approaches. The most commonly used are the time domain analysis and the frequency domain analysis, which can also be called power spectral density (PSD) [8].

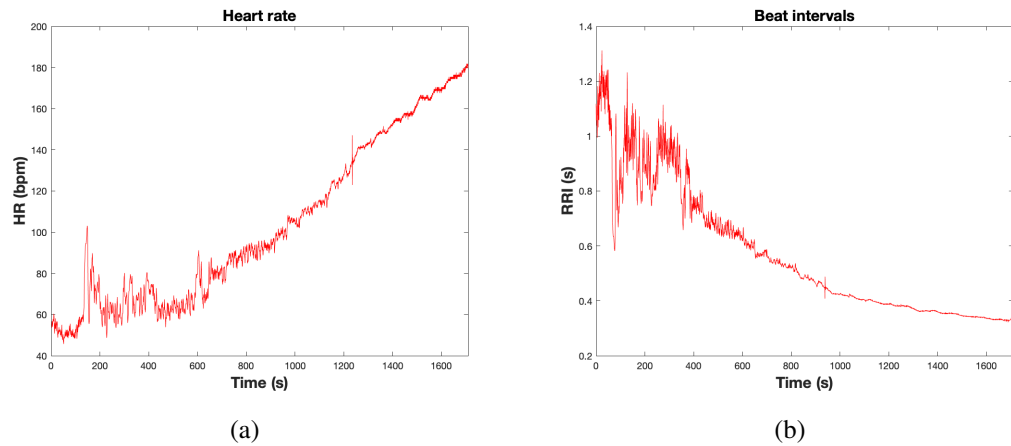


Figure 18. The effect of incremental exercise on beat interval and heart rate of a healthy subject. a) Heart rate (bpm) as a function of time b) Beat interval (s) as a function of time.

Time domain analysis

In the time domain methods, either the heart rate at any point in time or the intervals between normal successive sinus beats (RR intervals) are determined [84]. Some variables can be derived from the obtained values. Those are the mean value of RR intervals, the mean heart rate, and the difference between the longest and the shortest RR interval.

From a series of instantaneous heart rates or cycle intervals, particularly those recorded over longer periods (usually 24 hours), more complex statistical measures can be calculated. These can be divided into two categories: those derived from the direct measurements of the RR intervals or instantaneous heart rate, and those derived from the differences between RR intervals. These variables can be derived from the analysis of the total ECG recording or may be calculated using smaller segments of the recording period. The latter method allows comparison of the HRV to be made during different activities (for instance, rest, sleep, exercise, and so on). [83]

The standard deviation of RR intervals (SDNN) reflects the overall variation within the RR interval series. The root mean square of successive differences (RMSSD) can be used as a measure of the short-term variability. The standard deviation of the average RR intervals calculated over the 5 minute segments (SDANN) is an estimate of the changes in heart rate due to cycles longer than 5 minutes. Table 3 shows the normal values of the main time domain measures during a long-term recording according to the Task Force [83]. [85]

Table 3. Normal values of standard time domain measures of the heart rate variability

Measure	Value (mean \pm SD)	Unit
SDNN	141 \pm 39	ms
SDANN	127 \pm 35	ms
RMSSD	27 \pm 12	ms

Many of the time domain analysis measures are depended on the length of recording period. Thus, it is inappropriate to compare measures obtained from the recordings of different durations. Therefore, the international Task Force has published the standardized recording durations for short-term recordings (5 minutes) and nominal long-term recordings (24 hours) [83].

Frequency domain analysis

The HRV is consisted of four frequency components: high-frequency (HF) component, low-frequency (LF) component, very-low-frequency (VLF) component, and ultra-low-frequency (ULF) component. Parasympathetic nervous activity is seen in the high-frequency (HF) range (0.15-0.4 Hz), and this HF component is mediated almost solely by the PNS [73]. HF band is frequently called the respiratory band because it corresponds to the HR variations related to the respiratory cycle [8]. HR accelerates during inspiration and slows during expiration. These HR changes are known as the respiratory sinus arrhythmia (RSA).

The low-frequency (LF) component (0.04-0.15 Hz) is generally thought of being both of sympathetic and parasympathetic origin [73]. However, it has also been shown that the LF component could be used to assess sympathetic efferent activity, especially with the normalized value [86, 87]. In general, sympathetic effects are slow, on the time scale of seconds, and parasympathetic effects are fast, on the time scale of milliseconds. Therefore the parasympathetic influences are the only ones that can produce rapid changes in the beat to beat timing of the heart [73].

In addition to the HF and LF components, there is a very-low-frequency component (VLF) and an ultra-low-frequency component (ULF). Their frequency ranges are 0.0033-0.04 Hz and below 0.0033 Hz, respectively [8]. The pathological explanation and mechanisms involved in the generation of the VLF component have not been as well studied as the HF and LF components. Long-term regulation systems and the ANS activity related to thermoregulation and renin-angiosystem may contribute the VLF band [8, 88, 89]. However, it has been showed that VLF band has a strong association with mortality in the case of congestive heart failure and multiple organ dysfunction syndrome [90, 91]. In healthy individuals, the VLF power increases during the night and peaks before waking [8]. The ULF band reflects oscillations in the heart rhythm with a period of 5 minutes or greater. It can only be assessed with 24-hour and longer recordings. The circadian oscillation in the heart rate and other very slow regulatory processes are the primary source of the ULF band [8].

Power spectral density (PSD) analysis provides the information of how the power distributes as a function of frequency. The plot of the PSD as function of frequency is called a periodogram. Methods for calculating the PSD can be generally classified as a nonparametric and parametric [83].

In the nonparametric method, usually the fast Fourier transformation (FFT) is utilized, and is characterized with discrete peaks for the frequency components [92]. In the parametric approach, the autoregressive model (AM) estimation is often used. Other common parametric methods are the autoregressive moving average (ARMA) model and the moving average (MA) model. For the above methods, the FFT is a simpler and faster method, while the parametric methods have smoother spectral

components and other qualitative benefits [83]. However, the parametric methods are more complex and need verification of the suitability of the chosen model.

In short-term recordings, three spectral components are distinguished in a spectrum: the VLF, LF and HF components [92]. In long-term recordings, the ULF component is included. The measurement of the components is usually made in absolute values of power expressed as milliseconds squared (ms^2). The LF/HF ratio reflects the total ANS balance. Table 4 shows the normal values of the main frequency domain measures during a stationary supine 5-minute recording according to the Task Force [83].

Table 4. Normal values of standard frequency domain measures of heart rate variability

Measure	Value (mean \pm SD)	Unit
Total power	3466 \pm 1018	ms^2
LF	1170 \pm 416	ms^2
HF	975 \pm 203	ms^2
LF/HF ratio	1.5-2.0	

3.1.2. Morphological Variations

Identification of the normal QRS complex from the P- and T-waves does not create difficulties in normal sinus beats. The QRS complex has a characteristic waveform and dominating amplitude, which is usually about 1 mV in a normal heart [93]. The usual duration of the QRS complex is 0.08-0.09 s [7 p.440]. However, QRS complexes with a duration of 0.1 s can also be classified as normal [30, 94]. If the beat is normal, the P-wave has an amplitude of about 0.1 mV and duration of 0.1 s. For the T-wave, both of these values are about double. The T-wave follows the QRS complex after around 0.2 s.

The morphology of the ECG signal can be influenced by the fluctuations in the HR or artefacts caused by internal or external factors. In general, it can be said that when the HR is increased, the duration of the QRS complex is decreased [95]. The magnitude of the QRS complex is generally greatest for young males and the duration of the complex is shorter for females than for males. In addition, the QRS complex size is known to decrease with age. An important factor related to this study is that when the HR is very high, the T-wave and the successive P-wave from the next activation cycle may sometimes be very close to each other. This comes significant when choosing window lengths for the VCG templates later in this study.

Changes in the HR over time provide morphological information for different types of heartbeats (normal, ectopic, artefact, pacing). If the heart's rhythm changes rapidly, it causes an increase of the HRV and affects the HR signal seen in Figure 18.a. This leads to difficulties in the HR-level detection, which is discussed subsequently in this study. Different efforts, like averaging, beat correction, and outlier detection, can be made to reduce these effects. Those are included in this research and will be discussed later.

Heart's rhythm is regulated by the ANS, but several cardiac dysfunctions can affect the function of the heart. Cardiac rhythms can be divided into supraventricular

and ventricular rhythms. Next, various cardiac rhythms and their impact on the morphology of the ECG signal and HR are explained.

Supraventricular rhythms

The origin of the supraventricular rhythm is in the atria or AV junction [7 p.440]. In these situations, the activation proceeds to the ventricles along the conduction system in a normal way. In the normal sinus rhythm, the cardiac activation is triggered from the SA node [94]. The deflections, P-QRS-T, follow this order and are differentiable. The sinus rhythm can be seen as normal if its frequency is between 60 and 100 bpm.

Sinus bradycardia means a sinus rhythm, which frequency is below 60 bpm. Sinus bradycardia is common in normal individuals during sleep and in those with high PNS activity, such as athletes and young healthy adults [96]. When the sinus rhythm is higher than 100 bpm, it is called sinus tachycardia. It occurs in response to increased activity of the SNS. Usually, this is due to physical exercise or stress, but may also result from congestive heart failure [7 p.441-442].

If the sinus rhythm is considerably irregular, the situation is called sinus arrhythmia. In his book, Malmivuo defines a threshold value of 0.16 s between the longest PP interval or RR interval and the shortest interval [7 p.442-443]. If the threshold value is exceeded, the rhythm is defined as sinus arrhythmia. This arrhythmia is common in all age groups. In fact, according to Malmivuo, this kind of arrhythmia is so common in young people, that it is not even considered as a heart disease. Origin of this arrhythmia might be the vagus nerve, which mediated respiration as well as heart rhythm. This can be seen in the case of RSA.

Atrial contraction origin may be located somewhere else than the SA node. If the location is close to the AV node, the atrial depolarization occurs in the opposite direction than the normal one. As a consequence, the polarity of the P-wave is reversed. The origin of atrial contraction may also vary, in which case the polarity of the P-wave and the PQ interval will also vary. This is referred as wandering pacemaker. [7 p.443-444]

Atrial flutter is a situation when the heart rate is sufficiently elevated so that the isoelectric interval between the T-wave and P-wave is disappeared. The frequency of these fluctuations is around 220 to 300 bpm. The AV node and the ventricles are activated approximately by every second (2:1 heart block) or third atrial impulse (3:1 heart block). [7 p.444]

The activation in the atria can also be irregular and chaotic. In this case, the situation is called atrial fibrillation [97]. Irregular fluctuations appear at the baseline and ventricular rate is rapid and irregular, but usually the QRS contour is normal. Fibrillation waves have a frequency of 350 to 600 bpm. The amplitude can vary and may easily to be mistaken for flutter waves. Atrial fibrillation can occur as a consequence of rheumatic disease, hyperthyroidism, or atherosclerotic disease, but it can also occur in healthy subjects with a strong sympathetic activation [7 p.445].

If the origin of the cardiac rhythm is in the AV node, the heart rate is slow (around 40-55 bpm) and the P-waves are possibly hard to detect. However, the QRS complex is usually normal. This is called junctional rhythm, since the origin is in the junction between the atria and ventricles. The activation of the atria is in the opposite direction

and can occur before, simultaneously, or after the ventricles. If the activations happen at the same time, the P-wave will be superimposed and cannot be seen. [7 p.445]

Figure 19 shows different supraventricular rhythms. The signals are captured from the MIT-BIH Arrhythmia Database, and the presentations of the figures are based on the cardiologists' annotations. Some sub-figures contain a red triangular to demonstrate the location of that rhythm phenomenon.

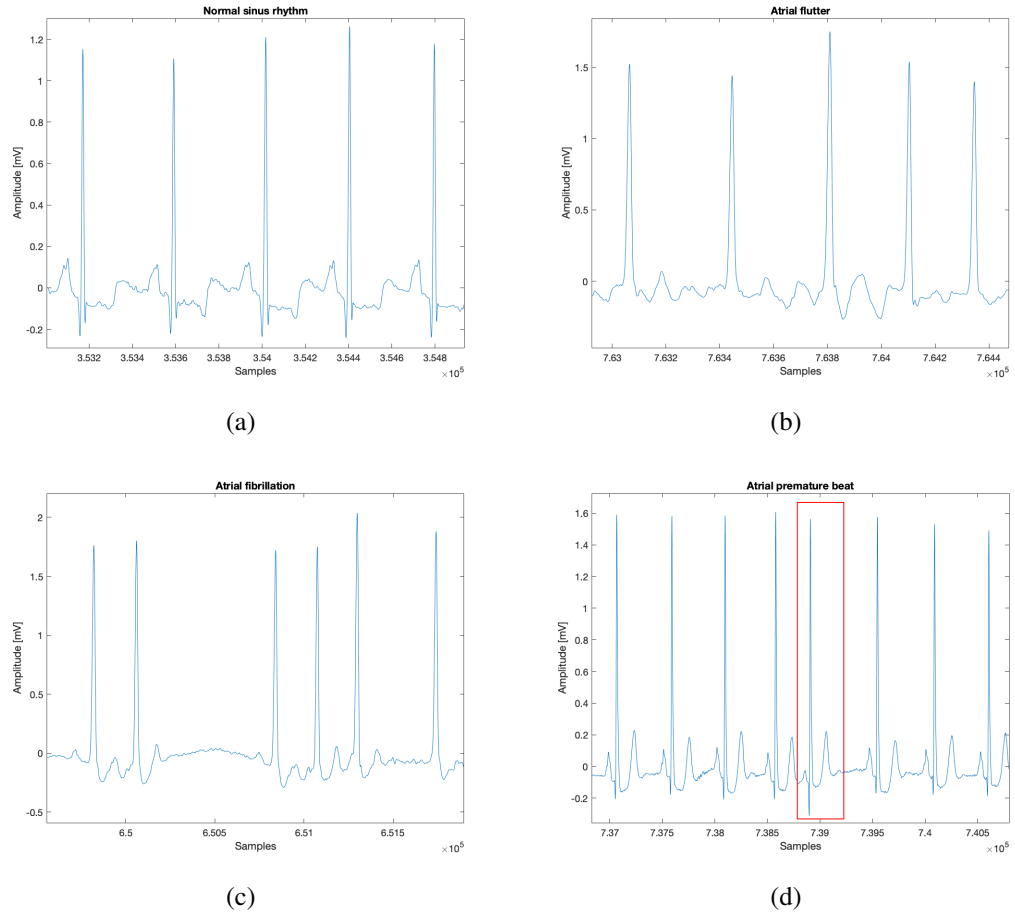


Figure 19. Supraventricular rhythms. a) Normal sinus rhythm b) Atrial flutter c) Atrial fibrillation d) Atrial premature beat. The original signals have been collected from Physionet's MIT-BIH Arrhythmia Database¹³.

Ventricular rhythms

In ventricular rhythms, the ventricular activation does not originate from the AV node and/or does not progress in the ventricles in a normal way. If the ventricular conduction system is broken or the ventricular activation starts far from the AV node, it takes longer time for the activation to progress throughout the ventricular mass. According to Malmivuo, the criterion for normal ventricular activation is a QRS interval shorter than 0.1 s [7 p.446]. If the interval is longer, it indicates abnormal ventricular activation.

In premature ventricular contraction (PVC), the beat occurs abnormally early [7 p.446]. Its origin can be in the atrium or in the AV node. In that case, the origin is

¹³[22, 23]

supraventricular, and the produced complex lasts less than 0.1 s. The origin can also be in the ventricular muscle, in which case the QRS complex has notably abnormal shape and it lasts longer than 0.1 s. Paired PVCs, also called ventricular couplet, is the situation where two PVCs occur in a row [98 p.92-120]. Multiform PVCs is used as a designation for the situation where PVCs look different on one another. If PVCs occur every other beat, phenomenon is called bigeminy, and if every third beat, it is called trigeminy.

A run of three or more consecutive PVCs is termed ventricular tachycardia [98 p.92-120]. Ventricular tachycardia is a result of the activation of the ventricular muscle at a high rate (>120 bpm) [7 p.446-447]. This causes rapid, abnormal, and wide QRS complexes. Ventricular tachycardia is often a consequence of ischemia and myocardial infraction.

If the ventricular depolarization occurs chaotically, the situation is called ventricular fibrillation. In the ECG signal this can be seen as irregular waving without the QRS complexes. The cause of fibrillation is usually involving diseased heart muscle, and the ventricular muscle is unable to pump the blood effectively. The lack of blood circulation causes to almost immediate loss of consciousness and death within minutes. This can be stopped with an external defibrillator pulse and appropriate medication. [7 p.447]

Sometimes an external pacemaker is needed to produce the contraction of the heart. The pacemaker is usually located in the right ventricle and activation does not involve the conduction system. Thus, a ventricular rhythm originating from a cardiac pacemaker is associated with the wide QRS complexes. The ventricular contraction is usually preceded by a clearly visible pacer impulse spike. Generally, the rhythm of the pacemaker is set to 72 bpm. [7 p.447-448]

Different ventricular rhythms can be seen in Figure 20. The original source of the signals is the same than in Figure 19, and the presentations of the figures are based on the cardiologists' annotations.

Activation sequence disorders

In addition to the above, the morphology of the ECG signal can be affected by disorders in the activation sequence. If the P-wave precedes the QRS complex with a PR interval of 0.12-0.2 s, the rhythm is considered as a normal sinus rhythm. If the PR-interval is shorter than normal, the origin of the impulse is closer to the ventricles, or the atrioventricular conduction is utilizing an abnormal tract, which causes pre-excitation of the ventricles. [7 p.448]

The PR interval can also be longer than normal, which is usually caused by some conduction block in the conduction system. If the block is located in the AV node, it is called atrioventricular block (A-V block) [7 p.448]. These can be further divided into the first-, second-, and third-degree A-V blocks. When the P-wave precedes the QRS complex, but the PR-interval is longer than 0.2 s, the situation is called first-degree atrioventricular block [96]. If the PR interval is longer than normal, and the QRS complex is not always following the P-wave, the atrioventricular block is second-degree. The PR interval can increase progressively, leading finally to the dropout of the QRS complex. In these situations, the phenomenon is called a Wenkenbach phenomenon. The third-degree A-V block is characterized by a complete lack of

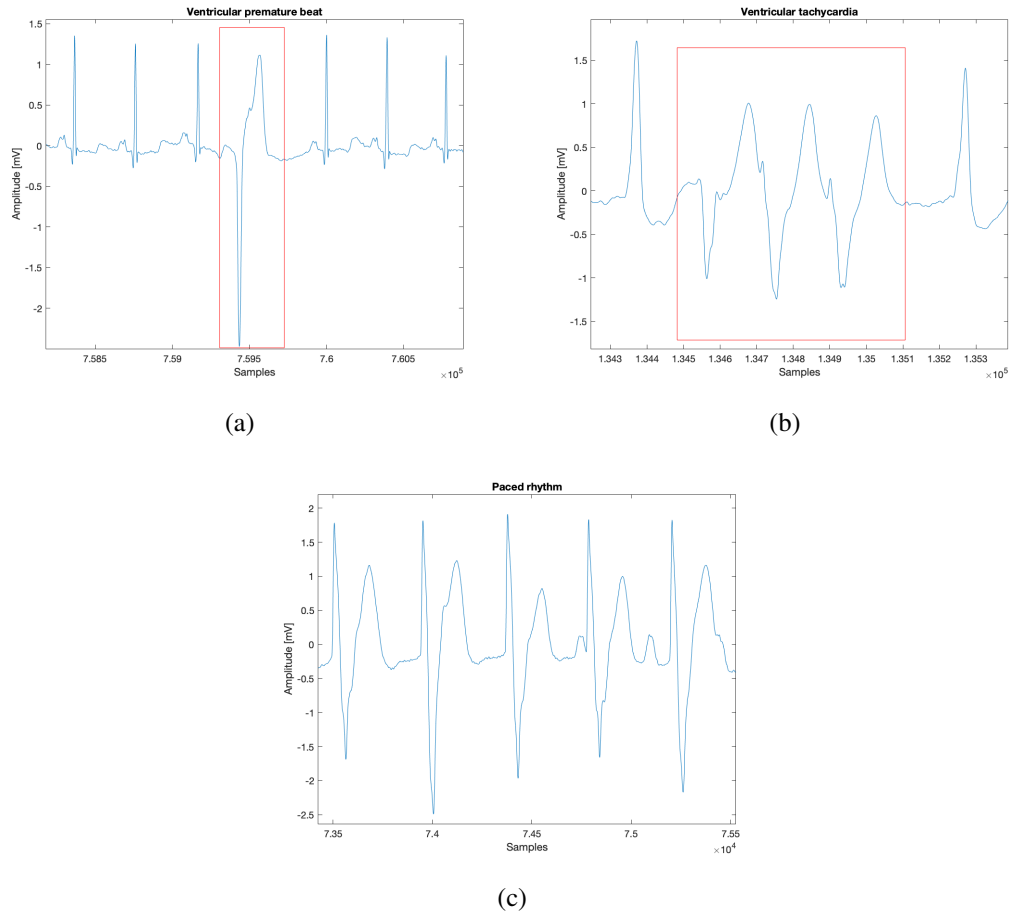


Figure 20. Ventricular rhythms. a) Ventricular premature beat b) Ventricular tachycardia c) Paced rhythm. The original signals have been collected from Physionet's MIT-BIH Arrhythmia Database¹⁵.

synchrony between the P-wave and the QRS complex. This can also be referred as a total block.

Bundle-branch block denotes conduction errors in the bundle branches. If the two bundle branches are blocked simultaneously, the progress of activation from the atria to the ventricles is completely inhibited. This is referred as the third-degree atrioventricular block, as discussed above. If the left or right bundle branch is blocked, the activation of the ventricle must await the activation by the opposite ventricle. This results much slower activation propagation, and the result is abnormally shaped and long QRS complexes. [7 p.450]

In the right bundle-branch block (RBBB), the electrical impulse cannot travel to the right ventricle via the bundle branch, and the activation proceeds to the right ventricle via the left ventricle. This propagation is much slower than through the normal conduction system, resulting in a QRS complex wider than 0.12 s [96]. Often the activation of the right ventricle is so much delayed, that it can be seen following the activation of the left ventricle. The RBBB is seen in the ECG signal as a broad S-wave or as a double R-wave, which is also named as RSR' complex [7 p.450-451].

¹⁵[22, 23]

The left bundle-branch block (LBBB) is similar to RBBB, but the activation propagation is opposite. Since the activation front is about to normal direction in the LBBB, the polarities of the signal are normal. Nonetheless, the activation point of the left ventricle occurs in a different location than normal, and this leads to a broad and tall R-wave in the signal. The duration criterion for the QRS complex in the LBBB is the same as in the RBBB (> 0.12 s). [7 p.451-452]

The RBBB and LBBB beats are represented in Figure 21. The same annotation data has been utilized as in the previous two figures.

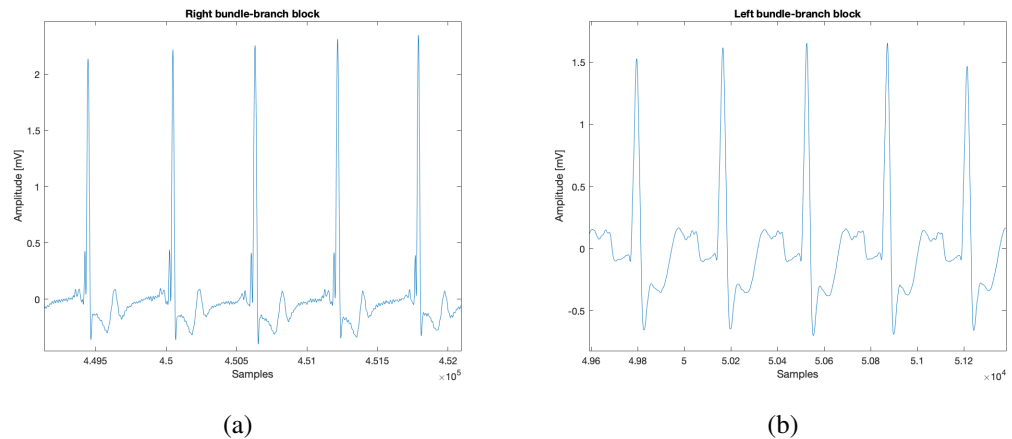


Figure 21. Bundle-branch block beats. a) Right bundle-branch block b) Left bundle-branch block. The original signals have been collected from Physionet's MIT-BIH Arrhythmia Database¹⁷.

Effects of physical stress and myocardial ischemia and infarction

Cardiac muscles enlarge with an increased amount of physical stress. The extra tension may be as a result of increased pressure load or volume load. Pressure overload is due to increased resistance in the outflow of the particular section. Volume overload is caused by the inefficient function of either the outflow or the inflow valve. The increase of the atrial or ventricular size is called atrial or ventricular enlargement. The increase of the wall thickness of either compartment is called hypertrophy of that area. Usually, both of these are referred as hypertrophy. Particularly in the ventricular hypertrophy, a tall S- or R-wave can be seen in the signal. [7 p.453-456]

If the blood flow of the coronary artery is disturbed, the transport of oxygen to the cardiac muscle is decreased. This causes an oxygen deficiency in the muscle, and is called ischemia. Ischemia affects the repolarization of the muscle cells, and is seen as changes in the T-wave. If the oxygen transport is terminated in a certain area, the heart muscle dies in that area. This is called infarction. The infarct area is electrically silent. [7 p.456-457]

¹⁷[22, 23]

3.2. Signal Preprocessing

ECG signal preprocessing includes several steps. This contains signal acquisition, data transformation and several filtering procedures. Subsequent signal processing steps include feature extraction (or QRS detection), which is further discussed in Section 3.3. The quality of the ECG beat detection and classification depends on the preprocessing stage [99]. Therefore, it is important to perform signal processing correctly and efficiently.

3.2.1. Sampling Rate Selection

Sampling rate (or sampling frequency, f_s) refers the number of samples taken per second (Hz). The correct sampling rate is crucial to avoid aliasing. The basic concept of choosing the sampling rate is that minimum sampling rate (also called Nyquist rate) must be twice as much as the highest frequency component of the signal. Nyquist frequency is half of the sampling frequency and it corresponds to the highest frequency which a sampled data can be reproduced without error.[100 p.81-83]

The selected sampling frequency must also meet the requirements of the filter specifications. The cut-off frequencies (f_c) of the filters set limitations for the sampling frequency. According to the Nyquist theorem, the sampling frequency must be twice as high than the desired high-frequency cut-off. However, the recommended sampling frequencies are set to be at 2 or 3 times the theoretical minimum [101]. For instance, a sampling frequency of at least 500 Hz is required for the 150 Hz cut-off. This sets the Nyquist frequency at 250 Hz.

Since a modern ECG device is a digital system, the analog signal is directly converted into a digital form. This is called analog-to-digital conversion (A/D). Oversampling technique is often used in the ECG devices, which refers to an initial A/D conversion sampling rate. Oversampling frequency is usually from 1000 Hz to 2000 Hz, but newest converters can sample at 10 000 Hz to 15 000 Hz [1]. This sampling rate is many times higher than the target final data resolution sampling rate, which is used further signal processing steps of the ECG signal. To convert an oversampled signal to a lower sampling rate, a decimation process is used, which includes low-pass filtering and resampling (downsampling) the resulting smoothed signal. [102]

3.2.2. Filter Selection

Filtering is used to attenuate or remove desired frequencies from the signal. An ideal filter removes the noise without changing the actual signal. In the real life scenario, desired signals and noise are often overlapping in the frequency range. This means that when the filter attenuates the frequency components of the noise, it can also affect the desired signal and induce magnitude distortion. Additionally to the magnitude response, filters have also a feature called the phase response. If filter's phase response is not a linear function of frequency, the phase distortion occurs. Filter's phase distortion can generate substantial influence on data accuracy. Digital filters can be

designed to have linear or zero-phase characteristics, and this avoids some of the distortion caused by the analog filters. [102]

In general, low-pass filters (LPF) are used to remove high-frequency noise and high-pass filters (HPF) low-frequency noise, respectively. The cut-off frequency (f_c) is the frequency where 3-decibel (3-dB) attenuation occurs. The cut is not perfect, and the filter has so called transition band from the passband to the stopband. Filter specification recommendations vary depending on whether the filter is analog or digital [1]. Principally, filter design is a tradeoff between magnitude distortion reduction and noise cancellation, while trying to minimizing the phase distortion. [102]

The low-frequency artefacts, such as baseline wandering and respiration, are removed by the HPF, but it can also change the important morphology of the ECG signal and generate time delay. The low-frequency cut-off is recommended to be 0.5 Hz for analog filter, and 0.67 Hz for linear digital filters with zero phase distortion [1]. The frequency of 0.67 Hz corresponds to a heart rate of 40 bpm. Heart rates below 40 bpm are uncommon in practice. Figure 22 shows removal of the baseline wandering introduced in Figure 11 using HPF. [102]

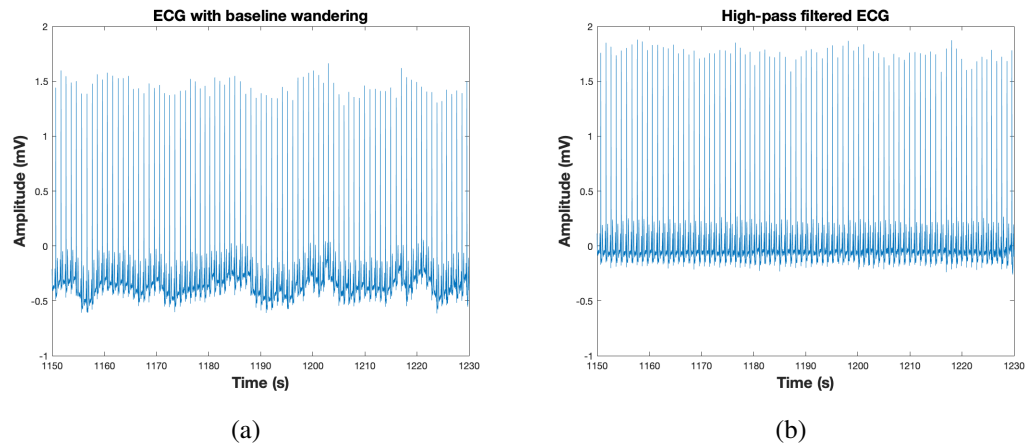


Figure 22. Correction of the baseline wandering. a) The distorted signal, which was introduced in Section 2.3.3. b) The high-pass filtered signal, where the baseline wandering is clearly attenuated.

The high-frequency artefacts (like EMG noise) are filtered by the LPF. Insufficient low-pass filtering can reduce the amplitude of the QRS complex and so impair the ability to detect small deviations. The high-frequency cut-off is recommended to be at least 150 Hz to reduce amplitude error measurements to around 1 % in adults [1]. An upper-frequency cut-off of 250 Hz is more suitable for infants. [102]

Some distortions do not fall within the limits of the HPF's low-frequency and the LPF's high-frequency cut-off in their frequency domain. One of this kind distortion is powerline interference, which occurs usually in the frequency band of 50-60 Hz. To be able to filter this kind of noise, a notch filter can be utilized. The notch filter, also called the line-frequency filter (LFF), is a bandstop filter, which have a low-frequency cut-off and a high-frequency cut-off, and its stopband is between this range. It is usually activated already during the ECG acquisition, but it can also be used in later signal processing stages. However, the LFF can induce ringing artefacts or so called oscillatory behavior, after the QRS complexes. The principle behind this is that if the

bandstop is steep in the frequency domain, ringing appears near the sharp edge of time domain. [102]

In addition to the above, various smoothing filters exist, which use different mathematical methods to smooth the data. They are, for instance, the median filter, the Savitzky-Golay filter [103], and the moving average filter. Smoothing the ECG data makes it more legible. The smoothing process changes the data points of the signal in such way that the noises are reduced, and the points that are lower than their neighboring points are increased. In practice, the result is amplified low-frequencies and attenuated high-frequencies. Figure 23 visualizes the effect of smoothing on the signal using the Savitzky-Golay filter. [104]

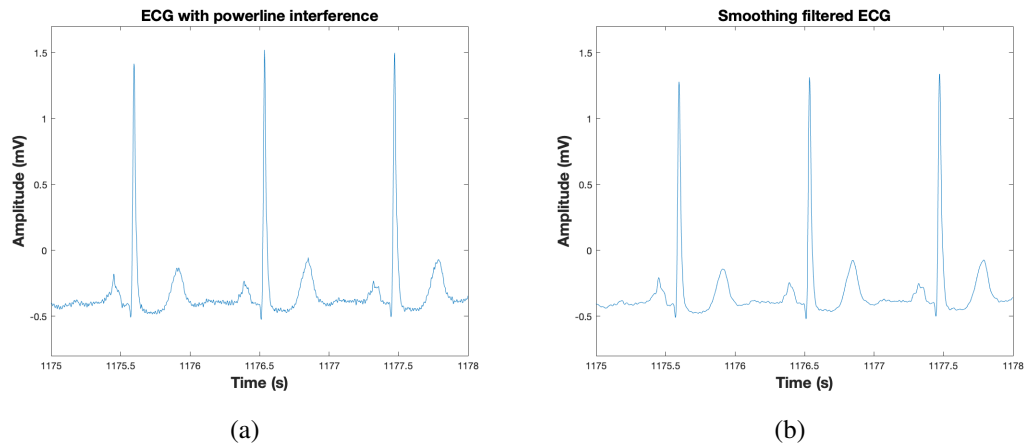


Figure 23. The effect of the smoothing filter. a) The noisy signal, which was introduced in Section 2.3.3. b) The smoothed signal with reduced noise.

3.3. QRS Detection

The heart beat period is commonly evaluated as the time difference between the detectable QRS complexes. A typical QRS detector consists of a preprocessing part followed by a decision rule. However, it is a challenging task to implement the algorithm to identify the QRS complexes since the QRS complexes are varying with the physiological variability and affected easily by the various noise sources. Therefore, several different QRS detectors have been proposed within last decades.

Benitez and his colleagues [105] presented the use of the first differential of the ECG signal and its Hilbert transformed data to locate the R-wave peaks in the ECG waveform. They used adaptive threshold detection to locate the peaks, which was based on the average RRI of the previous R-peaks located. The decision was made in base of the amplitude of the peak and its position in relation to the previous R-peak. They then used the second stage detector which utilized the information provided by the first detector. A window subset was used from the location of the detected R-peak to locate the actual point of the real peak. This was done by a simple amplitude comparison between the values of the subset.

Afonso with his associates [106] designed a signal processing algorithm based on the filter bank strategy. The algorithm incorporated a filter bank which decomposed

the ECG into sub-bands with uniform frequency bandwidths. The filter bank -based algorithm enabled independent time and frequency analysis. Features computed from a set of sub-bands and a heuristic detection strategy were used to fuse the decisions from multiple one-channel beat detection algorithms. Three different thresholds and timing information was used in detection. A threshold was used to decide if an incoming peak is due to the QRS complex or noise. If a period of time corresponding to the average heartbeat interval elapsed without a beat detection, the "search-bank" strategy was used to check the ECG again for the presence of a beat.

The morphological approaches, such as slope, amplitude, and width by derivative and squaring or integration processing were used by Pan and Tompkins [107], Hamilton and Tompkins [108], and Paoletti and Marchesi [109]. Pan and Tompkins developed an algorithm, which had four adaptive thresholds. The average RRI information was utilized for correction decisions in situations where the algorithm had not detected a possible peak. Hamilton and Tompkins further developed this algorithm and added more complex decision rules. Paoletti and Marchesi used four adaptive thresholds and their characterization was based on the clustering method called K-harmonic means. According to their paper, the algorithm was especially designed for noisy applications.

Christov [110] proposed a real-time detection method based on the moving average filter and adaptive threshold. The threshold was combined by three parameters: an adaptive steep-slope, an integrating value which rises when HF noise occurs, and a beat expectation value. Two algorithms were developed. The first one detected the current beat, and the second one performed the pseudo-real-time detection with additional triggering of potentially missed beats in the last interval by RRI analyses.

Sequential filtering method such as wavelet denoising, moving average filter based high-pass filter and low-pass filter was introduced by Chen et al. [111]. They applied adaptive threshold for decision-making to detect QRS complexes. Poli et al. [112] suggested the polynomial filter approach. Their algorithm consist of three modules, which were linear and nonlinear polynomial filter, and maxima detector. The maxima detector detected the maxima of the filter output and only amplitudes higher than the threshold were accepted. Their algorithm was genetic. Genetic algorithms are optimization and search procedures inspired by genetics and the process of natural selection in which the individuals of a population which are fitter with respect to the environment tend to survive and reproduce longer than the others.

Lee et al. [113] and Cvikl et al. [114] processed the ECG signals by phase portraits based on delay-coordinate mapping method. Lee and his associates used two multi-adaptive thresholds with refractory blanking rule and search back rule for the detector stage. Refractory blanking rule stated that any QRS complex within 200 ms followed by the detected QRS complex is ignored due to the physiological constraint imposed by a refractory period on a cardiac muscle. Search bank rule was that if any QRS detection occurred within 150% of the latest mean RRI, a new QRS search is attempted by applyin 80% and 50% of the previous threshold values. Cvikli and his research team applied a rule-based decision system. They also used refractory blanking rule and search back rule.

Choi et al. [115] designed ECG segmentation method using an irregular RRI checkup strategy into five sequential RRI patterns. Their algorithm, as usually all the algorithms designed for this purpose, was comprised of signal processing stage and

beat detection stage. The beat detection included the candidate ECG beat detection and segmentation by one threshold and by irregular RRI checkup strategy, respectively.

Pal and Mitra [116] presented a method based on empirical mode decomposition for detection of the QRS complexes in the ECG. They introduced empirical mode decomposition as a fully adaptive signal decomposition technique that generates Intrinsic Mode Functions as a decomposition output. They selected a set of Intrinsic Mode Functions that represented the QRS region and a nonlinear transformation was done for the QRS enhancement. According to Pal and Mirta, this method improved beat detection accuracy.

3.4. Clustering

Clustering aims to divide data sets into subsets (clusters), where objects in the same subset are similar to each other with a respect to a given similarity measure. On the contrary, objects in different clusters are dissimilar. Cluster analysis has been used in a large variety of fields, and various cluster analysis methods have been developed. In general, a clustering method is a multivariate statistical procedure that starts with a data set containing information about a samples and attempts to reorganize these samples into relatively homogeneous groups. Many medical or physiological analyses involve partitioning samples or variables into clusters.

Clustering quality can be checked using the within-cluster similarity and between-cluster similarity [117]. Within-cluster similarity is ideally high, whereas between-cluster similarity is ideally low. Sometimes it is better to plot the objects in order to determine how variable the objects are and whether there are clusters that have very similar profiles and therefore should be merged. Also exploration of the cluster quality as a function of the number of clusters is sometimes used.

Clustering methods

Clustering algorithms can be broadly divided into the hierarchical clustering and partitional clustering. The hierarchical clustering recursively find nested clusters either in the agglomerative mode or in the divisive mode. In the agglomerative mode, algorithm starts with each data point in its own cluster and merging the most similar pair of clusters successively to form a cluster hierarchy. In the case of the divisive mode, the algorithm starts with all the data points in one cluster and recursively dividing each cluster into smaller clusters. The most used hierarchical methods are the single-link and complete-link. Figure 24 demonstrates the complete linkage clustering of 5 objects. The results of the hierarchical clustering are usually introduced by a binary tree or dendrogram. [118]

In their book [119 p.31-62], Xu and Wunsch comprehensively describe the function of the hierarchical algorithms. The advantage of these methods is that they are conceptually easy to understand. The common criticism of the hierarchical clustering methods is a high computational complexity, which limits their use in the large-scale data sets. Another problem of these methods is that they make only one pass through the data, which means that once an object is assigned to a cluster, it will not be considered again. This leads to the fact that hierarchical clustering methods are not

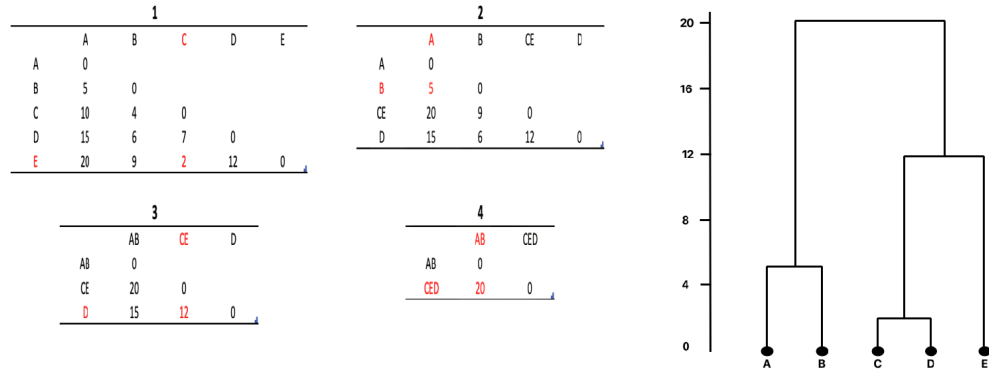


Figure 24. The demonstration of the complete linkage clustering of 5 objects. On the left (tables 1-4), pairwise maximum distances are used to merge objects. At each merging step (1-4), the shortest distance is selected (marked in red color). On the right, a dendrogram with a vertical axis indicating the distance between merged objects is showed. To create clusters, one can cut the tree at a fixed height.

capable of reacting to a possible previous misclassification. On the other hand, this has motivated the development of new, more efficient algorithms that have shown a great potential.

Partitional clustering directly divides data points into some predetermined number of clusters without the hierarchical structure. This procedure usually involves the optimization of a prespecified criterion function. The most popular partitional clustering method is the K-means, which is one of the best-known and most popular clustering algorithms. [119 p.63-67]

The K-means clustering is an iterative clustering method, which usually uses Euclidean or correlation distance [117]. It is known that the k-mean of a distribution can numerically be calculated by the Lloyd's iterative algorithm published by Lloyd [120]. The K-means algorithm starts by selecting the number of clusters (K). Then the algorithm selects K data points as the 'centroids'. These data points can be randomly generated within the range of the data, or some mathematical operations can be used for the cluster center initialization. In the first step, the similarity between each point and each centroid is computed, utilizing the selected measure method. Usually, the squared error between the mean μ_k of the cluster c_k and the points is defined as in Equation (7):

$$J(c_k) = \sum_{x_i \in c_k} ||x_i - \mu_k||^2 \quad (7)$$

where x_i represents the data points, $k = (1, \dots, K)$, and K is the number of clusters [118].

The points are then grouped with the nearest centroid. This follows the nearest-neighbor rule and therefore is the so-called Voronoi partition [119 p.68]. Subsequently, the position of each centroid is recalculated on the basis of the objects assigned to it, and the object assignment is repeated with the new centroid positions. These steps are

repeated until the centroids and clusters no longer change. Consequently, the goal of the K-means is to minimize the sum of the squared error over all K clusters computed like in Equation (8):

$$J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} ||x_i - u_k||^2 \quad (8)$$

where C represents all the clusters $C = \{c_k, k = 1, \dots, K\}$ [118]. Figure 25 visualizes the K-means algorithm technique.

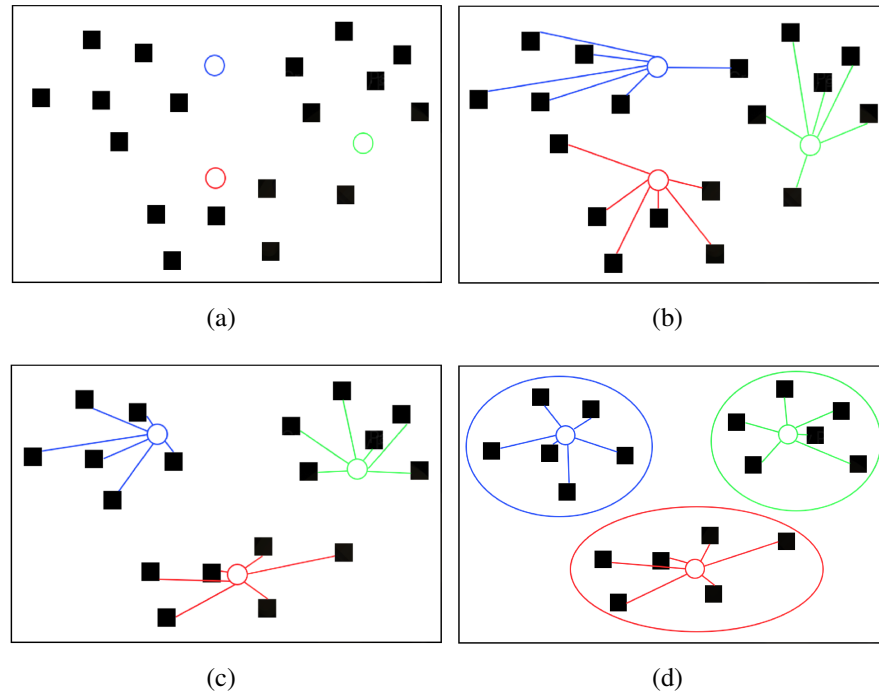


Figure 25. The K-means clustering visualization. Centroids are marked as circles and data points as squares. The number of clusters is 3. a) Starting point, initial centroids are randomly selected. b) The points are connected with the nearest centroid. c) The position of each centroid is recalculated and the step b) is repeated. d) The final location of the centroids is achieved and clusters are formed. Each cluster is indicated by the corresponding centroid color. In a real simulation, steps b) and c) are repeated multiple times before the step d) is achieved.

The K-means algorithm requires three specified parameters: the number of clusters K , cluster initialization, and criteria metric. The most critical is the K . This brings a limitation to the K-means technique since sometimes the number of clusters is unknown for some clustering practices, and estimating it becomes a problem for cluster validation [119 p.110]. In addition, different initializations can lead to different outcome since the K-means only converges to the local minima [118]. The basic K-means algorithm has been extended in many ways. Xu and Wunsch discuss different partitional clustering methods in their book [119 p.63-110].

Benabdellah et al. [121] compared clustering algorithms in their survey. The Self-Organization Map was the best performed algorithm, but the performance was decreased as the number of clusters increased. The hierarchical clustering was the

most sensitive to noisy dataset. Also, the K-means reached a higher performance for big data than several other algorithms. They made a conclusion, that none of the algorithms performed well for all the validity measures.

Yin et al. [122] tested the K-means algorithm along the Fuzzy C-means algorithm for automated determination of the arterial input function. In their study, the K-means performed more accurate and the results were more reproducible. However, the execution time was longer in the K-means than in the Fuzzy C-means.

Comparison between clustering algorithms is not a simple task. Each algorithm can be thought to have its own pros and cons. Although some comparisons are possible, the results depends highly on the specific scenario. These comparisons are mostly based on some specific applications, under certain conditions, and the results may become different if the conditions change. In general, it can be said that there is no clustering algorithm that can be universally used to solve all problems. A comprehensive review of clustering algorithms is offered by Xu and Wunsch [123], Ezugwu et al. [124], and Xu and Tian [125].

4. METHODS

The goal of this study was to develop heart rate dependent electrocardiogram templates that can be used in a body emulator environment. This contained the following areas:

1. Real-life ECG signal preprocessing
2. Development of a beat detection algorithm
3. HR-level clustering based on the detected beats
4. VCG template extraction for each HR-level

The goal was to develop templates for various heart rates. In this manner, these different templates could be used in the emulator to produce ECG signals for different heart rates. The extracted VCG signals represent a specific HR-level, and these VCG signals could be used as an input to the dipoles attached to the emulator. In addition to this, the aim was to develop a graphical user interface (GUI) for the above processes.

It must be noted, that because the intended use of the models developed in this study, the developed models are designed to work with already acquired ECG data. Therefore, the developed algorithms are not so called live algorithms, that can operate with the live ECG data. This is also relevant to the selections of the methods used in the following sections.

Each area and the techniques included in them are presented in this chapter in their own sections. The implementation of the work was performed with the MATLAB¹⁸ software (MATLAB_R2022b). The GUI was developed using the MATLAB App Designer.

4.1. Preprocessing Methods

A high-pass Butterworth infinite impulse response (IIR) digital filter was designed to reduce baseline wandering and other low-frequency noise in the signals. The filter order was selected at 12 and the normalized cut-off frequency at 0.00268. The normalized cut-off frequency was derived by taking the recommended low-frequency cut-off (0.67 Hz) [1] and dividing it by the Nyquist frequency like shown in Equation (9). This was done since the filter design was implemented by the MATLAB built in code¹⁹, and the normalized cut-off frequency was required. The sampling frequency f_s was 500 Hz, so the Nyquist frequency was 250 Hz.

$$f_c \text{ normalized} = \frac{0,67Hz}{f_s/2} = \frac{0,67Hz}{250Hz} \quad (9)$$

In addition, the zero phase distortion method was used in the design²⁰. In this way, the filtering was performed by processing the data in both the forward and reverse directions. Zero phase filtering helps to preserve features in a filtered time waveform exactly where they occur in the unfiltered signal. The transfer function $H(z)$ of the IIR filter is determined as in Equation (10):

¹⁸MATLAB - <https://www.mathworks.com/products/matlab.html>

¹⁹MATLAB designfilt function - <https://se.mathworks.com/help/signal/ref/designfilt.html>

²⁰MATLAB filtfilt function - <https://se.mathworks.com/help/signal/ref/filtfilt.html>

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^P b_i z^{-i}}{\sum_{j=0}^Q a_j z^{-j}} \quad (10)$$

where $Y(z)$ and $X(z)$ are the Z-transforms of the output and input, respectively, P is the feedforward filter order, Q is the feedback filter order, and b and a are the filter coefficients. The designed filter coefficients expressed as a second-order section are displayed in Table 5. In the second-order section, each row of the matrix is of the form $[b_0, b_1, b_2, a_0, a_1, a_2]$.

Table 5. The filter coefficients expressed as a second-order section

	1	2	3	4	5	6
1	0.9989	-1.9978	0.9989	1	-1.9977	0.9978
2	0.9968	-1.9935	0.9968	1	-1.9935	0.9936
3	0.9949	-1.9898	0.9949	1	-1.9897	0.9898
4	0.9933	-1.9867	0.9933	1	-1.9867	0.9867
5	0.9923	-1.9845	0.9923	1	-1.9845	0.9846
6	0.9917	-1.9834	0.9917	1	-1.9834	0.9834

With the matrix, the transfer function equation can be expressed as in Equation (11). For the clarity, only the first two rows in the matrix are displayed in the equation.

$$H(z) = \left(\frac{0.9989 - 1.9978z^{-1} + 0.9989z^{-2}}{1 - 1.9977z^{-1} + 0.9978z^{-2}} \right) \left(\frac{0.9968 - 1.9935z^{-1} + 0.9968z^{-2}}{1 - 1.9935z^{-1} + 0.9936z^{-2}} \right) \cdots \quad (11)$$

After the HPF, the signals were smoothed by the Savitzky-Golay finite impulse response (FIR) filter²¹. The Savitzky-Golay filter smooths out noise in the signal but preserves the high frequency signal components. In general, Savitzky-Golay filter consists of replacing each point of the signal by least-squares fitting an nth-order polynomial through the signal values in the window frame and taking calculated central point of the fitted polynomial curve as the new smoothed data point. The polynomial order was set to 3 and the frame length to 15.

4.2. Beat Detection Algorithm

One of the main focus area of this study was the development of a beat detection algorithm. The developed algorithm detects the R-peaks from the ECG signal and classifies possible beat artefacts. Different morphological variations of the ECG signal and their physiological background were discussed in Section 3.1.2. In addition, the beat detection algorithm can identify incorrectly detected R-peaks. Once the algorithm has classified morphological deviations and false detections, it makes the

²¹MATLAB sgolayfilt function - <https://se.mathworks.com/help/signal/ref/sgolayfilt.html>

right corrections to them. The goal was an ECG signal with corrected R-peak detections and smooth RR intervals without any distortions. In this manner the generated VCG signals correspond their respective HR-levels as closely as possible.

The architecture of the algorithm was inspired by Lipponen and Tarvainen [126]. Their novel algorithm performed well in standard database tests and therefore a similar approach was desired.

Next, the functionality of the developed beat detection algorithm is presented. The steps of the algorithm are briefly presented in Figure 26. The filtered ECG signal is the input of the algorithm. The output is the same signal with detected (and corrected) beats and calculated heart rates and RR intervals.



Figure 26. The overview of the algorithm.

Each subsection of this section contains details on the specified function listed above. The R-peak detection method is presented in Section 4.2.1, the decision algorithm and artefact classification are introduced in Section 4.2.2, and the artefact correction is showed in Section 4.2.3. The test results of the algorithm are discussed in Chapter 5.

4.2.1. R-Peak Detection

R-peaks were detected using the MATLAB `findpeaks` function²², which finds the local maxima in the given signal. Its properties can be modified by setting different parameters. The parameters used in the detection were the minimum peak height, the minimum peak distance, the reference height of width measurements, the maximum peak width, and the minimum peak prominence. The function returns the amplitudes of the peaks and their respective locations.

Minimum peak height determines how much the peak amplitude must be to be classified as an R-peak. Only peaks that have higher amplitude than the specified value are detected. Specifying the minimum peak height reduces the incorrect detection of tall T-waves and other disturbances, and lowers the processing time. The downside is that if the limit is set too high, low-amplitude R-peaks are not detected.

Minimum peak distance affects how close the acceptable R-peaks can be to each other. The system chooses the tallest peak in the signal and ignores all other peaks within this range. The function repeats the process for the longest remaining peak and iterates until it runs out of the peaks to consider. Like the minimum peak height, this parameter also reduces the change of false detection of non-R-waves. Large disturbances in the signal obviously affect the behaviour of this parameter. If there

²²MATLAB `findpeaks` function - <https://se.mathworks.com/help/signal/ref/findpeaks.html>

is a large spike in the signal, for example due to the movement of subject, the function can incorrectly detect the large spike and ignore the possible real R-wave close to it.

Reference height of width measurements defines the reference line from which the peak width is calculated. The reference method was selected as 'half height', which positions the reference line at one-half the peak height. This also has the effect of discarding peaks with height less than zero. Maximum peak width specifies the maximum peak width that is acceptable. Only those peaks that have the width at most of this value are selected. Practically, this parameter improves the accuracy of the detection, because the P- and T-waves have a longer duration.

Minimum peak prominence modifies the detection in a way that only the peaks that have a relative importance of at least this value are accepted. The prominence of a peak measures how much the peak stands out due to its intrinsic height and its location relative to the other peaks. Setting the minimum peak prominence excludes small peaks that are close to the bigger ones. This is desired in most of the cases, where for example an unusually high P- or T-wave is present. However, this makes the detection more vulnerable for the tall spikes caused by some distortion.

4.2.2. Artefact Classification

The algorithm uses the RRI information obtained from the initial beat detection. The function of the artefact classification is based on two different time series (dRRs and mRRs) and two time-varying adaptive thresholds (Th1 and Th2). The dRRs and mRRs series are defined as in Equation (12) and Equation (13), respectively:

$$dRRs(i) = RR(i) - RR(i - 1), \quad i = 2 \dots N \quad (12)$$

$$mRRs(i) = RR(i) - \text{median}[RR(i - 5 \dots i + 5)], \quad i = 1 \dots N \quad (13)$$

where N is the number of RR intervals and RR means the RR interval. The dRRs series is computed as the difference between two successive RR intervals (RR). Since the first interval is $R(2) - R(1)$, the first element of the dRRs series is $dRRs(1) = 0$. The mRRs series is defined as the difference between the individual RR intervals and an 11-beat median RR interval.

In the median calculation, some modifications had to be made in the case of data edges. The 11th order median filter was utilized via the `medfilt1` function²³. The function considers the signal to be 0 beyond the endpoints. Therefore, the padding parameter was selected to be 'truncate', which enabled the median calculation with smaller segments when approaching the edges.

The first threshold (Th1) was estimated from the dRRs series in order to adopt to different HRV levels. The threshold is computed as α times quartile deviation (QD) of the 91 surrounding beat interval differences as shown in Equation (14):

$$Th1(i) = \alpha QD[|dRRs(i - 45 \dots i + 45)|], \quad i = 1 \dots N \quad (14)$$

²³MATLAB `medfilt1` function - <https://www.mathworks.com/help/signal/ref/medfilt1.html>

where α is a scaling factor, which was selected to be 5.2 according to Lipponen and Tarvainen [126]. When $i < 46$ or $i > N - 45$, the equation was customized to perform the calculations with smaller segments. In order to define the QD, the interquartile range was calculated. The interquartile range is the difference between the 75th and 25th percentiles of the data. This was done by using the `iqr` function²⁴. The QD was derived by dividing the interquartile range by 2.

Once the Th1 was defined, the dRRs series were normalized with the Th1 like presented in Equation (15):

$$dRR(i) = \frac{dRRs(i)}{Th1}, \quad i = 1 \dots N \quad (15)$$

Each of the artefact beats had to fulfill the statement shown in Equation (16):

$$|dRR(i)| > 1 \quad (16)$$

Based on the estimate that the extra beat would divide the normal beat interval in a half: for the heart rate of 60 bpm, the mRRs is close to -0.5 s for extra beats and close to 1 s for missed beats. To be able to provide equal threshold for the extra and missed beats, the mRRs series were scaled as shown in Equation (17):

$$mRRs(i) = \begin{cases} 2mRRs(i), & \text{if } mRRs(i) < 0 \\ mRRs(i), & \text{if } mRRs(i) \geq 0 \end{cases}, \quad i = 1 \dots N \quad (17)$$

This set the mRRs close to 0 s for normal beats, close to 1 s for missed beats and close to -1 s for extra beats.

The second threshold (Th2) was computed from the mRRs series. This enabled adaptation to different HRV levels. Equation (18) presents the calculation of the Th2. The threshold is defined as α times the quartile deviation (QD) of the 91 surrounding beat differences.

$$Th2(i) = \alpha QD[|mRRs(i - 45 \dots i + 45)|], \quad i = 1 \dots N \quad (18)$$

As in the Th1 computation, equation customization was performed. The mRRs series were normalized with the Th2 like introduced in Equation (19):

$$mRR(i) = \frac{mRRs(i)}{Th2}, \quad i = 1 \dots N \quad (19)$$

The series and thresholds calculated above were used to identify artefact beats. Each artefact classification protocol is explained below.

Ectopic beats

Ectopic beats are consisted of premature atrial contraction and premature ventricular contraction. The algorithm differentiates ectopic beats from the other artefact beats and normal beats. Detection errors are not classified as ectopic beats.

²⁴MATLAB `iqr` function - <https://www.mathworks.com/help/matlab/ref/iqr.html>

Ectopic beats cause the negative-positive-negative and positive-negative-positive segments to the dRR series. This pattern was utilized for the classification. Equation (20) and Equation (21) show the formation of the two-dimensional space S1:

$$S11(i) = dRR(i), \quad i = 1 \dots N \quad (20)$$

$$S12(i) = \begin{cases} \max[dRR(i-1), dRR(i+1)], & \text{if } dRR(i) \geq 0 \\ \min[dRR(i-1), dRR(i+1)], & \text{if } dRR(i) < 0 \end{cases} \quad (21)$$

$S12$ and $-S11$ gets simultaneously large values for the segments of negative-positive-negative and small values for the segments of positive-negative-positive. Based on this knowledge, Equation (22) was build to classify ectopic beats:

$$Ectopic \ beat = \begin{cases} S11(i) > 1 \ \& \ S12(i) < -c1 \ S11(i) - c2 \\ \quad \quad \quad OR \\ S11(i) < -1 \ \& \ S12(i) > -c1 \ S11(i) + c2 \end{cases} \quad (22)$$

where $c1 = 0,13$ and $c2 = 0,17$ were estimated using simulations by Lipponen and Tarvainen [126]. The classification of an ectopic beat required that one of the two statements in Equation (22) was fulfilled and the beat satisfied the condition showed in Equation (16). The first and last value were forced to be 0 (= non-ectopic beat) for the sake of algorithm functionality.

Missed beats

A missed beat is a situation where an abnormally long RRI occurs. This can be due to an undetected R-peak or a real escape beat. Missed beats form the positive-negative segments to the dRR series. In order to detect missed (and extra beats), the second two-dimensional space S2 was created as in Equation (23) and Equation (24)

$$S21(i) = dRR(i), \quad i = 1 \dots N \quad (23)$$

$$S22(i) = \begin{cases} \min[dRR(i+1), dRR(i+2)], & \text{if } dRR(i) \geq 0 \\ \max[dRR(i+1), dRR(i+2)], & \text{if } dRR(i) < 0 \end{cases} \quad (24)$$

The $S22(1) = dRR(1)$ and $S22(2) = dRR(2)$ since the equation was inoperative for the last two elements.

Equation (25) shows the three statements that were used to determine whether a beat was a possible missed or extra beat. The fulfillment of one of the three statements was enough.

$$Possible \ missed \ or \ extra \ beat = \begin{cases} S21 > 1 \ \& \ S22 < -1 \\ \quad \quad \quad OR \\ S21 < -1 \ \& \ S22 > 1 \\ \quad \quad \quad OR \\ |mRR| > 3 \end{cases} \quad (25)$$

The last two beats were forced to be normal beats since the algorithm was unable to handle artefact beats at the end of the data.

A missed beat was classified if the beat satisfied the condition in Equation (16), any of the statements in Equation (25), and if the RRI satisfied the following condition in Equation (26):

$$Missed\ beat = \left| \frac{RR(i)}{2} - medianRR(i) \right| < Th2 \quad (26)$$

where the $medianRR$ was calculated using the same principle as in Equation (13).

Extra beats

Extra beats create the negative-negative-positive or negative-positive-positive segments to the dRR series. They are always originated from a detection error. A extra beat was classified if the beat satisfied the condition in Equation (16), any of the statements in Equation (25) and if the RRI satisfied the following condition in Equation (27):

$$Extra\ beat = |RR(i) + RR(i + 1) - medianRR(i)| < Th2 \quad (27)$$

Architecture

Figure 27 summarizes the architecture of the algorithm.

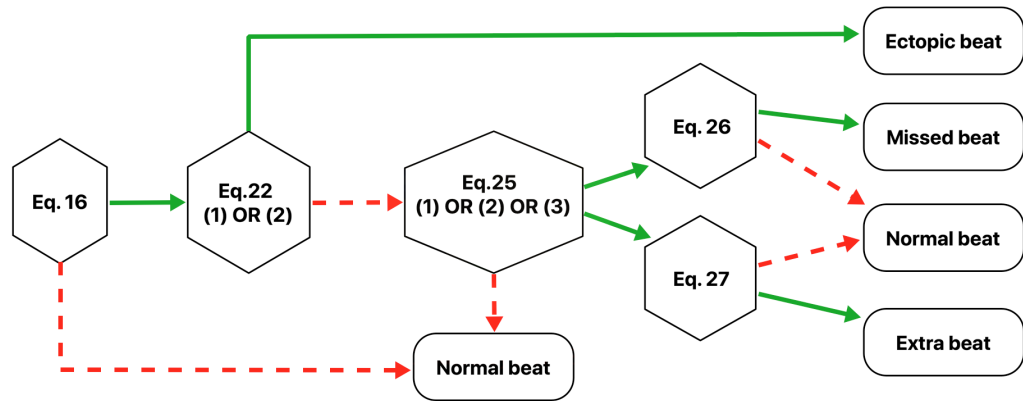


Figure 27. The architecture of the developed beat detection algorithm. The decision making starts from Equation (19) and continues in the direction indicated by the arrows. Solid arrows indicate satisfied condition and dashed arrows unsatisfied condition.

4.2.3. Artefact Correction

After artefact beat classification, the beats were corrected. The correction method depended on the type of the artefact beat. Extra beats were corrected by removing

the extra R-peak detection. Missed and ectopic beats were corrected by replacing the missing or distorted beat by an interpolated one.

Correction for the missed and ectopic beats was done in stages, first for the missed beats and then for the ectopic beats. Missed beat correction was started by inserting NaN (= Not a Number) values inside the beat array at the positions of the missing beats. The interpolation was performed over the NaN values using the `interp1` function²⁵. For the interpolation method, 'spline' was selected. The spline algorithm produce cubic interpolation to generate piecewise polynomials with continuous second-order deviates.

Ectopic beat correction was started by removing etopic beat detection points from the beat array and replacing them by NaN values. Finally, the interpolation was executed over the NaN values using the same function with the same properties. After all corrections, the RR intervals and heart rates were recalculated. The resulted ECG signal with detected R-peaks and artefact corrections was the output of the algorithm.

4.3. Template Extraction

After beat detection, VCG signal templates were extracted from the ECG signal to the correct HR-levels. The HR in bpm was calculated from the RRI information, which is shown in Equation (28):

$$HR(i) = \frac{60}{RR(i)}, \quad i = 1 \dots N \quad (28)$$

where N is the number of RR intervals. Since the HR was calculated from the RRI, the actual number of the beats, and heart rates, was $N + 1$. Therefore, the HR for the first beat was added and assumed to be the same as the second beat $HR(1) = HR(2)$.

The HR data was smoothed by the moving average filter using the `smooth` function²⁶. The span parameter was decided to be 11, which means the number of data points used for calculating the smoothed value.

4.3.1. Heart Rate Level-Based Clustering

The HR signal like the one visualized in Figure 18.a was utilized for the HR analysis. The initial and end part of the signal were omitted, since the initial part often contained a lot of noise and the end part contained the recovery phase after incremental exercise (not seen in Figure 18.b). The K-means clustering was used to detect different HR-levels in the signal. The function of the K-means clustering was described in Section 3.4. The number of clusters (i.e. HR-levels) was set to 18, and the `kmeans` function²⁷ was used to perform the clustering.

The `kmeans` function uses the squared Euclidean distance metric and the `k-means++` algorithm for the cluster center initialization. According to Arthur and Vassilvitskii

²⁵MATLAB `interp1` function - <https://www.mathworks.com/help/matlab/ref/interp1.html>

²⁶MATLAB `smooth` function - <https://www.mathworks.com/help/curvefit/smooth.html>

²⁷MATLAB `kmeans` function - <https://www.mathworks.com/help/stats/kmeans.html>

[127], the k-means++ algorithm improves the running time and the quality of the final solution of the Lloyd's algorithm. The k-means++ algorithm works as follows:

1. An observation is selected randomly from the data set. The chosen observation is the first centroid c_1 .
2. Distances from each data point to the c_1 are computed. The distance between the point x_m and the centroid c_j is denoted as $d(x_m, c_j)$.
3. The next centroid c_2 is selected randomly from the data set with the probability P which is derived from Equation (29):

$$P(c_2) = \frac{d^2(x_m, c_1)}{\sum_{j=1}^n d^2(x_j, c_1)} \quad (29)$$

where n is the number of data points.

4. To choose the centroid c_j , the distances from each data point to each centroid is computed. Each data point is assigned to its closest centroid. The centroid c_j is selected randomly from the data set with the probability P which is derived from Equation (30):

$$P(c_j) = \frac{d^2(x_m, c_p)}{\sum_{\{h; x_h \in C_p\}} d^2(x_h, c_p)} \quad (30)$$

where $m = 1, \dots, n$ and $p = 1, \dots, j - 1$. C_p is the set of all data points closest to the centroid c_p and x_m belongs to C_p .

5. The step 4 is repeated until the specified number of centroids is selected.

4.3.2. VCG and Inverse Dower's Matrix

An inverse Dower's transformation matrix was used to convert the ECG to VCG. Table 6 shows the transformation matrix, which is based on the mathematical pseudoinversion of the Dower's transformation matrix illustrated in Table 2. The leads X, Y and Z are expressed by a linear combination of 8 leads.

The VCG beat matrix was constructed using Equation (31):

$$\mathbf{VCG} = \mathbf{D} \cdot \mathbf{ECG} \quad (31)$$

where \mathbf{D} is the inverse Dower's transformation matrix. Templates were aligned to the R-peak locations 200 ms before and 300 ms after. Finally the VCG signals were combined to their corresponding HR-levels. It should be noted that each HR-level contained a different number of VCG signals. The number of VCG signals was specified by the clustering algorithm and was equal to the number of R-peaks determined to correspond to that HR-level.

Table 6. Inverse Dower's transformation matrix

Lead	X	Y	Z
I	0.1561	-0.2274	0.0217
II	-0.0099	0.8865	0.1021
V ₁	-0.1724	0.0572	-0.2289
V ₂	-0.0738	-0.0190	-0.3100
V ₃	0.1222	-0.1064	-0.2459
V ₄	0.2310	-0.0220	-0.0634
V ₅	0.2393	0.0409	0.0548
V ₆	0.1936	0.0483	0.1085

4.3.3. Template Verification

In order to improve the accuracy and reliability of the extracted VCG signal templates, an outlier detection algorithm was developed. First, the median VCG signal leads corresponding to each HR-level were computed. The pointwise absolute difference D between each VCG signal lead and the median $mVCG$ was determined as in Equation (32):

$$D_L(j, i) = |VCG_L(j, i) - mVCG_L(j, h)|, \quad i = 1, \dots, N \quad (32)$$

where N is the number of VCG signals at a given HR-level h , j is a data point, and L indicates a specific lead (X, Y, Z). The number of data points was 270, so $j = 1, \dots, 270$.

The difference percentages ($D\%$) for each lead were computed from the obtained results. Lastly, the total difference percent ($Total\ D\%$) for each VCG signal was obtained by combining the leads and calculating the mean difference percent like shown in Equation (33) and Equation (34):

$$D\%_L(i) = \frac{\sum_{j=1}^K D_L(j, i)}{\sum_{j=1}^K |mVCG_L(j, h)|} \quad (33)$$

$$Total\ D\%(i) = \frac{\sum_{L=1}^3 D\%(L, i)}{3} \cdot 100 \quad (34)$$

where $K = 270$ is the number of data points.

The decision was based on the specified threshold value, which was set to 100 (%). The total difference percent for each signal was compared to the threshold value, and if the value was exceeded, the signal was removed from the signal library. Histogram, as shown in Figure 28, was used for visualize the differences. Finally, the mean VCG signal for each HR-level was calculated to form the final templates.

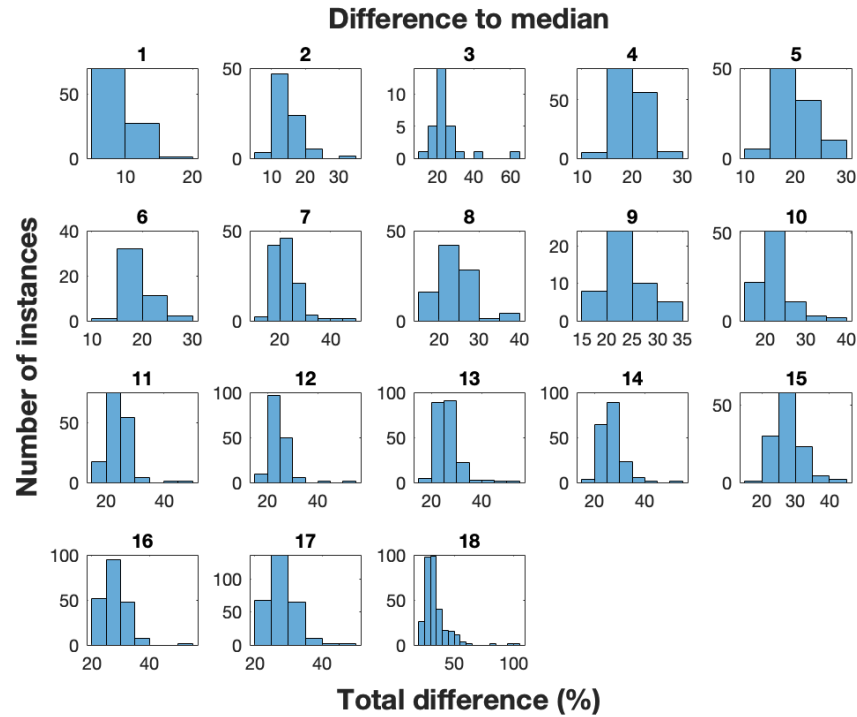


Figure 28. The outlier detection visualization. Histogram numbers (1-18) represent specific heart rate level. The total difference percent is shown on the x-axis and the number of signals is represented on the y-axis of the histograms.

4.4. Graphical User Interface

The GUI was developed for the processes mentioned above. The GUI enabled the user to easily customize the parameters used in different steps and instantly inspect the effects of new parameters. It consisted of two connected windows, that is, from the first window one could move to the second. The first window included functions for the preprocessing methods, beat detection, and HR-level clustering. The second window contained properties for the VCG templates. The individual components of the GUI are presented in Appendices 1 to 5.

An overview of the first window is presented in Figure 29. The window consisted of the ECG data information and individual subtabs for each process. In addition, the ECG signal and HR-levels were plotted to the top of the window. With the GUI, additional properties could be added for the processes. For the beat detection, manual R-peak detection was included. This allowed the user to manually add or remove any beat detection. In addition, the possibilities to modify the heart rate signal limits and view the high-pass filter were added.

The user could load a new ECG file or import the parameters and signals used in the selected previous session. The save option was added to save the parameters used in the session. The parameters and HR-levels with corresponding templates were stored as a MATLAB .mat file in the selected directory. When the user wanted to end the session, pressing the End button closed the GUI.

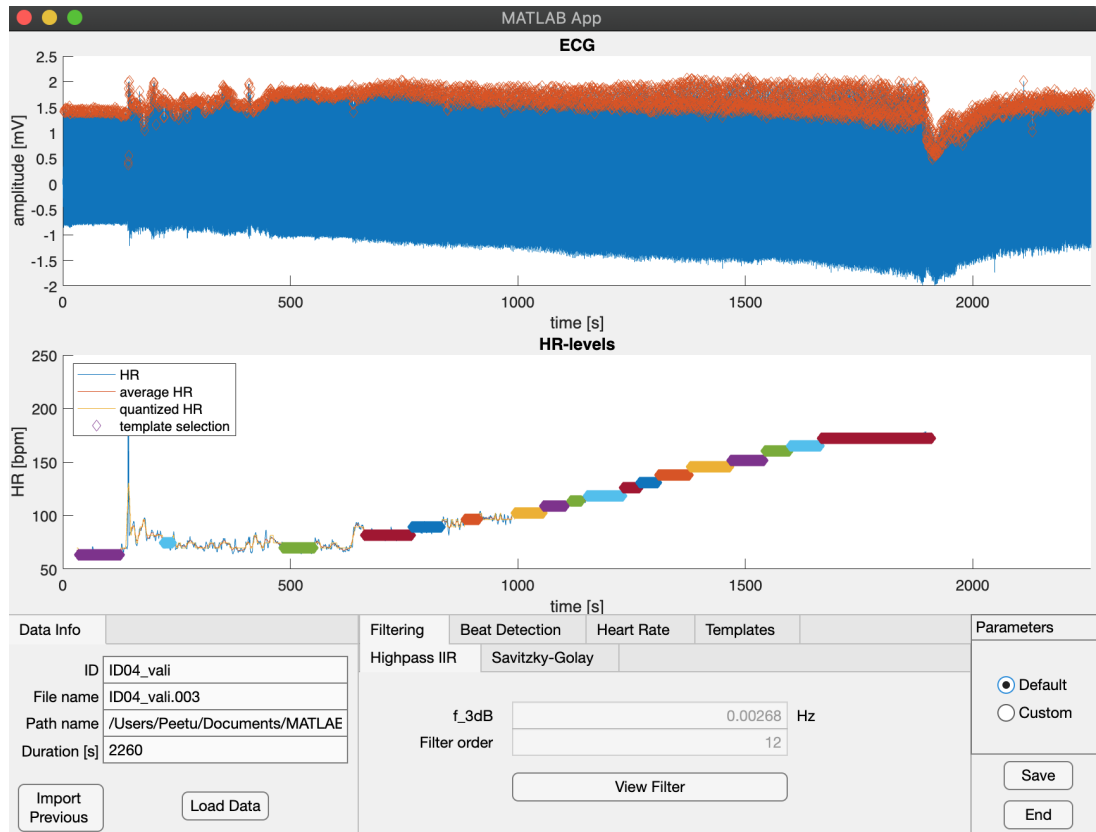


Figure 29. The first window of the graphical user interface. Electrocardiogram signal and heart rate graph are plotted to the top of the window.

The second window is presented in Figure 30. The second window contained two views which are represented in Appendix 4. The first view responded to the final template properties and the second view to the individual templates. The cut points were added to the templates, because especially for the high HR-levels, the next signal sometimes occurred at the end part of the template (or the previous signal was occurred at the initial part of the template). The cut points could be modified by the user.

In the individual template view, the user could inspect all the VCG signals from a specific HR-level. The user was able to drop signals out of the median calculations related to the outlier detection algorithm. In this way, if some signal was clearly differentiable to others, it did not affect to the median. Also, the threshold value used for decision-making for the outlier detection was editable. When the user wanted to stop the template customization and pressed the Return button, the software returned to the first window.

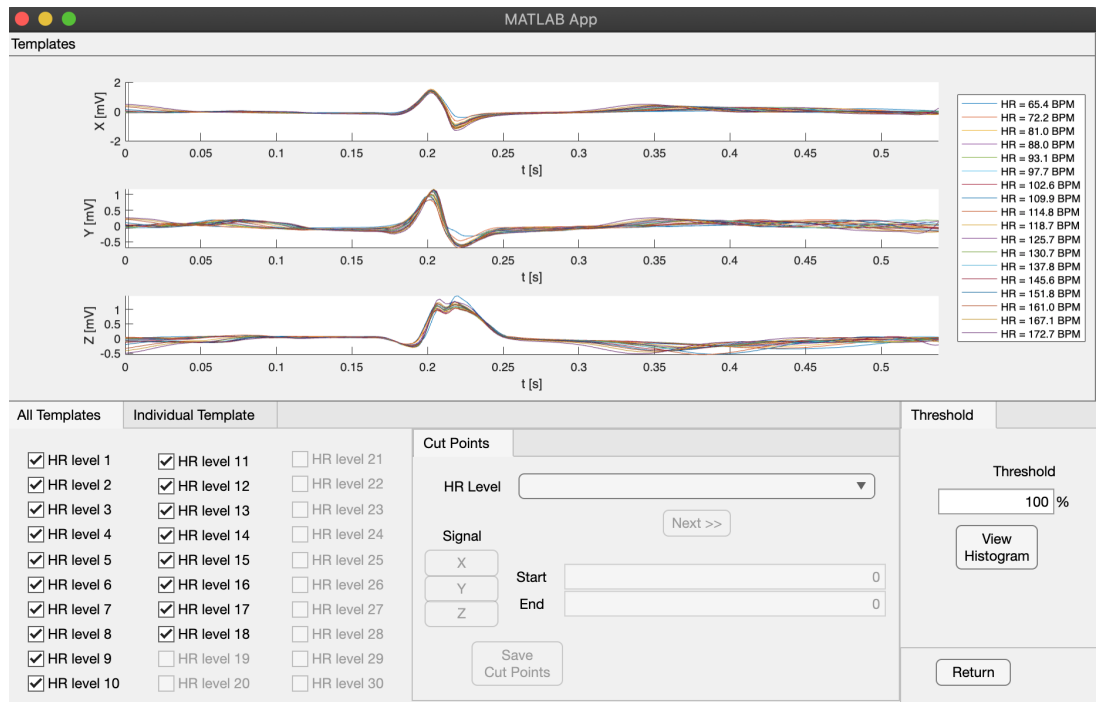


Figure 30. The second window of the graphical user interface. Vectorcardiogram templates are displayed to the top of the window.

5. RESULTS

In this chapter, the results for developed models are discussed. First, the data sets used for testing purposes are introduced. Subsequently, the results for each method are presented sequentially starting from the noise reduction methods and followed by the beat detection algorithm and template extraction.

5.1. Data Used in Tests

In this study, the ECG data from stress exercises collected by Kiviniemi et al. [82] in the year 2010 were utilized. The data set consisted of 42 separate ECG recordings from healthy subjects (21 women, 21 men) aged 28-40 years. All the subjects had a body mass index below $30 \text{ kg}\cdot\text{m}^{-2}$ and were non-smokers. Additionally, individuals with diabetes mellitus, asthma, or cardiovascular disorders were excluded. The subjects executed an incremental exercise test on a bicycle ergometer (Monark Ergomic 839 E; Monark Exercise AB, Vansbro, Sweden) until voluntary exhaustion. The measurements were performed in the Department of Exercise and Medical Physiology at Verve (Oulu, Finland).

Incremental exercise denotes an exercise where the stress level is gradually increased. In the aforementioned research, the starting resistance of the bicycle was set at 25 watts (W) and was followed by an increase of $1 \text{ W}\cdot 5 \text{ s}^{-1}$. The demonstration of the incremental stress exercise is shown in Figure 31. The ECG data from Kiviniemi et al. [82] was used to test HR-level clustering and VCG template extraction, since the ECG data from incremental exercise was a proper data set for this purpose. The ECG data were 8-channel recordings.

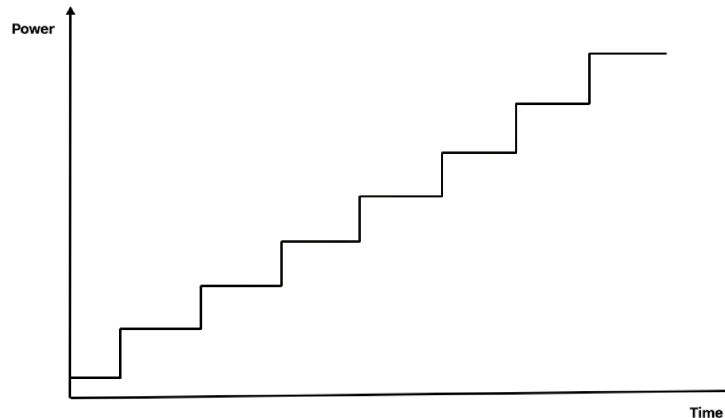


Figure 31. The demonstration of the incremental exercise.

The developed beat detection algorithm was tested using the MIT-BIH Arrhythmia Database [22, 23]. The database contains 48 half-hour two-channel ambulatory ECG recordings from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979. Half of the recordings became freely available in 1999, and remaining files were added to the freely available database in 2005. The recordings

include annotations from cardiologists and several arrhythmias are present. The MIT-BIH Arrhythmia Database is widely used database for evaluation of beat detection algorithms and was therefore selected for this study. Since the annotations were marked only for one of the two channels, only channel with annotations was included.

The records in the MIT-BIH Arrhythmia Database were originally sampled with the sampling frequency of 360 Hz. The sampling frequency used in this study was 500 Hz, so in order to use the data, resampling was implemented. Signal resampling was performed using the `interp1` function²⁸ with spline interpolation method. After resampling, the positions of the annotations were rounded to the nearest integer.

5.2. Outcome

Noise reduction

The preprocessing of the ECG signals was done using the HPF and Savitzky-Golay filter. The final output after filtering one example ECG signal from the MIT-BIH Arrhythmia Database (record 101) is presented in Figure 32. The figure contains a comparison between the original unfiltered and the filtered signal at three different segments.

Beat detection algorithm

The beat detection algorithm was tested with all 48 records from the MIT-BIH Arrhythmia Database. The performance of initial beat detection and after beat correction were evaluated for the comparison purposes. The performance was evaluated by the sensitivity (SEN), precision (PRE), and detection error rate (DER). The metrics were defined as in Equation (35), Equation (36), and Equation (37), respectively:

$$SEN = \frac{TP}{TP + FN} \quad (35)$$

$$PRE = \frac{TP}{TP + FP} \quad (36)$$

$$DER = \frac{FN + FP}{TP + FN} \quad (37)$$

where TP means the number of true positive detections, FN stands for the number of false negative detections, and FP is the number of false positive detections. Positive indicates a detected R-peak and negative denotes a data point without detection.

The metrics were determined by the decision-making rule that a beat is true when it is detected within 100 ms from the annotated time, otherwise false. The time limit was decided on the basis of other studies on the same topic [115, 128]. The sensitivity measured the ability of the algorithm to detect true beats, while the precision was used

²⁸MATLAB `interp1` function - <https://www.mathworks.com/help/matlab/ref/interp1.html>

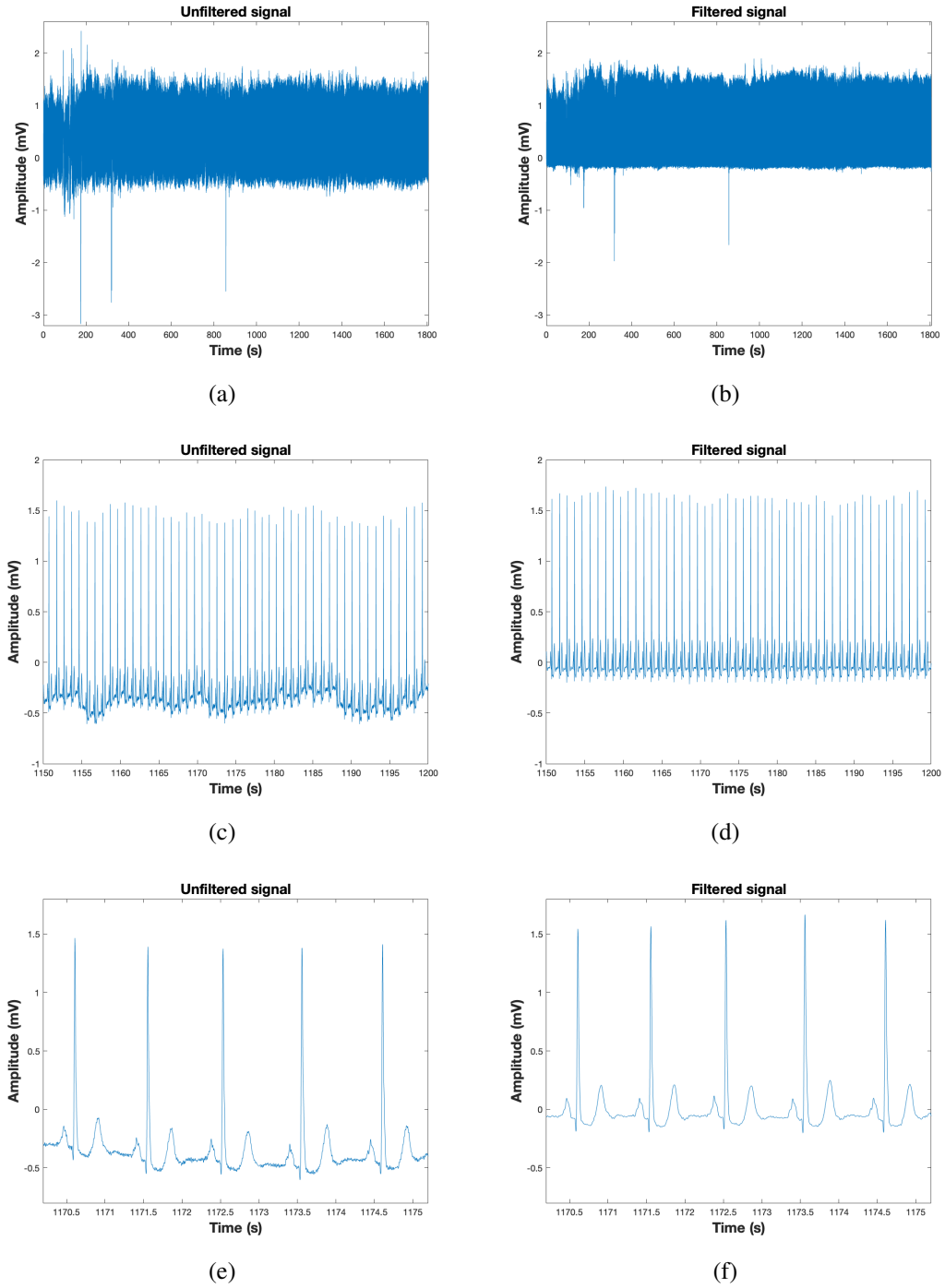
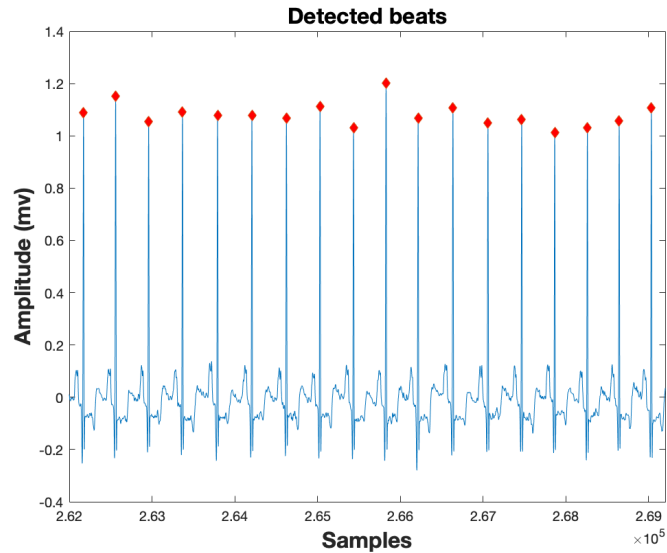


Figure 32. Filtering results for MIT-BIH recording 101. Unfiltered signals are on the left (a, c, e) and filtered signals on the right (b, d, f).

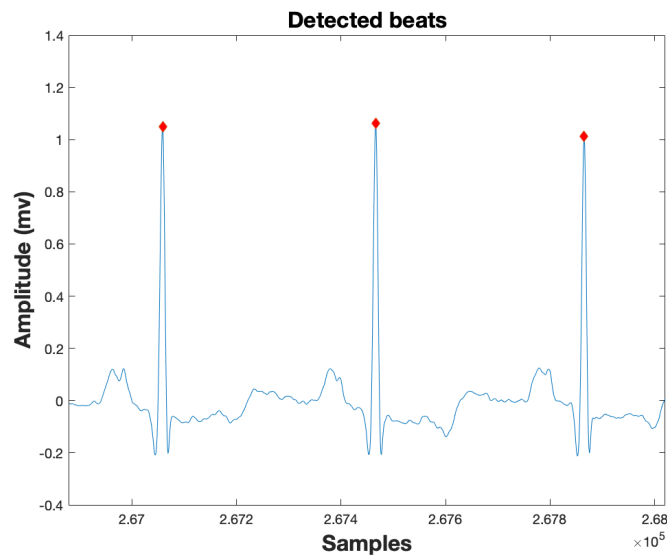
to evaluate the ability of the algorithm to distinguish between the true and false beats. The detection error rate was used to measure the accuracy of the algorithm.

Beats were detected like shown in Figure 33, where the detected R-peaks are marked with red marks. The figure presents the made detections in two separate segments for an example recording from the MIT-BIH Arrhythmia Database.

The overall results of the beat detection algorithm for the MIT-BIH Arrhythmia Database are presented in Appendix 6. The algorithm produced 630 false beats



(a)



(b)

Figure 33. Detected beats marked with red diamonds in two segments. a) The bigger segment b) The smaller segment.

including 382 false positive beats and 248 false negative beats, and succeeded in detecting correctly 99.42 % of the ECG beats. In detail, the sensitivity of 99.77 % and the precision of 99.65 % was achieved. The algorithm suffered the lowest performance for the records 203 and 207, and these two records covered 36.03 % of all detection errors.

The comparison between the initial beat detection and after artefact correction is presented in Table 7. In addition, the table contains the results for beat artefact classifications. A problem was encountered during the evaluation of the beat artefacts, which can be seen in the results of the beat artefact classifications, where especially the performance of the ectopic beat and missing beat classifications seems to be very

poor. This was explained by the fact that the interpolated (corrected) beats did not fall between the time limit. That is, the interpolated position of the beat often differed from the annotation position by more than the allowed time limit. In addition, missed beat classification might happen due to the unusually long RRI, which led the situation where added beat occurred in the position where none original beat was present. These aforementioned situations are demonstrated in Figure 34.

Table 7. Overall scores for the beat detections and beat artefact classifications

Method	ANN	TP	FP	FN	FP + FN	SEN	PRE	DER
Before algorithm	109,494	109,205	295	289	584	99.74	99.73	0.53
After algorithm	109,494	109,246	382	248	630	99.77	99.65	0.58
Ectopic beats	-	1561	1358	8456	9814	15.58	53.48	97.97
Missed beats	-	47	167	245	412	16.10	21.96	141.10
Extra beats	-	51	51	0	51	100.00	50.00	100.00

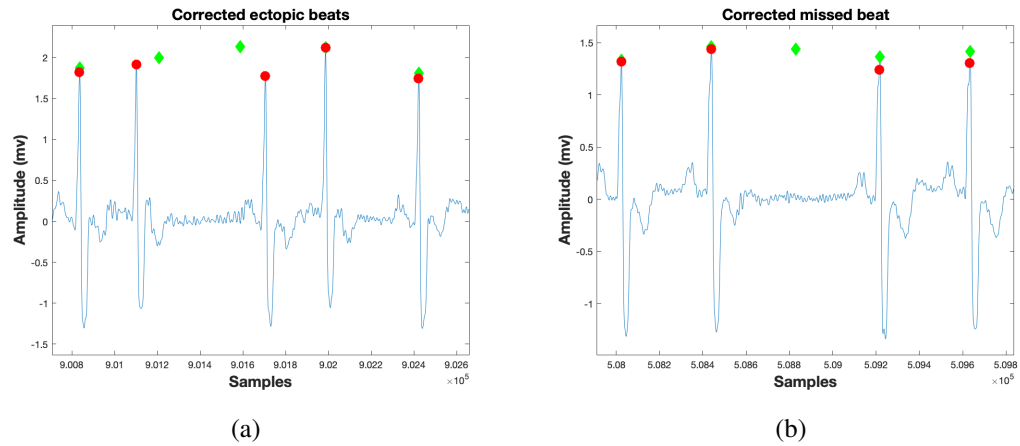


Figure 34. Placement differences between the corrected beats and annotations. Beats by algorithm are marked by green diamonds and annotations by red circles. a) Corrected ectopic beats b) Corrected missed beat.

Choi et al. [115] listed the performances of several developed QRS detection algorithms. Comparative results for the MIT-BIH Arrhythmia Database are summarized in Table 8. Furthermore, the comparison was made individually for the normal sinus rhythm records to check the reliability of the algorithm for normal beats and to minimize the effects of artefact beat time limit errors. Record 101 contained normal sinus rhythm with couple severe changes possibly caused by body movement. The comparison of the record 101 is summarized in Table 9. Record 105 consisted of normal sinus rhythm with a high amount of noise and artefacts. The performances of the record 105 are compiled in Table 10.

Template extraction

Data from Kiviniemi et al. [82] was used to test whether the HR-level clustering and template extraction were successful. The evaluation was based on the observations

Table 8. Comparison of QRS detection performance with several algorithms

Method	ANN	TP	FP	FN	FP + FN	SEN (%)	PRE (%)	DER (%)
Cvikl et al. (2007) [114]	109,494	109,294	200	200	400	99.82	99.82	0.37
Lee et al. (2002) [113]	109,486	109,151	137	335	472	99.69	99.87	0.43
Choi et al. (2010) [115]	109,494	109,118	218	376	594	99.66	99.80	0.54
Hamilton & Tompkins (1986) [108]	109,267	108,927	248	340	588	99.69	99.77	0.54
Present work	109,494	109,246	382	248	630	99.77	99.65	0.58
Christov (2004) [110]	110,050	109,762	387	287	674	99.74	99.65	0.61
Pan & Tompkins (1985) [107]	109,809	109,532	507	277	784	99.75	99.54	0.71
Afonso et al. (1999) [106]	90,909	90,535	374	406	780	99.55	99.59	0.86
Paoletti & Marchesi (2006) [109]	109,809	109,430	565	379	944	99.65	99.49	0.86
Poli et al. (1995) [112]	109,963	109,522	545	441	986	99.60	99.50	0.90
Chen et al. (2006) [111]	100,381	99,852	459	529	988	99.47	99.54	0.98
Benitez et al. (2001) [105]	45,856	45,828	30	28	58	99.94	99.93	0.13

Table 9. Performance comparison with several algorithms for the record 101

Method	ANN	TP	FP	FN	FP + FN	SEN (%)	PRE (%)	DER (%)
Choi et al (2010) [115]	1865	1865	1	0	1	100.00	99.95	0.05
Present work	1865	1865	1	0	1	100.00	99.95	0.05
Lee et al. (2002) [113]	1865	1865	2	0	2	100.00	99.89	0.11
Afonso et al. (1999) [106]	1523	1523	2	0	2	100.00	99.87	0.13
Benitez et al. (2001) [105]	1866	1865	3	1	4	99.95	99.84	0.21
Hamilton & Tompkins (1986) [108]	1859	1858	3	1	4	99.95	99.84	0.22
Cvikl et al. (2007) [114]	1865	1864	4	1	5	99.95	99.79	0.27
Christov (2004) [110]	1863	1862	4	1	5	99.95	99.79	0.27
Pan & Tompkins (1985) [107]	1865	1862	5	3	8	99.84	99.73	0.43
Chen et al. (2006) [111]	1865	1865	10	0	10	100.00	99.47	0.54
Paoletti & Marchesi (2006) [109]	1865	1860	7	5	12	99.73	99.63	0.64

Table 10. Performance comparison with several algorithms for the record 105

Method	ANN	TP	FP	FN	FP + FN	SEN (%)	PRE (%)	DER (%)
Benitez et al. (2001) [105]	2572	2569	7	3	10	99.88	99.73	0.39
Cvikl et al. (2007) [114]	2572	2567	34	5	39	99.81	98.69	1.52
Present work	2572	2554	24	18	42	99.30	99.07	1.63
Lee et al. (2002) [113]	2572	2568	41	4	45	99.84	98.43	1.75
Choi et al. (2010) [115]	2572	2541	21	31	52	98.79	99.18	2.02
Christov (2004) [110]	2567	2552	44	15	59	99.42	98.31	2.30
Hamilton & Tompkins (1986) [108]	2564	2542	53	22	75	99.14	97.96	2.93
Chen et al. (2006) [111]	2572	2567	78	5	83	99.81	97.05	3.23
Afonso et al. (1999) [106]	2139	2123	53	16	69	99.25	97.56	3.23
Pan & Tompkins (1985) [107]	2572	2550	67	22	89	99.14	97.44	3.46
Poli et al. (1995) [112]	2572	2567	86	5	91	99.81	96.76	3.54
Paoletti & Marchesi (2006) [109]	2572	2541	73	31	104	98.79	97.21	4.04

rather than on any metric values. Different HR-levels were detected and templates were extracted for every HR-level. Figure 35 shows the HR-level detection for one subject, where the number of clusters was 18.

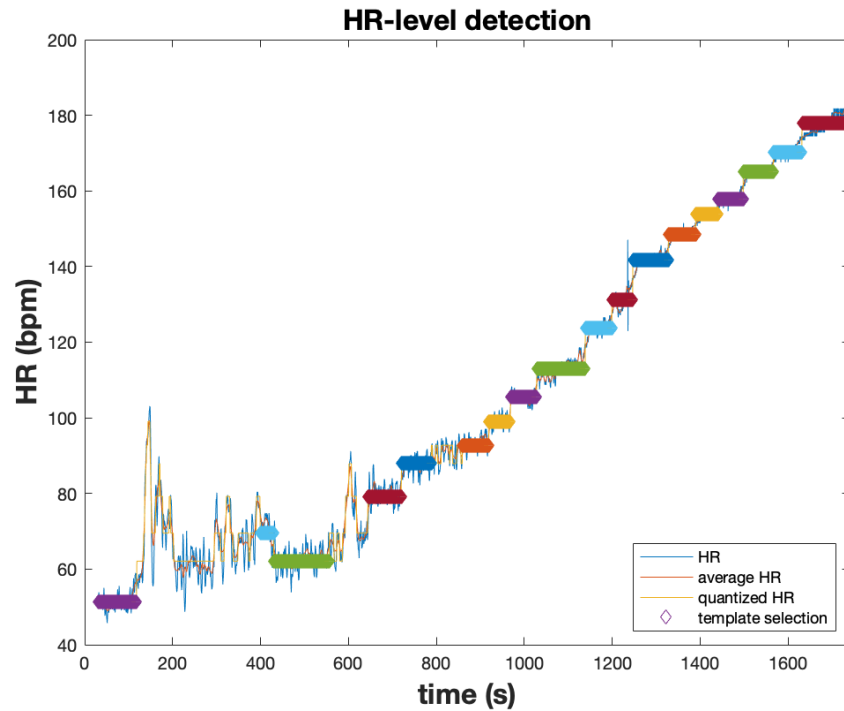


Figure 35. Heart rate level detection for 18 heart rate levels. Different clusters are marked with colors in the signal. The signal has been cut so that the initial and recovery parts were excluded.

Figure 36 shows all the signals before outlier detection and the final template for one HR-level for one example subject. The figure includes scalar vectorcardiogram signals differentiated with colors and respective final templates. An example of the final templates of all the HR-levels is presented in Figure 37. The figure shows extracted scalar vectorcardiogram templates for all the HR-levels for the same subject, where different HR-levels are marked with colors. In addition, different HR-levels are listed next to the templates.

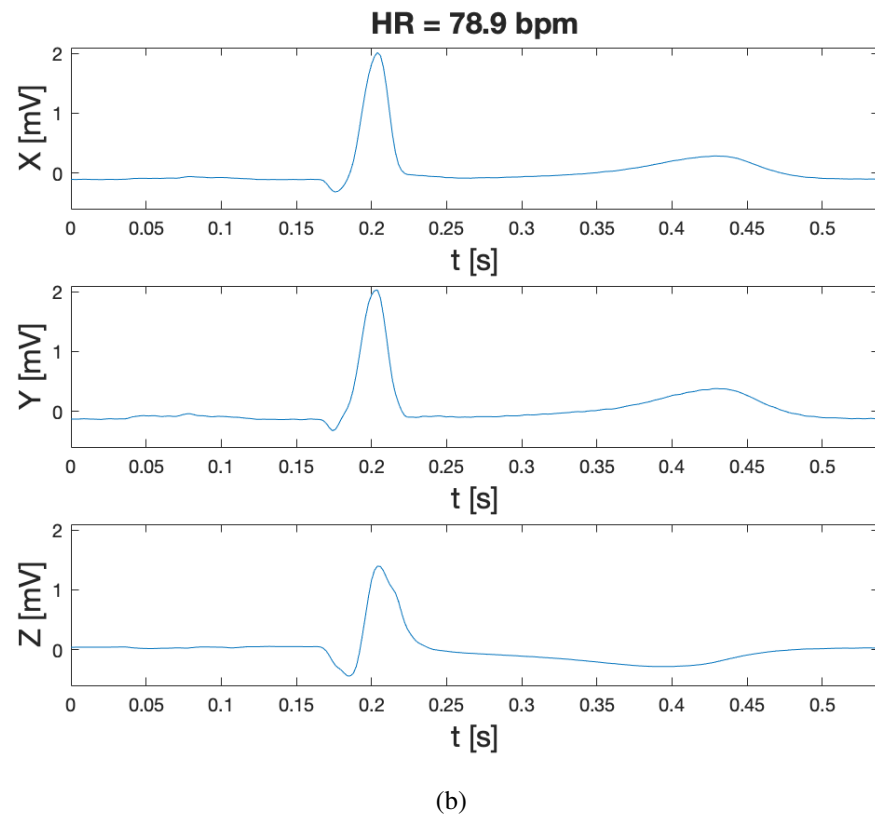
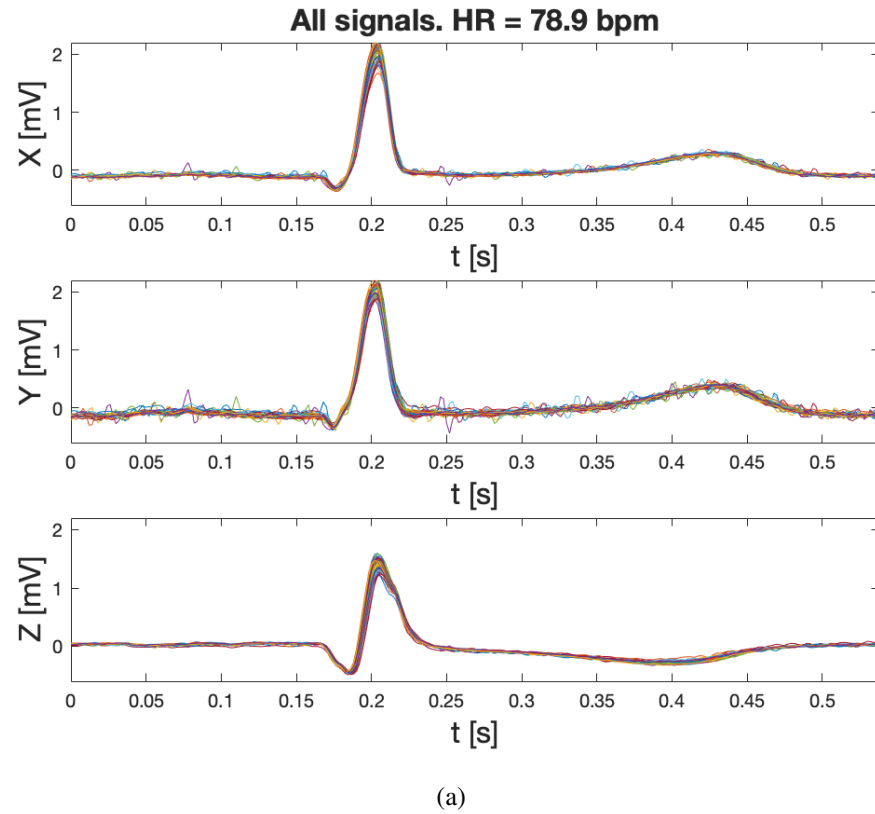


Figure 36. a) All vectorcardiogram signals (X, Y, X) before outlier detection. Different signals are marked with colors. b) The final template for one heart rate level.

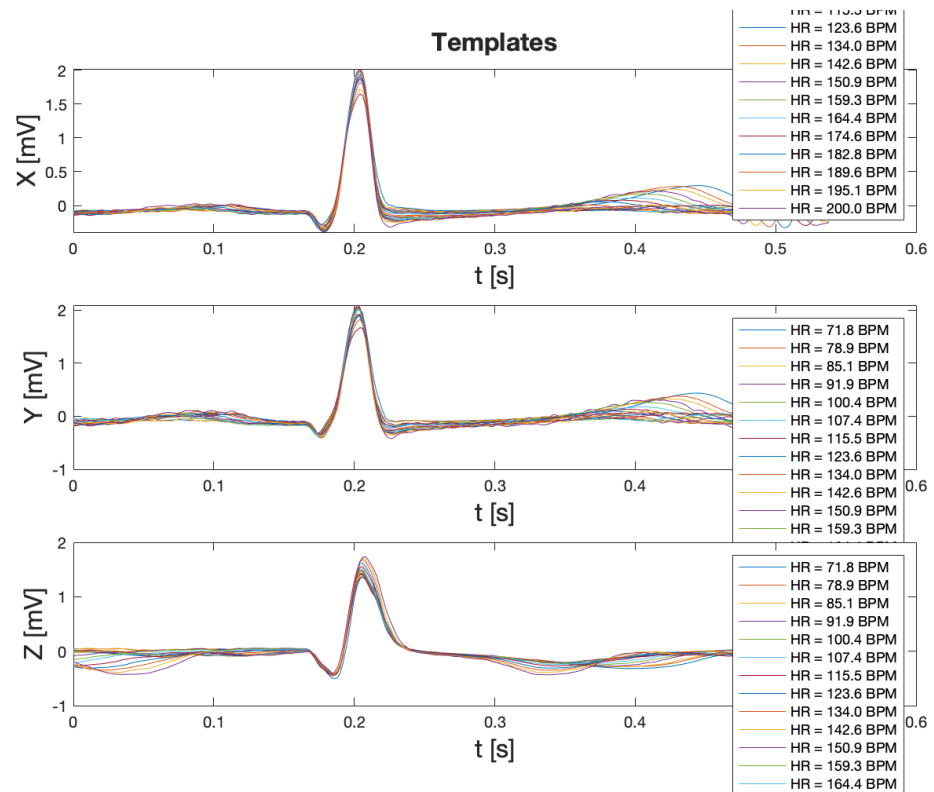


Figure 37. The final vectorcardiogram signal templates (X, Y, Z) for all detected HR-levels for one subject. Different HR-levels are marked with colors and listed next to the templates.

6. DISCUSSION

The goal of this study was to develop a beat detection algorithm and extract heart rate dependent ECG templates for a body emulator environment. Real-life ECG signals were preprocessed using the HPF and Savitzky-Golay filter. The developed beat detection algorithm was used to detect heartbeats from the ECG signals. Different HR-levels from the ECG signals were detected utilizing the K-means clustering algorithm. The ECG signals were converted to the VCG using the inverse Dower's transformation matrix, and the VCG signals were extracted to the respective HR-levels. The GUI was built to give a better view and customization properties to the processes. The developed beat detection algorithm was tested with the MIT-BIH Arrhythmia Database. The HR-level clustering and template extraction were validated using the incremental exercise ECG.

The noise was successfully reduced from the signals. From Figures 32.a and 32.b can be seen that the overall state of the signal is significantly smoother after filtering. Baseline wandering has been reduced, which can be noted from Figures 32.c and 32.d. The high-frequency noise has been remarkably attenuated as shown in Figures 32.e and 32.f.

At the beginning of the beat detection algorithm tests, fixed parameters were used to detect R-peaks in the ECG signal. It was quickly discovered that the records in the MIT-BIH Arrhythmia Database were too variable and the results differed significantly from each other. Consequently, it was decided to use variable parameters. This also described better the situation for which the algorithm was actually developed, since it is possible to modify the parameters in the GUI.

The QRS complexes were successfully detected in majority of the records (Appendix 6), particularly for normal sinus rhythm. In addition, the algorithm achieved good scores on records with LBBB, junctional premature beats, and PVCs (uni- and multiform). The lowest performance was resulted for the records 203 and 207. This might be due to the considerable noise, ventricular flutter, and idioventricular rhythm. Especially the record 203 contained axis shifts, which caused the QRS morphology changes. Cardiologists had marked the both records to an extremely difficult. In addition, lower performance for the records 124 and 228 could be due to the idioventricular rhythm and short occurrence tape slippages, respectively.

The comparison was made for the QRS detection performance with several algorithms (Table 8). For overall scores, the results of the algorithm were slightly better than the average. Especially the sensitivity of the algorithm was higher than for many others. The algorithm performed well with normal sinus rhythm and many high noise signals. The performance for the record 101 (Table 9), which was consisted of mainly normal sinus rhythm, was equal to the best of all the algorithms. Record 105 contained normal sinus rhythm with a high amount of noise and artefacts, and the developed algorithm produced the third best performance on this record (Table 10). These results show that the preprocessing methods (Figure 32) were sufficient and the algorithm was able to detect the QRS complexes from normal sinus rhythm with the high accuracy.

A problem was occurred related to the evaluation for the beat artefact classification and the results (Table 7) did not reflect the actual outcome of the classification performance. The interpolated beats did not fall between the time limits, which

resulted false detections. For instance, one corrected ectopic beat which exceeded the time limit resulted one false positive and one false negative detection (Figure 34). Missed beats were corrected by adding the new beat detection, which was sufficient in situations where detection error for the real beat was present. In contrast, if the missed beat classification happened due to an unusually long RR interval, the added detection was surely to be classified as a false positive beat. These difficulties might be due to the insufficient interpolation methods or that the evaluation method (i.e. the comparison to the original peak locations) was not suitable for this purpose.

Although the evaluation problems were present, there were also real classification errors. Most often the ectopic beat classification occurred when there was not a real ectopic beat present. Sometimes the missed beat classification happened when a real ectopic beat was present, and vice versa. Nonetheless, most often these beat corrections served their purpose of creating an ECG signal in which changes in the RR interval occurred as a result of heart rate variability and not due to arrhythmias, noise, or artefacts.

The initial beat detection resulted lower detection error rate than after the beat artefact correction (Table 7). However, it should be noted that even those evaluation errors presented above, the algorithm gained a better sensitivity score and more TP detections than the initial beat detection. Furthermore, based on the observations above, the performance of the algorithm can be supposed to be slightly better than the obtained scores.

Heart rate levels were detected using the K-means clustering algorithm (Figure 35). It can be seen that it is more challenging to detect HR-levels from the initial part of the signal since the HRV is higher than elsewhere. However, lower heart rates are also needed, and the figure shows that it was possible to detect HR-levels from different heart rates.

Vectorcardiogram templates were extracted to the correct heart rate levels (Figure 37). It can be noted that the templates are uniform, although small differences can be noticed close to the edges. That is normal because different heart rates have different signal morphology. In addition, it can be noticed in Figure 36 that even when there were some distorted signals involved, the final template was congruent to the majority of the signals. Therefore it can be said, that the outlier detection worked as planned.

The GUI was created successfully, and it gave many beneficial function possibilities. The manual beat detection and algorithm parameter customization properties enabled basically 100 % beat detection accuracy. The number of desired heart rate levels was easy to change, and the time limit modification helped to include only the relevant part of the signals. The cut points of the templates enabled the customization of the template lengths, which excluded disturbances from the adjacent beats.

The limitations of this study include the evaluation methods for artefact beat classification. While the beat detection algorithm obtained relatively good results, the assessment of the artefact beat classification remained unresolved. In addition, extracted templates were not tested in the phantom, and the functionality of the templates was only validated by a visual inspection. Although the templates seemed successful, a more reliable evaluation would have been achieved with a real test in the phantom.

The extracted templates in this study can be used as an input for the dipoles attached to the emulator to produce realistic modeling cardiac signals. The developed GUI

can be used to customize the template generation properties to obtain the desired templates for each situation. These developed models are part of the larger picture of the human emulator, that can emulate the physiological functions of the human body. The phantom can be used for testing medical devices and analysis methods, and could be one solution to the difficulties of human trials.

For the future developments, a re-evaluation of the artefact beat classification is needed, since the current effects of the classifications and thus the functionality of the algorithm are partially unknown. The evaluation method needs modification, since the interpolated beats did not fall between the time limits. One approach could be a wider time limit for artefact beats and a narrower time limit for normal beats. In addition, the development of interpolation methods used should be considered. Furthermore, in the future, the extracted templates could be tested in the phantom to assure the quality of the developed models.

7. SUMMARY

In this research, heart rate depended electrocardiogram templates were developed for a body emulator environment. The real-life ECG signals were preprocessed with two digital filters. The high-pass Butterworth IIR filter was used to reduce low-frequency noise in the signals, and the Savitzky-Golay FIR filter was used to attenuate high-frequency noise in the signals.

A beat detection algorithm was developed to detect QRS complexes in the ECG signal. It consisted of the initial QRS detection followed by the artefact beat classification and correction. The algorithm was based on the RR interval information and the classifications were made based on the irregular RR intervals. The developed algorithm was tested with the MIT-BIH Arrhythmia Database, and the results were compared with the state-of-the-art algorithms. Out of 109 494 beats, 99.42 % were correctly detected, and the algorithm resulted the sensitivity of 99.77 % and precision of 99.65 %. The developed QRS detection algorithm performed successfully in majority of the records, and was better for records contained normal sinus rhythm and PVCs. The comparison of the QRS detection performance with several state-of-the-art algorithms was made. The developed algorithm achieved the best results for a low noise ECG signal with normal sinus rhythm and the third best results for a high noise signal.

Heart rate levels were detected by the K-means clustering method. The K-means++ algorithm was used for the cluster center initialization. Vectorcardiogram signals were converted from the electrocardiogram signals using the inverse Dower's transformation matrix. Vectorcardiogram templates were extracted to the related heart rate levels. The outlier detection algorithm based on the mathematical median was used to improve the quality of the templates. The incremental stress exercise ECG data was used to validate the functionality of the methods.

Finally, a graphical user interface was created to enhance the process visibility and customization properties. In the GUI, manual beat selection was developed to enable the best possible outcome for the beat detection. Cut point modifications for the heart rate signal and VCG templates were introduced to improve the quality of the detected heart rate levels and final templates. The GUI was properly integrated into the processes and the final templates were successfully obtained.

These final VCG templates can be used to generate an electrical impulse in the phantom using dipoles placed at the heart location. The electrical impulses will propagate in the phantom by ions mixed to the torso and produce cardiac signals that emulate the real human heart. The signals from the heart emulator can be measured with electrodes placed on the torso as a conventional ECG measurement. The phantom will also emulate the motion of the human torso, such as respiratory and muscle movement. These will make the signals even more realistic. The phantom will be a platform for testing the ECG related devices and analysis methods instead of humans.

8. REFERENCES

- [1] Kligfield P., Gettes L.S., Bailey J.J., Childers R., Deal B.J., Hancock E.W., van Herpen G., Kors J.A., Macfarlane P., Mirvis D.M., Pahlm O., Rautaharju P. & Wagner G.S. (2007) Recommendations for the Standardization and Interpretation of the Electrocardiogram: part I: The Electrocardiogram and Its Technology: A Scientific Statement From the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology. *Circulation* 115(10), pp. 1306–1324. DOI: <https://doi.org/10.1161/CIRCULATIONAHA.106.180200>.
- [2] Waller A.D. (1887) A Demonstration on Man of Electromotive Changes accompanying the Heart's Beat. *The Journal of Physiology* 8(5), pp. 229–234. DOI: <https://doi.org/10.1113/jphysiol.1887.sp000257>.
- [3] Barold S.S. (2003) Willem Einthoven and the Birth of Clinical Electrocardiography a Hundred Years Ago. *Cardiac Electrophysiology Review* 7(1), pp. 99–104. DOI: <https://doi.org/10.1023/A:1023667812925>.
- [4] Besterman E. & Creese R. (1979) Waller - pioneer of electrocardiography. *Heart* 42, pp. 61–64. DOI: <https://doi.org/10.1136/hrt.42.1.61>.
- [5] Soliman E.Z., Backlund J.C., Bebu I., Orchard T.J., Zinman B., Lachin J.M. & the DCCT/EDIC Research Group (2017) Electrocardiographic Abnormalities and Cardiovascular Disease Risk in 1 Diabetes: The Epidemiology of Diabetes Interventions and Complications (EDIC) Study. *Diabetes Care* 40(6), pp. 793–799. DOI: <https://doi.org/10.2337/dc16-2050>.
- [6] Al-Sakini N. (2022) Anatomy of the heart. *Medicine (Abingdon. 1995, UK ed.)* 50(6), pp. 317–321. DOI: <https://doi.org/10.1016/j.mpmed.2022.03.014>.
- [7] Malmivuo J. & Plonsey R. (1995) Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields. Oxford University Press, New York, USA, 482 p.
- [8] Shaffer F., McCraty R. & Zerr C.L. (2014) A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology* 5, p. 1040. DOI: <https://doi.org/10.3389/fpsyg.2014.01040>.
- [9] Mahadevan V. (2018) Anatomy of the heart. *Surgery (Oxford)* 36(2), pp. 43–47. DOI: <https://doi.org/10.1016/j.mpsur.2017.11.010>.
- [10] Nerbonne J.M. & Kass R.S. (2005) Molecular Physiology of Cardiac Repolarization. *Physiological Reviews* 85(4), pp. 1205–1253. DOI: <https://doi.org/10.1152/physrev.00002.2005>.

- [11] Kleber A.G. & Wit A.L. (2020) The Interaction Between Na^+ and Ca^{2+} Inward Currents in Cardiac Propagation. *Circulation Research* 127(12), pp. 1549–1551. DOI: <https://doi.org/10.1161/CIRCRESAHA.120.318316>.
- [12] Kanno S. & Saffitz J.E. (2001) The Role of Myocardial Gap Junctions in Electrical Conduction and Arrhythmogenesis. *Cardiovascular Pathology* 10(4), pp. 169–177. DOI: [https://doi.org/10.1016/s1054-8807\(01\)00078-3](https://doi.org/10.1016/s1054-8807(01)00078-3).
- [13] Davis L.M., Kanter H.L., Beyer E.C. & Saffitz J.E. (1994) Distinct Gap Junction Protein Phenotypes in Cardiac Tissues With Disparate Conduction Properties. *Journal of the American College of Cardiology* 24(4), pp. 1124–1132. DOI: [https://doi.org/10.1016/0735-1097\(94\)90879-6](https://doi.org/10.1016/0735-1097(94)90879-6).
- [14] Monfredi O., Dobrzynski H., Mondal T., Boyett M.R. & Morris G.M. (2010) The Anatomy and Physiology of the Sinuatrial Node - A Contemporary Review. *Pacing and Clinical Electrophysiology* 33(11), pp. 1392–1406. DOI: <https://doi.org/10.1111/j.1540-8159.2010.02838.x>.
- [15] Waller A.D. (1889) On the electromotive changes connected with the beat of the mammalian heart, and of the human heart in particular. *Philosophical Transactions of the Royal Society of London. B.* 180, pp. 160–194. DOI: <https://doi.org/10.1098/rstb.1889.0004>.
- [16] Kahn J.K. & Howell J.D. (1987) Frank Norman Wilson. *Clinical Cardiology* 10(10), pp. 616–618. DOI: <https://doi.org/10.1002/clc.4960101022>.
- [17] Patterson R.P. (1978) The electrical characteristics of some commercial ECG electrodes. *Journal of Electrocardiology* 11(1), pp. 23–26. DOI: [https://doi.org/10.1016/S0022-0736\(78\)80025-9](https://doi.org/10.1016/S0022-0736(78)80025-9).
- [18] Francis J. (2016) ECG monitoring leads and special leads. *Indian Pacing and Electrophysiology Journal* 16(3), pp. 92–95. DOI: <https://doi.org/10.1016/j.ipej.2016.07.003>.
- [19] Mason R.E. & Likar I. (1966) A new system of multiple-lead exercise electrocardiography. *American Heart Journal* 71(2), pp. 196–205. DOI: [https://doi.org/10.1016/0002-8703\(66\)90182-7](https://doi.org/10.1016/0002-8703(66)90182-7).
- [20] Al Rahhal M.M., Bazi Y., AlHichri H., Alajlan N., Melgani F. & Yager R.R. (2016) Deep learning approach for active classification of electrocardiogram signals. *Information Sciences* 345, pp. 340–354. DOI: <https://doi.org/10.1016/j.ins.2016.01.082>.
- [21] Mincholé A., Camps J., Lyon A. & Rodríguez B. (2019) Machine learning in the electrocardiogram. *Journal of Electrocardiology* 57, pp. S61–S64. DOI: <https://doi.org/10.1016/j.jelectrocard.2019.08.008>.

- [22] Moody G.B. & Mark R.G. (2001) The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine* 20(3), pp. 45–50. DOI: <https://doi.org/10.1109/51.932724>.
- [23] Goldberger A.L., Amaral L.A.N., Glass L., Hausdorff J.M., Ivanov P.C., Mark R.G., Mietus J.E., Moody G.B., Peng C.K. & Stanley H.E. (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]* 101(23), pp. e215–e220. DOI: <https://doi.org/10.1161/01.CIR.101.23.e215>.
- [24] Zheng J., Zhang J., Danioko S., Yao H., Guo H. & Rakovski C. (2020) A 12-lead electrocardiogram database for arrhythmia research covering mote than 10,000 patients. *Scientific Data* 7(1), p. 48. DOI: <https://doi.org/10.1038/s41597-020-0386-x>.
- [25] Taddei A., Distanto G., Emdin M., Moody G.B., Zeelenberg C. & Marchesi C. (1992) The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *European Heart Journal* 13(9), pp. 1164–1172. DOI: <https://doi.org/10.1093/oxfordjournals.eurheartj.a060332>.
- [26] Yang X., Liu G., Tong Y., Yan H., Xu Z., Chen Q., Liu X., Zhang H., Wang H. & Tan S. (2015) The history, hotspots, and trends of electrocardiogram. *Journal of Geriatric Cardiology* 12(4), pp. 448–456. DOI: <https://doi.org/10.11909/j.issn.1671-5411.2015.04.018>.
- [27] Wilson F.N., Johnston F.D., Macleod A. & S. B.P. (1934) Electrocardiograms that represent the potential variations of a single electrode. *The American Heart Journal* 9(4), pp. 447–458. DOI: [https://doi.org/10.1016/S0002-8703\(34\)90093-4](https://doi.org/10.1016/S0002-8703(34)90093-4).
- [28] Wilson F.N., Johnston F.D., Rosenbaum F.F., Erlanger H., Kossmann C.E., Hecht H., Cotrim N., de Oliveira R.M., Scarsi R. & Barker P.S. (1944) The Precordial Electrocardiogram. *The American Heart Journal* 27(1), pp. 19–85. DOI: [https://doi.org/10.1016/S0002-8703\(44\)90603-4](https://doi.org/10.1016/S0002-8703(44)90603-4).
- [29] Garcia T.B. (2015) *The 12-lead ECG: The art of interpretation*. Jones & Bartlett Learning, Burlington, Massachusetts, USA, 680 p.
- [30] Sattar Y. & Chhabra L. (2023) Electrocardiogram. In: *StatPearls [Internet]*, StatPearls Publishing. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK549803/>.
- [31] Kihlgren M., Almqvist C., Amankhani F., Jonasson L., Norman C., Perez M., Ebrahimi A. & Gottfridsson C. (2023) The U-wave: A remaining enigma of the electrocardiogram. *Journal of Electrocardiology* 79, pp. 13–20. DOI: <https://doi.org/10.1016/j.jelectrocard.2023.03.001>.
- [32] Wagner P., Strodthoff N., Bousseljot R. Kreiseler D., Lunze F.I., Wojciech S. & Schaeffter T. (2020) PTB-XL, a large publicly available electrocardiography

dataset. *Scientific Data* 7(154). DOI: <https://doi.org/10.1038/s41597-020-0495-6>.

- [33] Wagner P., Strodthoff N., Bousseljot R., Samek W. & Schaeffter T. (2022), PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3.). *PhysioNet*. DOI: <https://doi.org/10.13026/kfzx-aw45>. Accessed 14.7.2023.
- [34] Aslanger E. (2010) An unusual electrocardiogram artifact in a patient with near syncope. *Journal of Electrocardiology* 43(6), pp. 686–688. DOI: <https://doi.org/10.1016/j.jelectrocard.2010.04.012>.
- [35] Uzelac B. (2018) ECG artefacts in emergency medicine. *ABC* (Belgrade, Serbia) 18(3), pp. 41–48. DOI: <https://doi.org/10.5937/abc1803035B>.
- [36] Baranchuk A., Shaw C., Alanazi H., Campbell D., Bally K., Redfearn D.P., Simpson C.S. & Abdollah H. (2009) Electrocardiography Pitfalls and Artifacts: The 10 Commandments. *Critical Care Nurse* 29(1), pp. 67–73. DOI: <https://doi.org/10.4037/ccn2009607>.
- [37] Rudiger A., Hellermann J.P., Mukherjee R., Follath F. & Turina J. (2007) Electrocardiographic artifacts due to electrode misplacement and their frequency in different clinical settings. *The American Journal of Emergency Medicine* 25(2), pp. 174–178. DOI: <https://doi.org/10.1016/j.ajem.2006.06.018>.
- [38] Hedén B., Ohlsson M., Holst H., Mjöman M., Rittner R., Pahlm O., Peterson C. & Edenbrandt L. (1996) Detection of frequently overlooked electrocardiographic lead reversals using artificial neural networks. *The American Journal of Cardiology* 78(5), pp. 600–604. DOI: [https://doi.org/10.1016/S0002-9149\(96\)00377-3](https://doi.org/10.1016/S0002-9149(96)00377-3).
- [39] Rudiger A., Schöb L. & Follath F. (2003) Influence of electrode misplacement on the electrocardiographic signs of inferior myocardial ischemia. *The American Journal of Emergency Medicine* 21(7), pp. 574–577. DOI: <https://doi.org/10.1016/j.ajem.2003.08.007>.
- [40] Guijjaro-Morales A., Gil-Extremuera B. & Maldonado-Martín A. (1991) ECG diagnostic errors due to improper connection of the right arm and leg cables. *International Journal of Cardiology* 30(2), pp. 233–235. DOI: [https://doi.org/10.1016/0167-5273\(91\)90103-V](https://doi.org/10.1016/0167-5273(91)90103-V).
- [41] Shin S., Kim K., Song C., Lee J., Kim J. & Jeung G. (2015) Removal of baseline wandering in ECG signal by improved detrending method. *Bio-medical Materials and Engineering* 26(Suppl 1(s1)), pp. S1087–S1093. DOI: <https://doi.org/10.3233/BME-151405>.
- [42] Rahman M.A., Milu M.M.H., Anjum A., Khanam F. & Ahmad M. (2017) Baseline wandering removal from ecg signal by wandering path finding

- algorithm. In: 3rd International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh. DOI: <https://doi.org/10.1109/EICT.2017.8275164>.
- [43] Pallas-Areny R., Colominas-Balague J. & Rosell F.J. (1989) The Effect of Respiration-Induced Heart Movements on the ECG. *IEEE Transactions on Biomedical Engineering* 36(6), pp. 585–590. DOI: <https://doi.org/10.1109/10.29452>.
 - [44] Minetti A.E., Cisotti C. & Mian O.S. (2011) The mathematical description of the body centre of mass 3D path in human and animal locomotion. *Journal of Biomechanics* 44(8), pp. 1471–1477. DOI: <https://doi.org/10.1016/j.jbiomech.2011.03.014>.
 - [45] Lee C.R. & Farley C.T. (1998) Determinants of the center of mass trajectory in human walking and running. *Journal of Experimental Biology* 201(21), pp. 2935–2944. DOI: <https://doi.org/10.1242/jeb.201.21.2935>.
 - [46] Wahabi S., Pouryayevali S. & Hatzinakos D. (2015) Posture-invariant ecg recognition with posture detection. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1812–1816. DOI: <https://doi.org/10.1109/ICASSP.2015.7178283>.
 - [47] Chowdhury M.H. & Cheung R.C.C. (2019) Reconfigurable Architecture for Multi-lead ECG Signal Compression with High-frequency Noise Reduction. *Scientific Reports* 9(1), pp. 17233–12. DOI: <https://doi.org/10.1038/s41598-019-53460-3>.
 - [48] Agrawal S. & Gupta A. (2013) Fractal and EMD based removal of baseline wander and powerline interference from ECG signals. *Computers in Biology and Medicine* 43(11), pp. 1889–1899. DOI: <https://doi.org/10.1016/j.compbiomed.2013.07.030>.
 - [49] Mann H. (1920) A method for analyzing the electrocardiogram. *Archives of Internal Medicine* (1960) 25(3), pp. 283–294. DOI: <https://doi.org/10.1001/archinte.1920.00090320054004>.
 - [50] Vondrak J. & Penhaker M. (2022) Review of Processing Pathological Vectorcardiographic Records for the Detection of Heart Disease. *Frontiers in Physiology* 13, p. 856590. DOI: <https://doi.org/10.3389/fphys.2022.856590>.
 - [51] Chou T. (1986) When is the vectorcardiogram superior to the scalar electrocardiogram? *Journal of the American College of Cardiology* 8(4), pp. 791–799. DOI: [https://doi.org/10.1016/S0735-1097\(86\)80419-3](https://doi.org/10.1016/S0735-1097(86)80419-3).
 - [52] Gamboa R., Hugenholtz P.G. & Nadas A.S. (1965) Comparison of Electrocardiograms and Vectorcardiograms in Congenital Aortic Stenosis. *British Heart Journal* 27(3), pp. 344–354. DOI: <https://doi.org/10.1136/hrt.27.3.344>.

- [53] Pérez-Riera A.R., Uchida A.H., Filho C.F., Meneghini A., Ferreira C., Schapacknik E., Dubner S. & Moffa P. (2007) Significance of Vectorcardiogram in the Cardiological Diagnosis of the 21st Century. *Clinical Cardiology* 30(7), pp. 319–323. DOI: <https://doi.org/10.1002/clc.14>.
- [54] Dellborg M., Malmberg K., Rydén L., Svensson A. & Swedberg K. (1995) Dynamic On-Line Vectorcardiography Improves and Simplifies In-Hospital Ischemia Monitoring of Patients With Unstable Angina. *Journal of the American College of Cardiology* 26(6), pp. 1501–1507. DOI: [https://doi.org/10.1016/0735-1097\(95\)00361-4](https://doi.org/10.1016/0735-1097(95)00361-4).
- [55] Pérez-Riera A.R., Barbosa-Barros R., Daminello-Raimundo R., de Abreu L.C. & Nikus K. (2021) The Vectorcardiogram and the Main Dromotropic Disturbances. *Current Cardiology Reviews* 17(1), pp. 50–59. DOI: <https://doi.org/10.2174/1573403X16666200810105504>.
- [56] Talbot S., Dreifus L.S., Watanabe Y., Chiang R., Morris K. & Reich M. (1976) Diagnostic criteria for computer-aided electrocardiographic 15-lead system. Evaluation using 12 leads and Frank orthogonal leads with vector display. *British Heart Journal* 38(12), pp. 1247–1261. DOI: <https://doi.org/10.1136/hrt.38.12.1247>.
- [57] Frank E. (1956) An Accurate, Clinically Practical System For Spatial Vectorcardiography. *Circulation (New York)* 13(5), pp. 737–749. DOI: <https://doi.org/10.1161/01.CIR.13.5.737>.
- [58] Maheshwari S., Acharyya A., Schiariti M. & Puddu P.E. (2016) Frank vectorcardiographic system from standard 12 lead ECG: An effort to enhance cardiovascular diagnosis. *Journal of Electrocardiology* 49(2), pp. 231–242. DOI: <https://doi.org/10.1016/j.jelectrocard.2015.12.008>.
- [59] Jaros R., Martinek R. & Danys L. (2019) Comparison of Different Electrocardiography with Vectorcardiography Transformations. *Sensors (Basel, Switzerland)* 19(14), p. 3072. DOI: <https://doi.org/10.3390/s19143072>.
- [60] Schmitt O.H. & Simonson E. (1955) The Present Status of Vectorcardiography. *A.M.A. archives of internal medicine* 96(5), pp. 574–590. DOI: <https://doi.org/10.1001/archinte.1955.00250160016002>.
- [61] McFee R. & Parungao A. (1961) An orthogonal lead system for clinical electrocardiography. *American Heart Journal* 62(1), pp. 93–100. DOI: [https://doi.org/10.1016/0002-8703\(61\)90488-4](https://doi.org/10.1016/0002-8703(61)90488-4).
- [62] Burger H.C., Van Milaan J.B. & Den Boer W. (1952) Comparison of Different Systems of Vectorcardiography. *British Heart Journal* 14(3), pp. 401–405. DOI: <https://doi.org/10.1136/hrt.14.3.401>.

- [63] Kors J.A., van Herpen G., Sittig A.C. & van Bommel J.H. (1990) Reconstruction of the Frank vectorcardiogram from standard electrocardiographic leads: diagnostic comparison of different methods. *European Heart Journal* 11(12), pp. 1083–1092. DOI: <https://doi.org/10.1093/oxfordjournals.eurheartj.a059647>.
- [64] Guillem M.S., Sahakian A.V. & S. S. (2006) Derivation of orthogonal leads from the 12-lead ecg. accuracy of a single transform for the derivation of atrial and ventricular waves. In: *Computers in Cardiology, Valencia, Spain*, pp. 249–252.
- [65] Dower G.E., Machado H.B. & Osborne J.A. (1980) On Deriving the Electrocardiogram from Vectorcardiogram Leads. *Clinical Cardiology* 3(2), pp. 87–95. DOI: <https://doi.org/10.1002/clc.1980.3.2.87>.
- [66] Edenbrandt L. & Pahlm O. (1988) Vectorcardiogram synthesized from a 12-lead ECG: Superiority of the inverse Dower matrix. *Journal of Electrocardiology* 21(4), pp. 361–367. DOI: [https://doi.org/10.1016/0022-0736\(88\)90113-6](https://doi.org/10.1016/0022-0736(88)90113-6).
- [67] Edenbrandt L., Houston A. & Macfarlane P.W. (1994) Vectorcardiograms synthesized from 12-lead ECGs: A new method applied in 1792 healthy children. *Pediatric Cardiology* 15(1), pp. 21–26. DOI: <https://doi.org/10.1007/BF00797001>.
- [68] Wulan N., Wang W., Sun P., Wang K., Xia Y. & Zhang H. (2020) Generating electrocardiogram signals by deep learning. *Neurocomputing* 404, pp. 122–136. DOI: <https://doi.org/10.1016/j.neucom.2020.04.076>.
- [69] McSharry P.E., Clifford G.D., Tarassenko L. & Smith L.A. (2003) A Dynamical Model for Generating Synthetic Electrocardiogram Signals. *IEEE Transactions on Biomedical Engineering* 50(3), pp. 289–294. DOI: <https://doi.org/10.1109/TBME.2003.808805>.
- [70] Yang H., Liu J., Zhang L., Li Y. & Zhang H. (2021) ProEGAN-MS: A Progressive Growing Generative Adversarial Networks for Electrocardiogram Generation. *IEEE Access* 9, pp. 52089–52100. DOI: <https://doi.org/10.1109/ACCESS.2021.3069827>.
- [71] Xia Y., Wang W. & Wang K. (2023) ECG signal generation based on conditional generative models. *Biomedical Signal Processing and Control* 82, p. 104587. DOI: <https://doi.org/10.1016/j.bspc.2023.104587>.
- [72] Adib E., Fernandez A.S., Afghah F. & Prevost J.J. (2023) Synthetic ECG Signal Generation using Probabilistic Diffusion Models. *IEEE Access* , pp. 1–1 DOI: <https://doi.org/10.1109/ACCESS.2023.3296542>.
- [73] Thayer J.F., Åhs F., Fredrikson M., Sollers III J.J. & Wager T.D. (2012) A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Reviews* 36(2), pp. 747–756. DOI: <https://doi.org/10.1016/j.neubiorev.2011.11.009>.

- [74] Acharya R.U., Joseph K.P., Kannathal N., Lim C.M. & Suri J.S. (2006) Heart rate variability: a review. *Medical Biological Engineering Computing* 44(12), pp. 1031–1051. DOI: <https://doi.org/10.1007/s11517-006-0119-0>.
- [75] Berntson G.G., Bigger Jr. T., Eckberg D.L., Grossman P., Kaufmann P.G., Malik M., Nagaraja H.N., Porges S.W., Saul J.P., Stone P.H. & Van Der Molen M.W. (1997) Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology* 34(6), pp. 623–648. DOI: <https://doi.org/10.1111/j.1469-8986.1997.tb02140.x>.
- [76] Ernst G. (2017) Heart-Rate Variability - More than Heart Beats? *Frontiers in Public Health* 5, p. 240. DOI: <https://doi.org/10.3389/fpubh.2017.00240>.
- [77] Schmidt G., Malik M., Barthel P., Schneider R., Ulm K., Rolnitzky L., Camm A.J., Bigger Jr. J.T. & Schömig A. (1999) Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction. *The Lancet (British Edition)* 353(9162), pp. 1390–1396. DOI: [https://doi.org/10.1016/S0140-6736\(98\)08428-1](https://doi.org/10.1016/S0140-6736(98)08428-1).
- [78] Carney R.M., Blumenthal J.A., Stein P.K., Watkins L., Catellier D., Berkman L.F., Czajkowski S.M., O'Connor C., Stone P.H. & Freedland K.E. Depression, Heart Rate Variability, and Acute Myocardial Infarction. *Circulation* 104(17), pp. 2024–2028. DOI: <https://doi.org/10.1161/hc4201.097834>.
- [79] Carney R.M., Blumenthal J.A., Freedland K.E., Stein P.K., Howells W.B., Berkman L.F., Watkins L.L., Czajkowski S.M., Hayano J., Domitrovich P.P. & Jaffe A.S. (2005) Low Heart Rate Variability and the Effect of Depression on Post-Myocardial Infarction Mortality. *Archives of Internal Medicine* 165(13), pp. 1486–1491. DOI: <https://doi.org/10.1001/archinte.165.13.1486>.
- [80] Bonnemeier H., Wiegand U.K.H., Brandes A., Kluge N., Katus H.A., Richardt G. & Potratz J. (2003) Circadian Profile of Cardiac Autonomic Nervous Modulation in Healthy Subjects: Differing Effects of Aging and Gender on Heart Rate Variability. *Journal of Cardiovascular Electrophysiology* 14(8), pp. 791–799. DOI: <https://doi.org/10.1046/j.1540-8167.2003.03078.x>.
- [81] Nagy E., Orvos H., Bárdos G. & Molnár P. (2000) Gender-Related Heart Rate Differences in Human Neonates. *Pediatric Research* 47(6), pp. 778–780. DOI: <https://doi.org/10.1203/00006450-200006000-00016>.
- [82] Kiviniemi A.M., Hautala A.J., Kinnunen H., Nissilä J., Virtanen P., Karjalainen J. & Tulppo M.P. (2010) Daily Exercise Prescription on the Basis of HR Variability among Men and Women. *Medicine & Science in Sports & Exercise* 42(7), pp. 1355–1363. DOI: <https://doi.org/10.1249/MSS.0b013e3181cd5f39>.

- [83] Camm A.J., Malik M., Bigger Jr J.T., Breithardt G., Cerutti S., Cohen R.J., Coumel P., Fallen E.L., Kennedy H.L., Kleiger R.E., Lombardi F., Malliani A., Moss A.J., Rottman J.N., Schmidt G., Schwartz P.J. & Singer D.H. (1996) Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *Circulation* 93(5), pp. 1043–1065. DOI: <https://doi.org/10.1161/01.cir.93.5.1043>.
- [84] Laitio T., Jalonen J., Kuusela T. & Scheinin H. (2007) The Role of Heart Rate Variability in Risk Stratification for Adverse Postoperative Cardiac Events. *Anesthesia and Analgesia* 105(6), pp. 1548–1560. DOI: <https://doi.org/10.1213/01.ane.0000287654.49358.3a>.
- [85] Tarvainen M.P., Niskanen J., Lipponen J.A., Ranta-aho P.O. & Karjalainen P.A. (2014) Kubios HRV - Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine* 113(1), pp. 210–220. DOI: <https://doi.org/10.1016/j.cmpb.2013.07.024>.
- [86] Pagani M., Montano N., Porta A., Malliani A., Abboud F.M., Birkett C. & Somers V.K. (1997) Relationship Between Spectral Components of Cardiovascular Variabilities and Direct Measures of Muscle Sympathetic Nerve Activity in Humans. *Circulation* 95(6), pp. 1441–1448. DOI: <https://doi.org/10.1161/01.CIR.95.6.1441>.
- [87] Furlan R., Porta A., Costa F., Tank J., Baker L., Schiavi R., Robertson D., Malliani A. & Mosqueda-Garcia R. (2000) Oscillatory Patterns in Sympathetic Neural Discharge and Cardiovascular Variables During Orthostatic Stimulus. *Circulation* 101(8), pp. 886–892. DOI: <https://doi.org/10.1161/01.CIR.101.8.886>.
- [88] Akselrod S., Gordon D., Ubel F.A., Shannon D.C., Berger A.C. & Cohen R.J. (1981) Power Spectrum Analysis of Heart Rate Fluctuations: A Quantitative Probe of Beat-to-Beat Cardiovascular Control. *American Association for the Advancement of Science* 213(4504), pp. 220–222. DOI: <https://doi.org/10.1126/science.6166045>.
- [89] Claydon V.E. & Krassioukov A.V. (2008) Clinical correlates of frequency analyses of cardiovascular control after spinal cord injury. *American Journal of Physiology* 294(2), pp. 668–678. DOI: <https://doi.org/10.1152/ajpheart.00869.2007>.
- [90] Hadase M., Azuma A., Zen K., Asada S., Kawasaki T., Kamitani T., Kawasaki S., Sugihara H. & Matsubara H. (2004) Very Low Frequency Power of the Heart Rate Variability is a Powerful Predictor of Clinical Prognosis in Patients With Congestive Heart Failure. *Circulation Journal* 68(4), pp. 343–347. DOI: <https://doi.org/10.1253/circj.68.343>.
- [91] Schmidt H., Müller-Werdan U., Hoffmann T., Francis D.P., Piepoli M.F., Rauchhaus M., Prondzinsky R., Loppnow H., Buerke M., Hoyer D. & Werdan K. (2005) Autonomic dysfunction predicts mortality in patients with multiple organ dysfunction syndrome of different age groups. *Critical Care*

- Medicine 33(9), pp. 1994–2002. DOI: <https://doi.org/10.1097/01.CCM.0000178181.91250.99>.
- [92] Sztajzel J. (2004) Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Medical Weekly* 134, pp. 514–522.
- [93] Xie L., Li Z., Zhou Y., He Z. & Zhu J. (2020) Computational Diagnostic Techniques for Electrocardiogram Signal Analysis. *Sensors (Basel)* 20(21), p. 6318. DOI: <https://doi.org/10.3390/s20216318>.
- [94] Spiers C. & Stinchcombe E. (2002) An introduction to cardiac monitoring and rhythm interpretation. *British Journal of Anaesthetic and Recovery Nursing* 3(1), pp. 8–14. DOI: <https://doi.org/10.1017/S1742645600000760>.
- [95] Pilhall M., Riha M. & Jern S. (1992) Exercise-induced QRS changes in healthy men and women: a multivariate analysis on their relation to background data and exercise performance. *European Heart Journal* 13(10), pp. 1316–1324. DOI: <https://doi.org/10.1093/oxfordjournals.eurheartj.a060060>.
- [96] Da Costa D., Brady W.J. & Edhouse J. (2002) ABC of Clinical Electrocardiography: Bradycardias and Atrioventricular Conduction Block. *BMJ* 324(7336), pp. 535–538. DOI: <https://doi.org/10.1136/bmj.324.7336.535>.
- [97] Goodacre S. & Irons R. (2002) ABC of Clinical Electrocardiography: Atrial Arrhythmias. *BMJ* 324(7337), pp. 594–597. DOI: <https://doi.org/10.1136/bmj.324.7337.594>.
- [98] Coviello J.S. (2017) ECG interpretation made incredibly easy!: Pocket guide (third edition). Wolters Kluwer, Philadelphia, USA, 309 p.
- [99] Martis R.J., Acharya U.R. & Adeli H. (2014) Current methods in electrocardiogram characterization. *Computers in Biology and Medicine* 48(1), pp. 133–149. DOI: <https://doi.org/10.1016/j.compbiomed.2014.02.012>.
- [100] Leis J.W. (2011) Digital Signal Processing Using MATLAB for Students and Researchers. John Wiley & Sons, Inc., New Jersey, USA, 396 p.
- [101] Bailey J.J., Berson A.S., Garson Jr A., Horan L.G., Macfarlane P.W., Mortara D.W. & Zywiec C. (1990) Recommendations for standardization and specifications in automated electrocardiography: Bandwidth and digital signal processing. A report for health professionals by an Ad Hoc writing group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology, American Heart Association. *Circulation* 81(2), pp. 730–739. DOI: <https://doi.org/10.1161/01.CIR.81.2.730>.

- [102] Luo S. & Johnston P. (2010) A review of electrocardiogram filtering. *Journal of Electrocardiology* 43(6), pp. 486–496. DOI: <https://doi.org/10.1016/j.jelectrocard.2010.07.007>.
- [103] Savitzky A. & Golay M.J.E. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 36(8), pp. 1627–1639. DOI: <https://doi.org/10.1021/ac60214a047>.
- [104] Kawala-Sterniuk A., Podpora M., Pelc M., Blaszczyzyn M., Gorzelanczyk E.J., Martinek R. & Ozana S. (2020) Comparison of Smoothing Filters in Analysis of EEG Data for the Medical Diagnostics Purposes. *Sensors* 20(3), p. 807. DOI: <https://doi.org/10.3390/s20030807>.
- [105] Benitez D., Gaydecki P.A., Zaidi A. & Fitzpatrick A.P. (2001) The use of the Hilbert transform in ECG signal analysis. *Computers in Biology and Medicine* 31(5), pp. 399–406. DOI: [https://doi.org/10.1016/S0010-4825\(01\)00009-9](https://doi.org/10.1016/S0010-4825(01)00009-9).
- [106] Afonso V.X., Tompkins W.J., Nguyen T.Q. & Luo S. (1999) ECG Beat Detection Using Filter Banks. *IEEE Transactions on Biomedical Engineering* 46(2), pp. 192–202. DOI: <https://doi.org/10.1109/10.740882>.
- [107] Pan J. & Tompkins W.J. (1985) A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering* BME-32(3), pp. 230–236. DOI: <https://doi.org/10.1109/TBME.1985.325532>.
- [108] Hamilton P.S. & Tompkins W.J. (1986) Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database. *IEEE Transactions on Biomedical Engineering* BME-33(12), pp. 1157–1165. DOI: <https://doi.org/10.1109/TBME.1986.325695>.
- [109] Paoletti M. & Marchesi C. (2006) Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis. *Computer Methods and Programs in Biomedicine* 82(1), pp. 20–30. DOI: <https://doi.org/10.1016/j.cmpb.2006.01.005>.
- [110] Christov I.I. (2004) Real time electrocardiogram QRS detection using combined adaptive threshold. *Biomedical Engineering Online* 3(1), p. 28. DOI: <https://doi.org/10.1186/1475-925X-3-28>.
- [111] Chen S., Chen H. & Chan H. (2006) A real-time QRS detection method based on moving-averaging incorporating with wavelet denoising. *Computer Methods and Programs in Biomedicine* 82(3), pp. 187–195. DOI: <https://doi.org/10.1016/j.cmpb.2005.11.012>.
- [112] Poli R., Cagnoni S. & Valli G. (1995) Genetic Design of Optimum Linear and Nonlinear QRS Detectors. *IEEE Transactions on Biomedical Engineering* 42(11), pp. 1137–1141. DOI: <https://doi.org/10.1109/10.469381>.

- [113] Lee J., Kim K., Lee B., Lee B. & Lee M. (2002) A Real Time QRS Detection Using Delay-Coordinate Mapping for the Microcontroller Implementation. *Annals of Biomedical Engineering* 30(9), pp. 1140–1151. DOI: <https://doi.org/10.1114/1.1523030>.
- [114] Cvikl M., Jager F. & Zemva A. (2007) Hardware Implementation of a Modified Delay-Coordinate Mapping-Based QRS Complex Detection Algorithm. *EURASIP Journal on Advances in Signal Processing* 2007(1), pp. 1–13. DOI: <https://doi.org/10.1155/2007/57286>.
- [115] Choi S., Adnane M., Lee G., Jang H., Jiang Z. & Park H. (2010) Development of ECG beat segmentation method by combining lowpass filter and irregular R-R interval checkup strategy. *Expert Systems with Applications* 37(7), pp. 5208–5218. DOI: <https://doi.org/10.1016/j.eswa.2009.12.069>.
- [116] Pal S. & Mitra M. (2012) Empirical mode decomposition based ECG enhancement and QRS detection. *Computers in Biology and Medicine* 42(1), pp. 83–92. DOI: <https://doi.org/10.1016/j.combiomed.2011.10.012>.
- [117] Altman N. & Krzywinski M. (2017) Clustering. *Nature Methods* 14(6), pp. 545–546. DOI: <https://doi.org/10.1038/nmeth.4299>.
- [118] Jain A.K. (2010) Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31(8), pp. 651–666. DOI: <https://doi.org/10.1016/j.patrec.2009.09.011>.
- [119] Xu R. & Wunsch D. (2009) Clustering. IEEE Press, New Jersey, USA, 368 p.
- [120] Lloyd S.P. (1982) Least Squares Quantization in PCM. *IEEE Transactions of Information Theory* 28(2), pp. 129–137. DOI: <https://doi.org/10.1109/TIT.1982.1056489>.
- [121] Benabdellah A.C., Benghabrit A. & Bouhaddou I. (2019) A survey of clustering algorithms for an industrial context. *Procedia Computer Science* 148, pp. 291–302. DOI: <https://doi.org/10.1016/j.procs.2019.01.022>.
- [122] Yin J., Sun H., Yang J. & Guo Q. (2014) Comparison of K-means and Fuzzy c-Means Algorithm Performance for Automated Determination of the Arterial Input Function. *PloS One* 9(2), p. e85884. DOI: <https://doi.org/10.1371/journal.pone.0085884>.
- [123] Xu R. & Wunsch D. (2005) Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks* 16(3), pp. 645–678. DOI: <https://doi.org/10.1109/TNN.2005.845141>.
- [124] Ezugwu A.E., Ikotun A.M., Oyelade O.O., Abuligah L., Agushaka J.O., Eke C.I. & Akinyelu A.A. (2022) A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence* 110,

p. 104743. DOI: <https://doi.org/10.1016/j.engappai.2022.104743>.

- [125] Xu D. & Tian Y. (2015) A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science* 2(2), pp. 165–193. DOI: <https://doi.org/10.1007/s40745-015-0040-1>.
- [126] Lipponen J.A. & Tarvainen M.P. (2019) A robust algorithm for heart rate variability time series artefact correction using novel beat classification. *Journal of Medical Engineering & Technology* 43(3), pp. 173–181. DOI: <https://doi.org/10.1080/03091902.2019.1640306>.
- [127] Arthur D. & Vassilvitskii S. (2007) K-means++: The advantages of careful seeding. In: *SODA '07: Proceeding of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035.
- [128] Meyer C., Gavela J.F. & Harris M. (2006) Combining Algorithms in Automatic Detection of QRS Complexes in ECG Signals. *IEEE Transactions on Information Technology in Biomedicine* 10(3), pp. 468–475. DOI: <https://doi.org/10.1109/TITB.2006.875662>.

9. APPENDICES

Appendix 1	Graphical User Interface, First Window Overview
Appendix 2	Graphical User Interface, First Window Subtabs
Appendix 3	Graphical User Interface, First Window Components
Appendix 4	Graphical User Interface, Second Window Overview
Appendix 5	Graphical User Interface, Second Window Components
Appendix 6	Overall Performance of the Beat Detection Algorithm

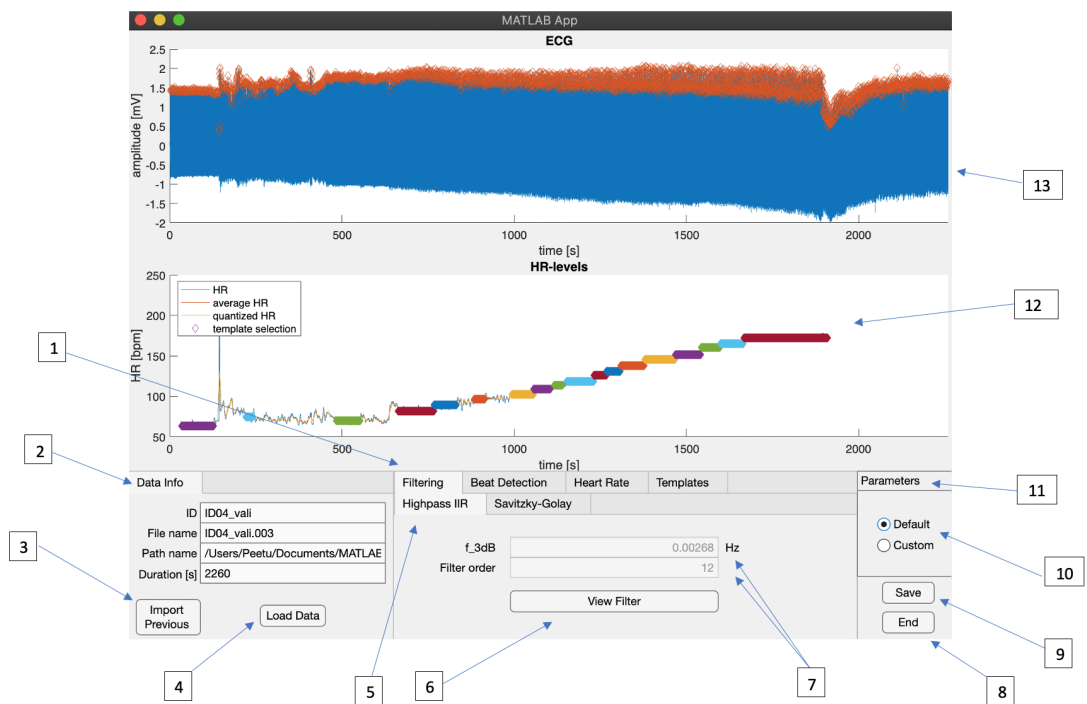


Figure A1.1. The main view of the first window of the graphical user interface. The components are marked with numbers.

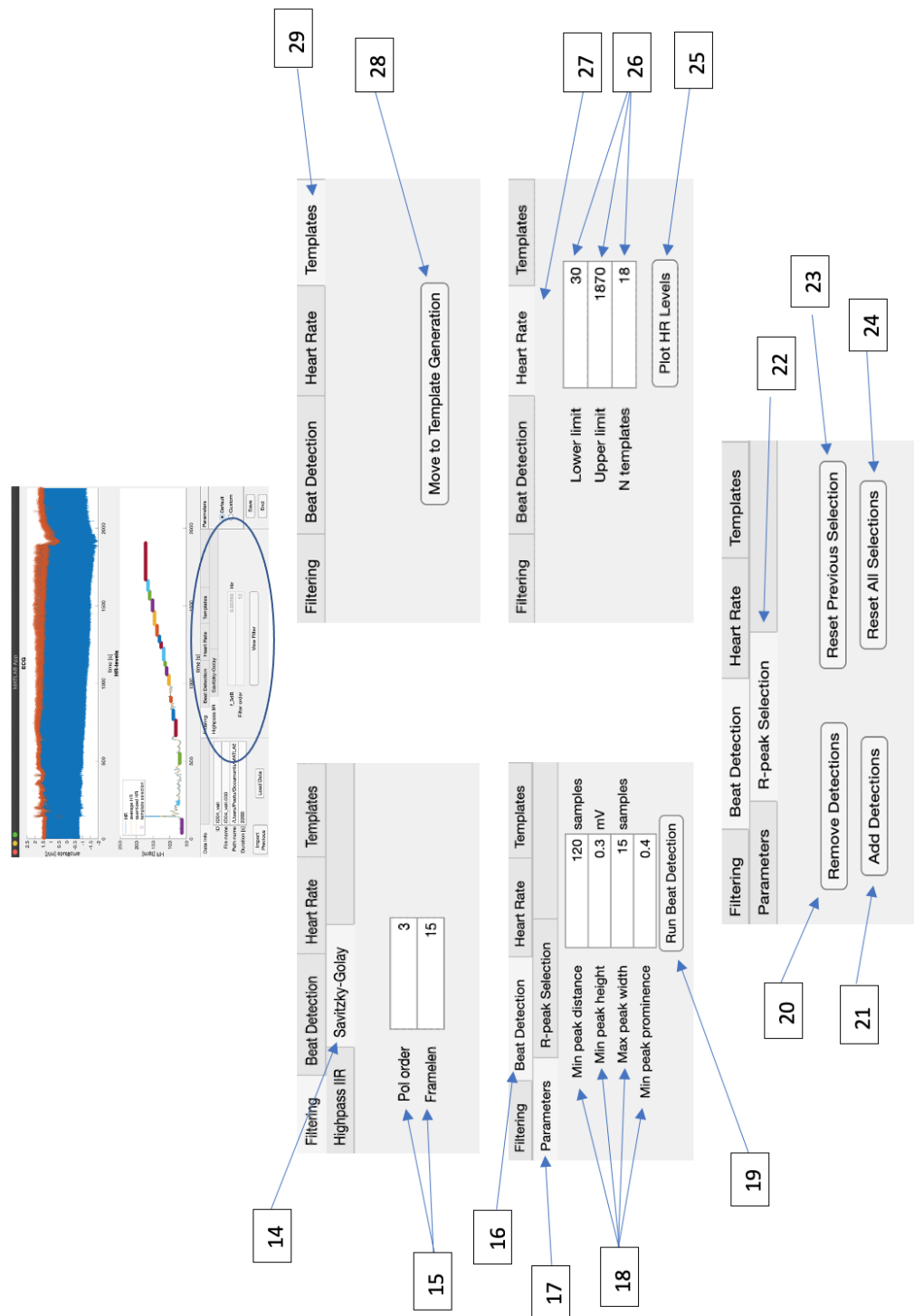
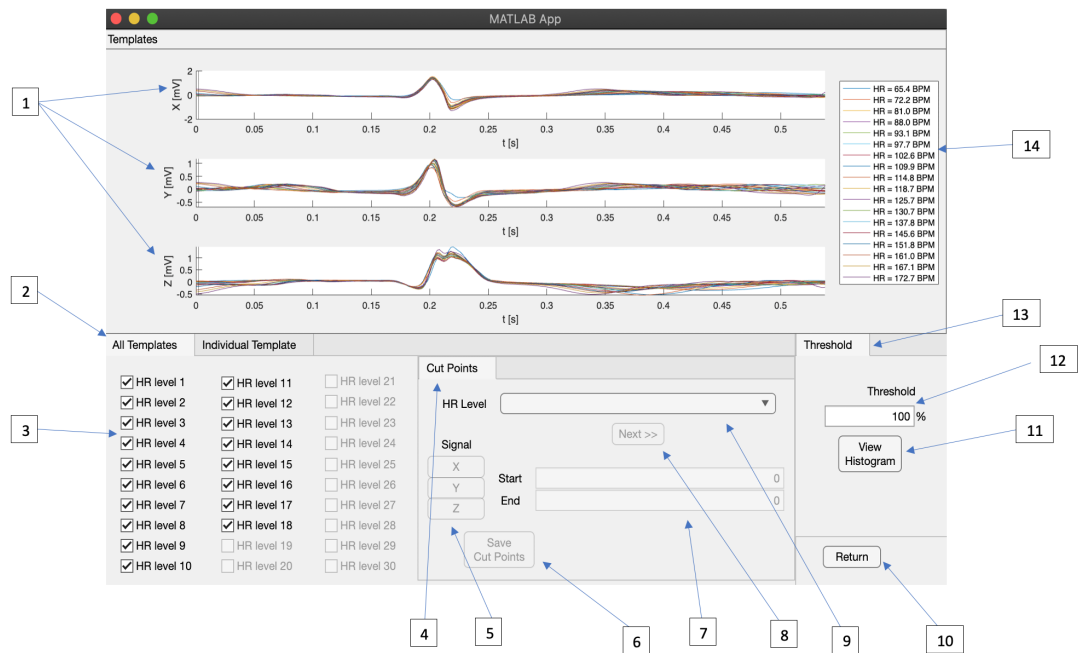


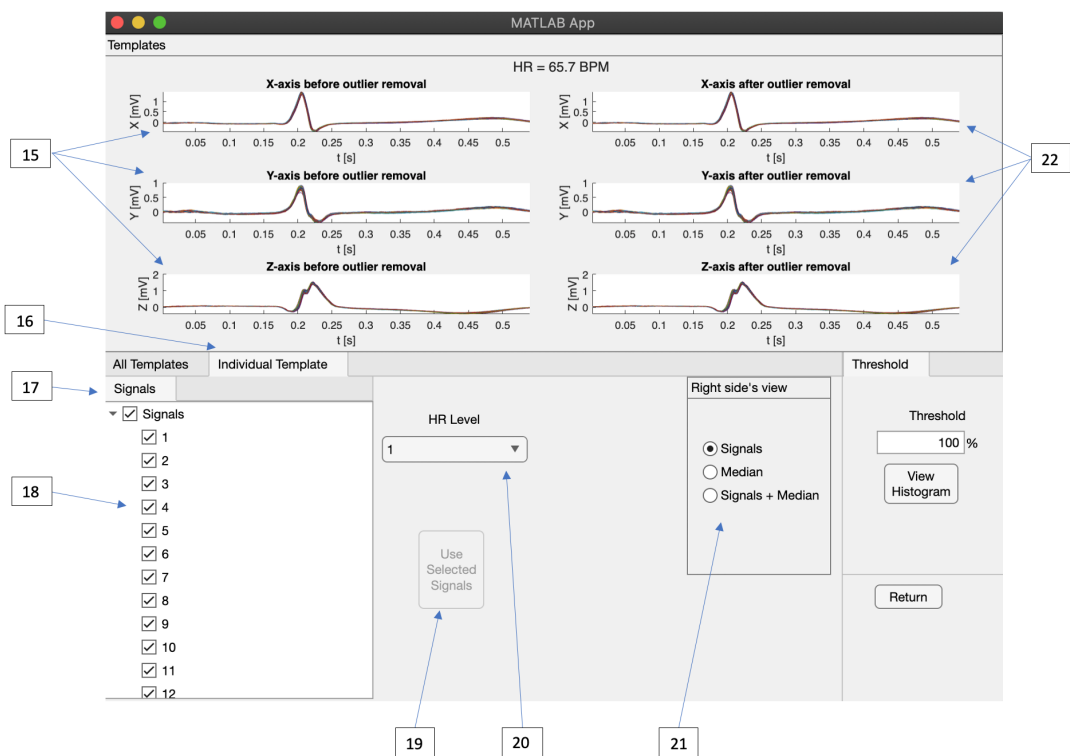
Figure A2.1. The subtabs of the first window. Each subtab contains properties for a specific process.

Table A3.1. The components of the first window of the graphical user interface

Index	Component	Information
1	Filtering tab	Contains filtering parameters in two subtabs
2	Data Info section	Contains information about the data
3	Import Previous button	Downloads settings from the previous session
4	Load Data button	Downloads the new ECG data
5	Highpass IIR subtab	Contains the parameters used in high-pass IIR filter
6	View Filter button	Opens a new window for filter visualization
7	Highpass IIR parameters	The parameters used in the high-pass IIR filter
8	End button	Closes the app
9	Save button	Saves the parameters to a .MAT file
10	Parameter editing options	Default: default parameters. Custom: enables parameter customization
11	Parameters section	Contains parameter editing options
12	Heart rate graph	Heart rate as a function of time. HR-levels showed with colors
13	ECG plot	ECG as a function of time. Detected beats showed with red marks
14	Savitzky-Golay subtab	Contains the parameters used in the Savitzky-Golay filter
15	Savitzky-Golay parameters	Parameters used in the Savitzky-Golay filter
16	Beat Detection tab	Contains beat detection parameters and manual options
17	Parameters subtab	Contains the parameters used in beat detection
18	Beat detection parameters	Parameters used in beat detection
19	Run Beat Detection button	Runs beat detection and shows the ECG signal with detections
20	Remove Detections button	Enables the user to remove beat detections manually
21	Add Detections button	Enables the user to add beat detections manually
22	R-peak Selection subtab	Contains manual beat detection options
23	Reset Previous Selection button	Resets the previous manual beat detection function
24	Reset All Selections button	Resets all manual beat detection functions
25	Plot HR levels button	Runs heart rate level clustering and plots the heart rate graph
26	Heart rate parameters	Parameters used in heart rate level clustering
27	Heart Rate tab	Contains the parameters used in heart rate level clustering
28	Move to Template Generation button	Opens the template generation window
29	Templates tab	Contains the button to template generation



(a)



(b)

Figure A4.1. Two views of the second window of the graphical user interface. a) The first view, where the final templates are plotted to the top of the window b) The second view, where the individual signals of a specific template are visible before (left) and after (right) the outlier detection algorithm.

Table A5.1. The components of the second window of the graphical user interface

Index	Component	Information
1	VCG templates	VCG template signals. Cut points are shown as lines on the x-axis
2	All Templates tab	Contains HR level selection and cut point customization properties
3	HR level checkboxes	All HR levels. Checked boxes are displayed
4	Cut Points section	Contains cut point customization options
5	Cut point selection buttons	Enables the user to select cut points
6	Save Cut Points button	Saves the selected cut points
7	Cut point edit fields	Shows current cut points. The user can edit the cut points by typing in the field
8	Next » button	Displays the next template
9	HR level dropdown	The list of all HR levels. HR level can be displayed by selecting it from the list
10	Return button	Returns to the first window and stores template setting into temporary memory
11	View Histogram button	Opens a new window for the histogram visualization
12	Threshold edit field	The user can edit the threshold value by typing in the field
13	Threshold section	Contains threshold customization and histogram visualization options
14	HR level legends	Shows template colors and respective HR levels that are currently displayed
15	VCG signals	X, Y, and Z signals before outlier detection algorithm
16	Individual Template tab	Contains signal selection options for a specific template
17	Signals section	Contains all signals for a specific template
18	Signal checkboxes	All signals for a specific template. Checked signals are displayed
19	Use Selected Signals button	Uses selected signals for outlier detection calculations
20	HR Level dropdown	The list of all HR levels. The user can display respective signals by selecting HR level
21	Right side's view section	Display options for the right side's plots
22	VCG signals	X, Y, and Z signals after the outlier detection algorithm

Table A6.1. The overall performance of the beat detection algorithm

Record	ANN	TP	FP	FN	FP + FN	SEN (%)	PRE (&)	DER (%)
100	2273	2273	0	0	0	100.00	100.00	0.00
101	1865	1865	1	0	1	100.00	99.95	0.05
102	2187	2187	0	0	0	100.00	100.00	0.00
103	2084	2084	0	0	0	100.00	100.00	0.00
104	2229	2226	4	3	7	99.87	99.82	0.31
105	2572	2554	24	18	42	99.30	99.07	1.63
106	2027	2020	10	7	17	99.65	99.51	0.84
107	2137	2133	3	4	7	99.81	99.86	0.33
108	1763	1749	9	14	23	99.21	99.49	1.30
109	2532	2532	0	0	0	100.00	100.00	0.00
111	2124	2124	0	0	0	100.00	100.00	0.00
112	2539	2539	0	0	0	100.00	100.00	0.00
113	1795	1795	2	0	2	100.00	99.89	0.11
114	1879	1879	2	0	2	100.00	99.89	0.11
115	1953	1953	0	0	0	100.00	100.00	0.00
116	2412	2398	2	14	16	99.42	99.92	0.66
117	1535	1535	0	0	0	100.00	100.00	0.00
118	2278	2278	6	0	6	100.00	99.74	0.26
119	1987	1987	1	0	1	100.00	99.95	0.05
121	1863	1863	0	0	0	100.00	100.00	0.00
122	2476	2476	0	0	0	100.00	100.00	0.00
123	1518	1515	0	3	3	99.80	100.00	0.20
124	1619	1596	9	23	32	98.58	99.44	1.98
200	2601	2595	23	6	29	99.77	99.12	1.12
201	1963	1958	7	5	12	99.75	99.64	0.61
202	2136	2135	8	1	9	99.95	99.63	0.42
203	2980	2932	46	48	94	98.39	98.46	3.15
205	2656	2652	2	4	6	99.85	99.92	0.23
207	1860	1816	89	44	133	97.63	95.33	7.15
208	2955	2946	14	9	23	99.70	99.53	0.78
209	3005	3005	0	0	0	100.00	100.00	0.00
210	2650	2646	18	4	22	99.85	99.32	0.83
212	2748	2748	0	0	0	100.00	100.00	0.00
213	3251	3251	0	0	0	100.00	100.00	0.00
214	2262	2259	4	3	7	99.87	99.82	0.31
215	3363	3363	8	0	8	100.00	99.76	0.24
217	2208	2206	1	2	3	99.91	99.95	0.14
219	2154	2148	32	6	38	99.72	98.53	1.76
220	2048	2040	3	8	11	99.61	99.85	0.54
221	2427	2427	4	0	4	100.00	99.84	0.16
222	2483	2480	13	3	16	99.88	99.48	0.64
223	2605	2604	2	1	3	99.96	99.92	0.12
228	2053	2039	20	14	34	99.32	99.03	1.66
230	2256	2256	0	0	0	100.00	100.00	0.00
231	1571	1570	7	1	8	99.94	99.56	0.51
232	1780	1777	7	3	10	99.83	99.61	0.56
233	3079	3079	1	0	1	100.00	99.97	0.03
234	2753	2753	0	0	0	100.00	100.00	0.00
48 records	109494	109246	382	248	630	99.77	99.65	0.58