Czech Technical University in Prague
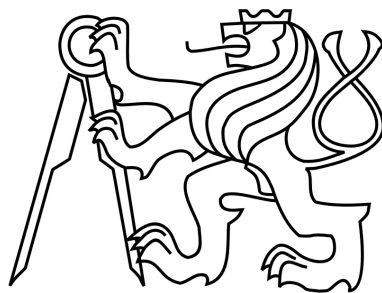
Faculty of Electrical Engineering

Diploma Thesis
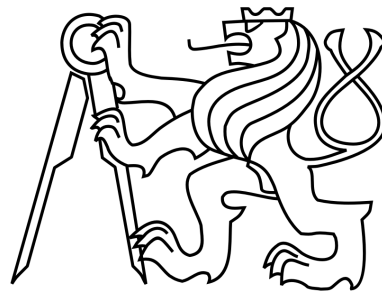
2009   Ondřej Mašek

Czech Technical University in Prague

Faculty of Electrical Engineering
Department of Cybernetics

Heart Rate Variability Analysis

May 2009   Author: Ondřej Mašek

                Supervisor: Ing. Václav Chudáček

Prohlášení

Prohlašuji, že jsem svou diplomovou .práci vypracoval samostatně a použil jsem pouze podklady (literaturu, projekty, SW atd.) uvedené v přiloženém seznamu.

V Praze dne .11. 5. 2009.....   ....
                                                         podpis

i

I would like to thank Václav Chudáček, my supervisor, for his many suggestions and constant support in my work.

Prague, Czech Republic

Ondřej Mašek

May, 2009

České vysoké učení technické v Praze
Fakulta elektrotechnická

Katedra kybernetiky

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Ondřej M a š e k

Studijní program: Elektrotechnika a informatika (magisterský), strukturovaný

Obor: Biomedicínské inženýrství

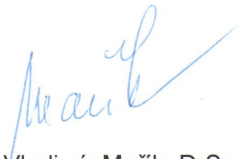**Název tématu:**   Analýza variability srdečního rytmu

Pokyny pro vypracování:

1. Seznamte se s problematikou analýzy variability srdečního rytmu.
2. Navrhněte několik možných metod analýzy variability srdečního rytmu.
3. Na základě dat z 1.LF UK, případně z MIT databáze vyhodnoťte a porovnejte navržené metody.

Seznam odborné literatury: Dodá vedoucí práce.

Vedoucí diplomové práce: Ing. Václav Chudáček

Platnost zadání: do konce zimního semestru 2008/2009

prof. Ing. Vladimír Mařík, DrSc.
vedoucí katedry

L.S.

doc. Ing. Boris Šimák, CSc.
děkan

V Praze dne 3.9.2007

Abstract

This diploma thesis deals with assessment of different heart rate analysis methods for distinguishing between atrial fibrilation and normal sinus rhythm. Many linear and nonlinear methods, such as NN50, pNN50, RMSSD, SDNN, SDSD, short term variability, long term irregularity, SDANN, SDNNind, delta, interval index, coefficient of variation, mean of RR intervals, mean of heart rate, Poincaré plot, detrended fluctuation analysis, approximate entropy, sample entropy, Shannon entropy, box counting dimension and Sevcik estimate for fractal dimension were tested and the reached results are presented. As the most promising features the following were found: slow detrended fluctuation analysis and ratio of axes in the Poincaré plot $PcPSD12$. Both features had high level of significance according to Mann-Whitney test. In this work not only theoretical analysis of the methods is provided, also practical implementation of the methods was done in form of program with user-friendly graphical interface.

Finally, the simple classification of data was undertaken for evaluation of the practical usability of the features. As a result classification using arbitrary decision tree in WEKA brought us results of 92,1 % specificity and 92 % sensitivity in distinguishing between atrial fibrillation and normal sinus rhythm - results comparable to those in recent articles.

Abstrakt

Tato diplomová práce se zabývá zhodnocením různých metod analýzy srdečního rytmu, které by mohly být použity pro rozpoznání fibrilace síní od sinusového rytmu. Otestoval jsem mnoho lineárních a nelineárnich metod, konkrétně: NN50, pNN50, RMSSD, SDNN, SDSD, krátkodobá variabilita , dlouhodobá nepravidelnost, SDANN, SDNNind, delta, interval index, koeficient variace, střední hodnota RR intervalů, střední hodnota srdečního rytmu, Poincaré plot, DFA, aproximativní entropie, vzorová (sample) entropie, Shannonova entropie, box counting dimenze a Sevcikův odhad fraktální dimenze a prezentoval dosažené výsledky. Mezi nejúspěšnější příznaky patří pomalé DFA a poměr os Poincaré plotu $PcPSD12$. Oba příznaky měly vysokou významnost založenou na Mann-Whitneyho testu. V této práci není provedena pouze teoretická analýza metod, ale také jejich praktické použití formou programu s grefickým rozhraním.

Nakonec byla vyzkoušena jednoduchá klasifikace dat pro zhodnocení praktického použití příznaků. Výsledná klasifikace na základě libovolného rozhodovacího stromu nám přinesla výsledky se specifičností 92,1 % a citlivostí 92 % pro rozeznání mezi síňovou fibilací a normální srdeční činností - výsledky srovnatelné s údaji v nedávno publikovaných článcích.

Contents

List of Figures

List of Tables

Glossary

AF atrial fibrillation.

NSR normal sinus rhythm.

HRV heart rate variability

HR heart rate

NN normal-to-normal.

PcP Poincaré plot.

SD standard deviation.

ApEn approximate entropy.

SampEn sample entropy

CV coefficient of variation

SevcikD Sevcik estimate for fractal dimension

BoxCountD box counting dimension

II interval index

STV short term variability

LTI long term irregularity

Chapter 1

Introduction

The atrial fibrillation (AF) is very extended disease. About 0,5 % of whole adult population, 4 % of persons over sixty years and more then 10 % of people over seventy years old are afflicted with this disease. The treatment of AF is based on the three methods. Rate control treatment is based on slow down the heart beats by medicaments (beta-blocker drugs). Rhythm control treatment using electrical impulse to reverting the erratic heart beats back to the normal regular rhythm also called cardiversion. Peacemakers and defibrillators are implanted to the body for that purpose. Anticoagulation is treatment with medicaments that reduce chance of forming a blood clot. Untreated atrial fibrillation may leads to the sudden death, therefore early diagnosis is very important to successful treatment.

The AF may be predicted with the help of heart rate analysis (HRV). Firstly were used linear methods but the results weren't satisfactory. Nowadays nonlinear methods are very popular. Efficiency and complexity of these methods are used for classification of AF and other diseases detectable on the HRV.

This thesis focuses generally on acquirement of knowledge from state of the art papers about HRV analysis. From these papers are chosen features that may be useful for distinguishing between atrial fibrillation (AF) and normal sinus rhythm (NSR). The assessment of these features is in the chapter Results. They were implemented by me into toolbox with graphical user interface for easier manipulation. Real data were used from two Physionet database, first database containing subjects

with atrial fibrillation (AF) and second database containing subjects with normal sinus rhythm (NSR). At the end of this work the results are compared.

- **Medical Background**: This chapter serves as an introduction to atrial fibrillation (AF), its classification, causes, symptoms and treatment.

- **HRV analysis methods:** This chapter describes all HRV features, such as $NN50$, $pNN50$, $RMSSD$, standard deviation of normal-to-normal, $SDSD$, short term variability, long term irregularity, $SDANN$, $SDNNind$, delta, interval index, coefficient of variation, mean value of RR intervals, mean value of heart rate, Poincaré plot, detrended fluctuation analysis, approximate entropy, sample entropy, Shannon entropy, box counting dimension and Sevcik estimate for fractal dimension, used for comparison their significance in recognition of AF.

- **Implementation of methods:** This chapter describes input data, preprocessing, implementation of all features and toolbox with GUI.

- **Results:** This chapter deals about reached results, what kind of segmentation is the best choice, which features are useful and which not.

- **Conclusion:** Conclusion of this thesis.

Chapter 2

Medical background

This chapter extracts information from [10].

2.1 Atrial fibrillation

If you have atrial fibrillation (AF) then:

- Your heart rate is usually a lot faster than normal.

- Your heart beat is irregular. This is called an abnormal heart rhythm or an arrhythmia.

- The force of each heart beat can vary in intensity.

What happens is that the normal controlling pacemaker in the heart is overridden by lot of random electrical impulses that originate from the heart muscle in the atria. This means that the atria only partially contract - but very rapidly (up to 400 times per minute) - they fibrillate. Only some of these impulses pass through to the ventricles. Therefore, the ventricles contract anywhere between 50 and 180 times a minute, but usually between 140 and 180 times a minute. However, the ventricles contract in an irregular way and with varying force.

Therefore, if you have AF and feel your pulse, you may count up to 180 beats per minute. Also, the force of each beat can vary, and the pulse feels erratic. Atrial fibrillation shows the figure 2.1
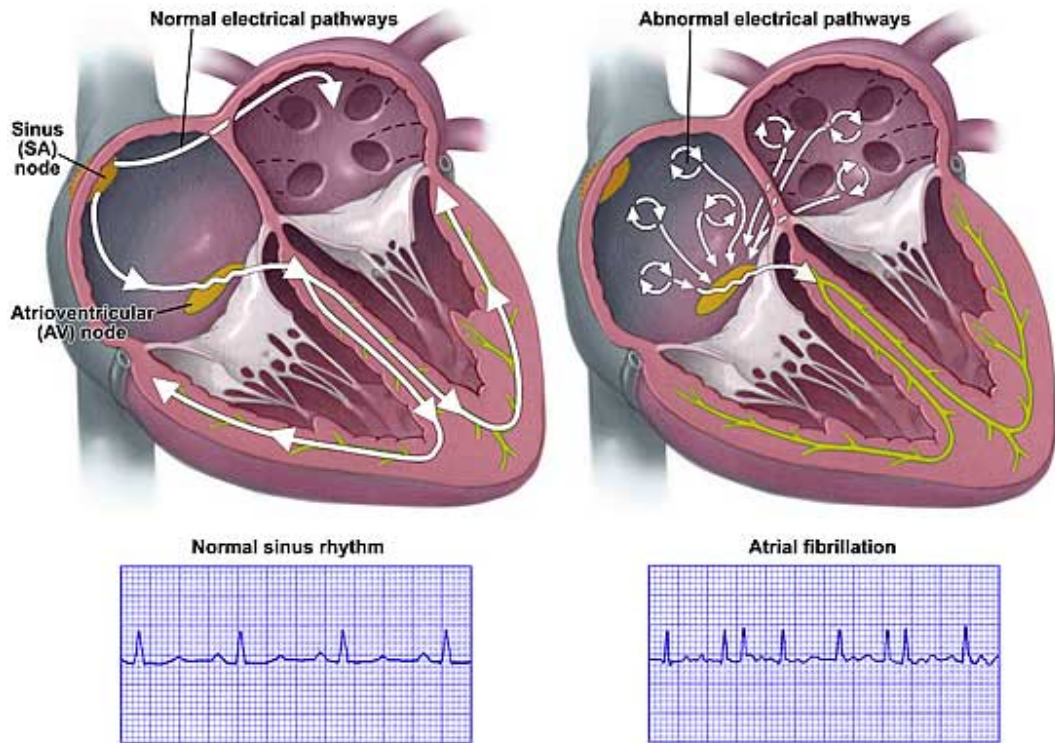
Figure 2.1: Normal sinus rhythm vs. atrial fibrillation.

2.1.1 Classification of AF

AF is commonly classified in the following 3 way.

Paroxysmal AF

The word paroxysmal means recurring sudden episodes of symptoms. If you have paroxysmal AF it means that you have episodes of AF that come and go. Each episode comes on suddenly, but will stop without any treatment within seven days (usually within two days). Each episode stops just as suddenly as it starts and the heart beat goes back to a normal rate and rhythm. The period of time between each episode (each paroxysm) can vary greatly from case to case. Although paroxysmal AF means that it will stop on its own, some people with paroxysmal AF take treatment as soon as the AF develops to stop it as quickly as possible after it starts.

Persistent AF

This means AF that lasts longer than seven days and is unlikely to revert back to normal without treatment. However, the heart beat can be reverted back to a normal rhythm with cardioversion treatment (see later). Persistent AF tends to be recurrent so it may come back again at some point after successful cardioversion treatment.

Permanent AF

This means that the AF is present long-term and the heart beat has not been reverted back to a normal rhythm. This may be because cardioversion treatment was tried and was not successful, or because cardioversion has not been tried. People with permanent AF are treated to bring their heart rate back down to normal, but the rhythm remains irregular (see below). Permanent AF is sometimes called established AF.

Most people with AF have permanent AF.

2.1.2 Causes of AF

High blood pressure is the most common cause. (High blood pressure puts a strain on the heart muscle.)

AF is a common complication of various heart conditions. For example, AF is a complication of ischaemic heart disease. This is the condition that causes angina and heart attacks and is common in older people. Various other heart problems may also trigger AF to develop. For example, AF occurs in some people with heart valve problems, cardiomyopathy, and pericardial disease.

Other conditions and situations that may trigger AF to develop include: an overactive thyroid gland (hyperthyroidism); pneumonia; pulmonary embolus; obesity; lung cancer; drinking a lot of alcohol; drinking a lot of caffeine (tea, coffee, etc).

In about 1 of 9 cases of AF there is no apparent cause. The heart is otherwise

fine and there are no other diseases to account for it. This is called lone AF.

2.1.3 Symptoms of AF

Symptoms often develop quickly, soon after the AF develops. Possible symptoms include:

- **Palpitations.** This means that you become aware of your heart. You may feel it beating in a fast and irregular way.

- **Dizziness.**

- **Angina (chest pains)** may develop. In particular, the pains tend to occur when you exert yourself, but they may occur even when you are resting.

- **Breathlessness** is often the first symptom that develops. It may occur all the time, but you may become breathless just when you exert yourself such as when you walk up stairs.

The reason why breathlessness, dizziness, and angina may develop is because when the heart beats too fast, it becomes less efficient. Small amounts of blood pumped faster by the heart are not as good as larger amounts that are pumped at the slower normal rate. This can lead to a pooling of blood in the veins of the lungs, and a reduced output of blood from the heart which can lead to these symptoms.

However, many people with AF have no symptoms, particularly if their heart rate is not very fast. The AF may then be diagnosed by chance when a doctor or nurse feels your pulse.

2.2 Treatment options for atrial fibrillation

2.2.1 Rate control treatment

If the heart rate is brought down to normal the heart becomes efficient again, and the symptoms usually improve. The pulse may still feel irregular, but not fast.

Several drugs can slow the heart rate down. They include beta-blocker drugs (such as atenolol, metoprolol and propranolol), diltiazem, verapamil, and digoxin. These drugs work by interfering with the electrical impulses of the heart. The drug chosen may depend on factors such as other heart problems that you may have.

In untreated AF, the heart rate may be as fast as 180 beats per minute, although it is more commonly between 120 and 160 beats per minute. The aim of medication is to bring the heart rate back down to normal (ideally, to less than 90 beats per minute when resting). Treatment is usually successful, but the dose needed can vary from case to case. Also, in some cases a combination of drugs may be needed if the heart rate is not brought down low enough with a single drug.

2.2.2 Rhythm control treatment

Rhythm control means reverting the erratic heart beat back to a normal regular rhythm. This is called cardioversion. This treatment is not tried in most cases as treatment to control the rate of the heart (described above) usually works well to control symptoms. However, cardioversion may be considered in certain situations. For example, if drugs to control the fast heart rate do not work very well, or if the irregular heart beat is causing unpleasant symptoms.

One method of cardioversion is to give your heart an electric shock. Another method is to use drugs that may convert the heart rhythm back to a regular beat. Both of these methods have only limited success. For example, after cardioversion, within a year in about half of cases the heart has reverted back to AF.

Cardioversion is more likely to be considered in certain situations, for example:

- If your AF developed recently.

- If you are younger than 65. (Age is no bar to cardioversion, but it is less likely to an option the older you become.)

- If an underlying cause of the AF has been successfully treated (and so AF is unlikely to come back again once the normal heart rhythm has been restored).

- If you have no other heart abnormality. (That is, if you have lone AF described earlier.)

- If you have acute heart failure or unstable angina which is being made worse by the irregular heart beat of AF.

Cardioversion is usually not an option in certain situations. For example:

- If you have certain heart diseases that include a structural problem to the heart (for example, certain valve problems such as mitral stenosis).

- If you have had AF a long time (usually more than 12 months).

- If you have had several previous attempts at cardioversion which have not worked, or only worked for a short time before the heart reverted back to AF.

A newer technique to restore the heart rhythm is called catheter ablation. In this procedure a catheter (a long thin wire) is passed into the chambers of the heart via an artery in a groin. The tip of the catheter can destroy tiny sections of heart tissue that may be the source or trigger of the abnormal electrical impulses. This treatment is only suitable in certain cases and is not a routine treatment. It does not always work and there is a small risk of serious complications.

2.2.3   Anticoagulation

Anticoagulation means that you take a drug to reduce the chance of forming a blood clot. Therefore, anticoagulation helps to prevent a stroke from occurring. Some people call anticoagulation thinning the blood although the blood is not actually made any thinner. The most commonly used anticoagulant drug is called warfarin. Warfarin interferes with certain chemicals in the blood to prevent blood clots forming so easily.

Overall, warfarin reduces the risk of stroke by nearly two-thirds. In other words, warfarin treatment can prevent about 6 of 10 strokes that would have occurred in

people with AF. The greatest benefit is seen in those people who are in the high risk category of having a stroke (described above).

As with all treatments, there is a small risk if you take warfarin. The main risk is that a bleeding complication may develop as the blood will not coagulate so well. Over a period of one year of treatment, about nine in a thousand people who take warfarin for AF are likely to have a serious bleeding problem. For example, some people may develop a serious bleeding ulcer in the gut. If you have a serious bleed you are likely to need to be admitted to the hospital, often needing a blood transfusion, and it can even result in death.

Most people with AF who have a high or medium risk of having a stroke are advised to take warfarin. However, some people with a moderate risk may be treated with aspirin rather than warfarin (see below), particularly if the risks of taking warfarin are higher than average. People with a low risk of having a stroke are not usually advised to take warfarin. This is because the benefit does not usually outweigh the risk of serious bleeding problems with taking warfarin. In short, the decision to take warfarin is a joint decision between you and your doctor. It involves weighing up the risk of having a stroke against the small risk of a complication from taking warfarin.

If you take warfarin you will need regular blood tests to check on how quickly your blood coagulates. Blood tests may be needed quite often at first, but should become less often quite quickly. The aim is to get the dose of warfarin just right so your blood does not clot as easily as normal, but not so much as to cause bleeding problems.

Aspirin is another drug that helps to prevent blood clots forming. It is not as effective as warfarin, but is less likely to cause problems. It is usually advised if you only have a low risk of stroke, or if you cannot take warfarin.

2.2.4 Other treatments

Other treatments may be advised, depending on the need to treat any underlying problems such as angina, heart valve problems, high blood pressure, overactive

thyroid, etc.

Chapter 3

HRV analysis methods

In this chapter I characterize methods used for extraction of heart rate variability (HRV) features. The clinical relevance of heart rate variability was first appreciated in 1965 when Hon and Lee [14] noted that fetal distress was preceded by alterations in interbeat intervals before any appreciable change occurred in the heart rate itself. Twenty years ago, Sayers and others [27] focused attention on the existence of physiological rhythms imbedded in the beat-to-beat heart rate signal. During the 1970s, Ewing et al. [11] devised a number of simple bedside tests of short-term RR differences to detect autonomic neuropathy in diabetic patients. The association of higher risk of post-infarction mortality with reduced HRV was first shown by Wolf et al. in 1977 [31]. In 1981, Akselrod et al. [1] introduced power spectral analysis of heart rate fluctuations to quantitatively evaluate beat-to-beat cardiovascular control.

About the time when HRV was being increasingly discussed as a tool for risk stratification after myocardial infarction [16], chaos theory was being propagated in popular science publications. Striking about most of the later works was that, although heart rate regulation was the focus of nonlinear dynamics and was seen as a particularly lucid example of chaos in physiology, there were very few scientific papers on which such claims could be based. In the following years increased effort was put into determination the chaotic nature of cardiac activity by applying analysis methods from nonlinear systems theory. For instance, on the presumption

of an underlying chaotic attractor, scaling properties of the generating dynamics were estimated on the basis of dimensional analysis. It was however soon realized that there are fundamental difficulties involved which include the noisy nature of biological signals, the restricted length of the data available and the problems with non-stationarity. This has led to a shift of the notation from chaos to complexity, irregularity, or randomness and has resulted in the development of measures and analysis techniques which are deemed more appropriate and more practical in application with respect to HRV [30].

Up to date, thousands of articles were written about linear and nonlinear methods for HRV analysis. In guidelines for HRV analysis [21] there are many linear methods described ($NN50$, $pNN50$, $SDNN$, $SDANN$, $RMSSD$, frequency domain methods, etc.). In 1991 Steave Pincus was between founders who developed nonlinear method for HRV analysis called approximate entropy (ApEn) [24]. In the following years many nonlinear methods were developed. For example adjusted ApEn called sample entropy is derived from approaches developed by Grassberger and co-workers [26]. Poincaré plot was another nonlinear method developed by Henry Poincaré for analyzing complex system. This method is used for quantifying the HRV in many articles, for example [25, 18] . Calculation of fractal dimension was explained by Hausdorff [9]. The derivation of Hausdorff dimension are Box count dimension [3] and Sevcik estimate for fractal dimension [28]. These methods will be described in detail further.

All these methods are used for prediction of many heart diseases. Examples are predictors of atrial fibrillation [15, 7], characterization of fetal heart rate [8, 13].

3.1 Heart rate and heart rate variability

Heart beats or contractions are caused by electrical depolarization of the heart muscles. Electrical depolarization of different parts of heart can be observed on an electrocardiogram (ECG). We are interested in the depolarization of lower chambers, known as ventricles, which creates the QRS complex (see Fig. 3.1 used from [25]).

The interval between two consecutive successive heart beats is called RR interval. It is the distance between two consecutive QRS complexes, usually measured as the distance between the RR waves. Mathematically we can describe RR interval like this:

$$RR = RR_{i+1} - RR_i[ms; ms, ms] \tag{3.1}$$

In healthy people the sinus node is the main rhythm generator or the primary pacemaker, with its own intrinsic activity of about 90-100 beats per minute.
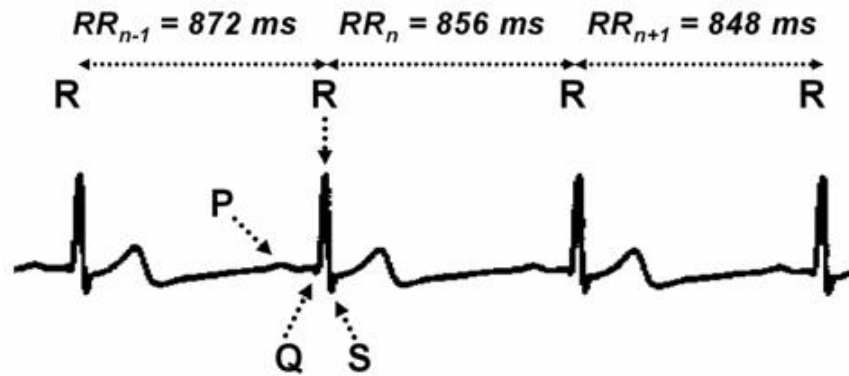
Figure 3.1: A strip of ECG presenting heart's electrical activity recorded in a healthy person. The P wave shows the depolarization of cardiac atria. The QRS complex characterizes depolarization of cardiac ventricles. R waves of the QRS complex are always directed upward and therefore are usually used for the identification of each heart beat and its duration called the RR interval (from one R to the consecutive R wave). In this sample there are 4 heart beats and 3 RR intervals between them.

The momentary heart rate and the duration of the RR interval is a consequence of constant interaction between the intrinsic activity of the sinus node and the influence of the autonomic nervous system, various substances circulating in the blood and present in the heart tissues. Breathing appears to be most important factor

modulating heart rate. This rhythmic phenomenon is known as respiratory sinus arrhythmia (RSA). It causes heart rate acceleration during inspiratory and its deceleration during expiratory. The changes in blood pressure modulated by baroreflex are another example of a separate system regulating the heart rate. The control of heart rate is modulated by both sympathetic and parasympathetic branches of autonomic nervous system as well as many other autonomic reflexes (e.g.: chemoreflexes). All these systems and reflexes are responsible for changing of the duration of RR interval from one beat to another and this phenomenon is called heart rate variability (HRV).

$$HR = 60 \cdot 1000 \cdot \frac{1}{RR_i}[1/min; -, ms] \tag{3.2}$$

The RR intervals characteristic for incorrect beats are usually shorter (eg.: $RRi < 300ms$) or longer ($RRi > 2000ms$) than physiologically acceptable for healthy people. Another characteristics of incorrect beats may be a change of more than 20 % from RR_i to RR_{i+1}.

In the following text I describe 22 methods for HRV analysis, 15 linear and 7 nonlinear methods. From these methods I got 26 features and I compare them.

3.2 Linear methods for HRV analysis

Linear methods for HRV analysis are created by features in time and frequency domain. They describe the characteristics of the signals but they don't say anything about manners of the signals. The main problem is the non-stationarity and the noisy nature.

3.2.1 NN50

This method is based on computing count of adjacent RR intervals differing by more than 50 ms in the entire analysis interval. It's used for classification of the segment longer or at least 5 minutes. This feature is introduced in article [7] and we can describe it as:

$$NN50 = \sum_{i=1}^{N} \{|RR_{i+1} - RR_i| > 50ms\}[count; -] \tag{3.3}$$

where N is total number of all RR intervals in segment

3.2.2 pNN50

This analysis rests in the NN50 count divided by total number of all RR intervals in selected segment [7]. It means $NN50$ in heart rate (HR) (%).

$$pNN50 = \frac{NN50}{N} \cdot 100[\%; -] \tag{3.4}$$

where N is total number of all RR intervals in segment

3.2.3 Standard deviation NN (SDNN)

SDNN means standard deviation of all RR intervals in selected segment. In statistics, standard deviation is a simple measure of the variability or dispersion of a data set. A low standard deviation indicates that the data points tend to be very close to the same value (the mean) - lower is HRV, while high standard deviation indicates that the data are "spread out" over a large range of values - higher is HRV. This function is also used in article [7].

The standard deviation of a discrete random variable is the root-mean-square (RMS) deviation of its values from the mean.

This calculation is described by the following formula:

$$SDNN = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (RR_i - \overline{RR})}[ms; -, ms, ms] \tag{3.5}$$

where $\overline{RR}$ is the arithmetic mean of the values RR_i, defined as:

$$\overline{RR} = \frac{RR_1 + RR_2...R_N}{N} = \frac{1}{N} \sum_{i=1}^{N} RR_i[ms; -, ms] \tag{3.6}$$

where N is total number of all RR intervals in segment

3.2.4 RMSSD

RMSSD is the square root mean of successive RR intervals. It can be found in article [7] and it is described by following equation:

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (RR_{i+1} - RR_i)^2} [ms; -, ms, ms] \qquad (3.7)$$

where N is total number of all RR intervals in segment

3.2.5 SDNNind

The SDNN index is the mean of the standard deviation (see Eq. 3.5) of all RR intervals for all 1 minute sections in selected segments of long term signal. Standard deviation was explained in text before. This feature says us variation of signal in short intervals.

$$SDNNind = \frac{1}{N} \sum_{i=1}^{N} SDNN_i [ms; -, ms] \qquad (3.8)$$

where N is total number of all 1 minute sections RR intervals in selected segment

3.2.6 SDANN

SDANN is the standard deviation (see Eq. 3.5) of the averages of RR intervals in all 1 minute section which they divide selected segments of long term signal.

$$SDANN = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\overline{RR_i} - \overline{\overline{RR}})^2} [ms; -, ms, ms] \qquad (3.9)$$

where N is total number of 1 minute sections RR intervals in selected segment,

$\overline{RR_i}$ is mean of RR intervals in 1 minute section,

$\overline{\overline{RR}}$ is mean of all means of RR intervals in all 1 minute sections.

3.2.7 SDSD

SDSD is counted like standard deviation of the temporal differences of consecutive RR intervals. We can formalize it with following formula:

$$SDSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} \left(|RR_i - RR_{i+1}| - \overline{RRdif}\right)^2} [ms; -, ms, ms, ms] \quad (3.10)$$

where N is total number of all RR intervals and
$\overline{RRdif}$ is described by following equation:

$$\overline{RRdif} = \frac{|RR_1 - RR_2| + ... + |RR_{N-1} - RR_N|}{N-1}$$

$$= \frac{1}{N-1} \sum_{i=1}^{N-1} (|RR_i - RR_{i+1}|) [ms; -, ms, ms] \quad (3.11)$$

3.2.8 RR mean

This feature is the standard statistical indicator. Mean is a parameter of distribution random variable, which is defined as a weighted average this distribution. The mean of all RR intervals is denumerable by following equation:

$$\overline{RR} = \frac{RR_1 + RR_2...R_N}{N} = \frac{1}{N} \sum_{i=1}^{N} RR_i [ms; -, ms] \quad (3.12)$$

where N is total number of all RR intervals.

3.2.9 HR mean

It is mean of heart rate. Enumeration of heart rate was mentioned above with equation 3.2. Mean of heart rate is similar like RR mean described in subsection 3.2.8:

$$\overline{HR} = \frac{1}{N} \sum_{i=1}^{N} HR_i [ms; -, ms] \quad (3.13)$$

where N is total number of all RR intervals.

3.2.10 Coefficient of Variation

CV is a normalized measure of dispersion of a probability distribution. It is defined as the ratio of the standard deviation $SDNN$ (see Eq. 3.5) to the mean $\overline{RR}$ (see Eq. 3.12).

$$CV = \frac{SDNN}{\overline{RR}}[-; ms, ms] \tag{3.14}$$

3.2.11 Short-term Variability

The Short-term Variability is described in article [12]. It's used for new born babies but I tried it also on adults. In first step we must resample the heart rate signal with interpolation to four Hertz sampled signal. Then we can use the following formula to compute STV:

$$STV = \frac{\sum_{i=1}^{24M} |sRR_{i+1} - sRR_i|}{24 \cdot M}[ms; ms, ms, -] \tag{3.15}$$

where sRR_i is the value of the signal RR_i taken every 2.5 s (i.e. once every ten samples).

$$sRR_i = RR(10 \cdot (i - 1) + 1)[ms; ms] \tag{3.16}$$

and M is count of all minutes in selected segment. The number 24 means count of sRR_i in one minute.

3.2.12 Long Time Irregularity

Long term irregularity (LTI) [12], defined as interquartile range [1/4, 3/4] (between 25 % and 75 %) of the distribution values m with:

$$m = \sqrt{RR_i^2 + RR_{i+1}^2}[ms; ms, ms] \tag{3.17}$$

This feature represents long term variability.

3.2.13 Interval Index

Interval index is the value denumerable like quotient of STV to standard deviation of all sRR in selected segment.

$$II = \frac{STV}{std(sRR_i)}[-; ms, ms] \tag{3.18}$$

where $std(sRR_i)$ is standard deviation described in subsection 3.2.3.

3.2.14 Delta

The Delta index shows the mean of differences between maximal and minimal values in one-minute long segments [12]. It is described by following equation:

$$Delta = \frac{\sum_{k=1}^{m}\left(\max_{i\epsilon[(k-1)\cdot240+1,k\cdot240]}(RR_i) - \min_{i\epsilon[(k-1)\cdot240+1,k\cdot240]}(RR:i)\right)}{m}[ms; ms, ms, -] \tag{3.19}$$

where m is the time of duration of the recording in minutes, the number 240 means 240 samples, it is one-minute of the signal by four Hertz sampled.

3.2.15 Delta total

The Delta total is difference between maximal and minimal value in selected segment [12].

$$Delta\_total = (\max_{i\epsilon[1,N]}(RR_i) - \min_{i\epsilon[1,N]}(RR_i)[ms; ms, ms] \tag{3.20}$$

where N is total number of all RR intervals in selected segment

3.3 Nonlinear methods for HRV analysis

3.3.1 Poincaré plot

The Poincaré plot, a nonlinear method is created by plotting by all RR intervals in two dimensional system. Method is described by following formula: Two adjacent

RR intervals represent one point in the plot. The first RR interval (RR_i) represents the x-coordinate, the second interval (RR_{i+1}) represents y-coordinate [15, 7, 18]. The Fig. 3.2 shows a Poincaré plot of a healthy patient.
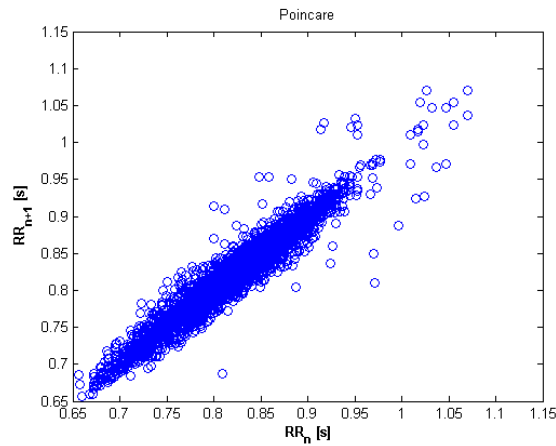
Figure 3.2: Poincaré plot of a healthy patient.

The Poincaré plot is characterized by a number of descriptors, some of which are presented in Fig 3.3 [25].
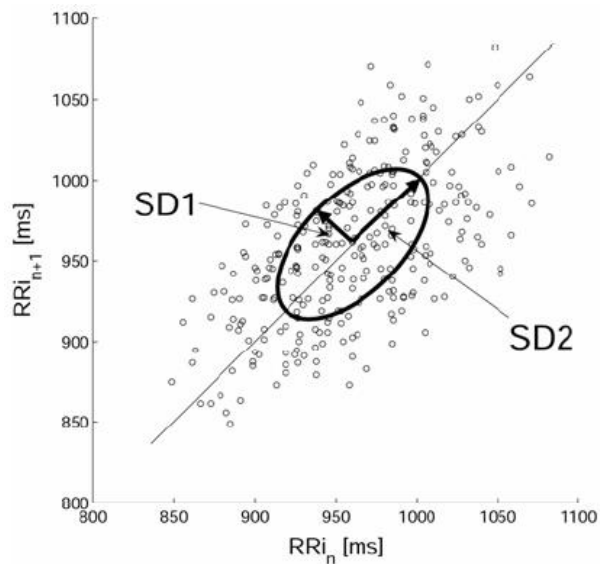
Figure 3.3: Some Poincaré plot descriptors and the PP ellipse.

$SD1$ and $SD2$ are two standard Poincaré plot descriptors [25, 7]. $SD2$ is defined

as the standard deviation of the projection of the Poincaré plot on the line of identity $(y = x)$, and $SD1$ is the standard deviation of projection of the PcP on the line perpendicular to the line of identity $(y = -x)$ [25]. Both parameters we may define as:

$$SD1 = \sqrt{Var(x_1)}[ms; ms], \ SD2 = \sqrt{Var(x_2)} \ [ms; ms] \qquad (3.21)$$

where $Var(x)$ is the variance of x, and

$$x_1 = \frac{\overrightarrow{RR_i} - \overrightarrow{RR_{i+1}}}{\sqrt{2}}, \ x_2 = \frac{\overrightarrow{RR_i} + \overrightarrow{RR_{i+1}}}{\sqrt{2}} \ [ms; ms, ms] \qquad (3.22)$$

$\overrightarrow{RR_i}$ and $\overrightarrow{RR_{i+1}}$ are vectors defined as:

$$\overrightarrow{RR_i} = (RR_1, R_2, ..., RR_{N-1}), \ \overrightarrow{RR_{i+1}} = (RR_2, RR_3, ..., RR_N) \qquad (3.23)$$

In the other words it means, that x_1 and x_2 correspond to the rotation of $\overrightarrow{RR_i}$ and $\overrightarrow{RR_{i+1}}$ by angle $\frac{\pi}{4}$.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} \cdot \begin{bmatrix} RR_i \\ RR_{i+1} \end{bmatrix} \qquad (3.24)$$

Another parameter which we may obtain from Poincaré plot is S. This parameter reflects the total variability:

$$S = \pi \cdot SD1 \cdot SD2 \ [ms^2; -, ms, ms] \qquad (3.25)$$

which is the area of the ellipse.

This parameter seems to be a better measure of total variability than $SDNN$, as in the case of bigeminy, when our data vector would be similar to x = (a, b, a, b, a, b ...) $SDNN$ would measure a substantial variability, whereas S would be equal to 0, which agrees with the intuitive understanding of "variability" [25].

In many articles, for example [18, 7], we can find another one parameter obtained from Poincaré plot. It is the $SD1/SD2$ ratio.

$$SD12 = \frac{SD1}{SD2}[-; ms, ms] \tag{3.26}$$

3.3.2   Detrended Fluctuation Analysis (DFA)

This method was first proposed in [5]. A detailed description of the algorithm and its application to physiologic signals can be found in [4]. The method of detrended fluctuation analysis [17] has proven useful in revealing the extent of long-range correlations in time series. Briefly, the time series to be analyzed (with N samples) is first integrated. Next, the integrated time series is divided into boxes of equal length n. In each box of length n, a least squares line is fit to the data (representing the trend in that box). The y coordinate of the straight line segments is denoted by $y_n(k)$. Next, we detrend the integrated time series, $y(k)$, by subtracting the local trend, $y_n(k)$, in each box. The root-mean-square fluctuation of this integrated and detrended time series is calculated by:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n(k)]^2} \tag{3.27}$$

This computation is repeated over all time scales (box sizes) to characterize the relationship between $F(n)$, the average fluctuation, and the box size n. Typically, $F(n)$ will increase with box size.

I used box size 4 beats for fast time scale and 64 beats for slow time scale.

3.3.3   Shannon entropy

In information theory, entropy is a measure of the uncertainty associated with a random variable. The term by itself in this context usually refers to the Shannon entropy, which quantifies, in the sense of an expected value, the information contained in a message, usually in units such as bits. Equivalently, the Shannon entropy is a measure of the average information content one is missing when one does not know the value of the random variable. The concept was introduced by

Claude E. Shannon in his 1948 paper "A Mathematical Theory of Communication" [29].

Shannon entropy represents an absolute limit of the best possible lossless compression of any communication, under certain constraints: treating messages to be encoded as a sequence of independent and identically-distributed random variables, Shannon's source coding theorem shows that, in the limit, the average length of the shortest possible representation to encode the messages in a given alphabet is their entropy divided by the logarithm of the number of symbols in the target alphabet.

Definition

The entropy H of a discrete random variable X with possible values $x_1, ..., x_n$ is:

$$H(X) = E(I(X)) \tag{3.28}$$

Where E is the expected value function, and $I(X)$ is the information content or self-information of X. $I(X)$ is itself a random variable. If p denotes the probability mass function of X then the entropy can explicitly be written as:

$$H(X) = \sum_{i=1}^{n} p(x_i) \cdot I(x_i) = -\sum_{i=1}^{n} p(x_i) \cdot \log_b p(x_i) \tag{3.29}$$

where b is the base of the logarithm used. Common values of b are 2, Euler's number e, and 10, and the unit of entropy is bit for $b = 2$, nat for $b = e$, and dit (or digit) for $b = 10$.

In the case of $p_i = 0$ for some i, the value of the corresponding summand $0 \log_b 0$ is taken to be 0, which is consistent with the limit:

$$\lim_{p \to 0^+} p \log p = 0 \tag{3.30}$$

Characterization

Shannon entropy is characterized by a small number of criteria, listed below. It can be shown that any definition of entropy satisfying these assumptions has the

form:

$$- K \cdot \sum_{i=1}^{n} p(x_i) \cdot \log_b p(x_i) \qquad (3.31)$$

where K is a constant corresponding to a choice of measurement units.

In the following, $p_i = P_r(X = x_i)$ and $H_n(p_1, ..., p_n) = H(X)$.

Continuity

The measure should be continuous, so that changing the values of the probabilities by a very small amount should only change the entropy by a small amount.

Symmetry

The measure should be unchanged if the outcomes xi are re-ordered.

$$H_n(p_1, p_2, ...) = H_n(p_2, p_1, ...) \; etc. \qquad (3.32)$$

Maximum

The measure should be maximal if all the outcomes are equally likely (uncertainty is highest when all possible events are equiprobable).

$$H_n(p_1, ..., p_n) \leq H_n(\frac{1}{n}, ..., \frac{1}{n}) \qquad (3.33)$$

For equiprobable events the entropy should increase with the number of outcomes.

$$H_n(\underbrace{\frac{1}{n}, ..., \frac{1}{n}}_{n}) < H_{n+1}(\underbrace{\frac{1}{n+1}, ..., \frac{1}{n+1}}_{n+1}) \qquad (3.34)$$

Additivity

The amount of entropy should be independent of how the process is regarded as being divided into parts.

This last functional relationship characterizes the entropy of a system with sub-systems. It demands that the entropy of a system can be calculated from the entropies of its sub-systems if the interactions between the sub-systems are known.

Given an ensemble of n uniformly distributed elements that are divided into k boxes (sub-systems) with $b_1, b_2, ..., b_k$ elements each, the entropy of the whole ensemble should be equal to the sum of the entropy of the system of boxes and the individual entropies of the boxes, each weighted with the probability of being in that particular box. For positive integers b_i where $b_1 + ... + b_k = n$:

$$H_n(\frac{1}{n}, ..., \frac{1}{n}) = H_k(\frac{b_1}{n}, ..., \frac{b_k}{n}) + \sum_{i=1}^{k} \frac{b_i}{n} H_{bi}(\frac{1}{b_i}, ..., \frac{1}{b_i}) \tag{3.35}$$

Choosing $k = n$, $b_1 = ... = b_n = 1$ this implies that the entropy of a certain outcome is zero:

$$H_1(1) = 0 \tag{3.36}$$

3.3.4 Approximate entropy (ApEn)

An interesting measure to assess the degree of randomness of sequences of numbers was proposed by Steave Pincus [24] and described in many works (i.e. [8, 13]) . ApEn is related to the probability that segments of m data samples which are similar (i.e., closer each other then given distance r) remain similar when the segment length increase to $m + 1$ [6]. Input parameter m is the value of size of vectors for comparison in selected segment of RR intervals. I set $m = 2$. It creates $N - m + 1$ vectors of m components from input data:

$$R_m(i) = [RR(i), RR(i + 1), ..., RR(i + m)]. \tag{3.37}$$

The vector $R_m(i)$ represents the sequence of m consecutive RR_i values starting at the beat i. Two vectors $R_m(i)$ and $R_m(j)$ are similar if the absolute differences between each couple of corresponding scalar components are less then distance $r \cdot SDNN$.

The second input parameter r, mentioned above, designates the threshold for comparison differences between R_m vectors. This parameter is usually chosen from values between 0.1 and 0.25, the recommended range in many articles. ApEn is calculated for input parameters $m = 2$ and $r = 0.2$ in almost all HRV studies. I used these values also. The article [6], conversant about the right choice of threshold r, deals with the relation between ApEn and values of r. It is shown in Fig. 3.4 used from [6].
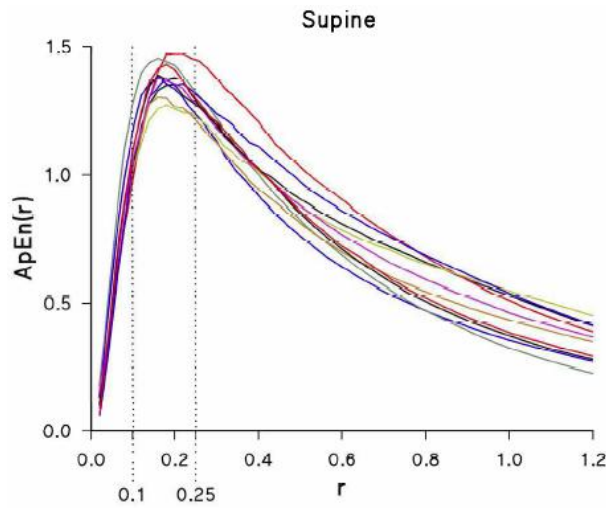
Figure 3.4: Individual ApEn(r) functions in 10 healthy volunteers during supine rest; the two dotted vertical lines delimitate the recommended r range.

If we call $n_i^m(r)$ the number of $N - m + 1$ vectors $R_m(j)$ which are similar to $R_m(i)$, then:

$$C_i^m(r) = \frac{n_i^m(r)}{N - m + 1} \; [-;-,-] \tag{3.38}$$

is the probability to find a sequence of m beats similar to the sequence represented by $R_m(i)$. $C^m(r)$, defined as the mean of all $C_i^m(r)$, quantifies the prevalence of similar strings of m beats. ApEn(r) is calculated as:

$$ApEn(r, m) = \ln[\frac{C^m(r)}{C^{m+1}(r)}] \; [-;-,-] \tag{3.39}$$

A high degree of regularity means that sequences which are similar for m points are likely to be similar also for the next $m + 1$ point, while this is unlikely to occur for irregular time series. Thus low values of ApEn reflect high regularity.

3.3.5 Sample entropy (SampEn)

Sample entropy is function very similar to ApEn. The major difference between ApEn and SampEn that SampEn does not count self-matches, having reduced bias when compared to ApEn.

3.3.6 Fractal dimension

Studying living systems as nonlinear dynamic systems (chaotic systems as they are commonly called) is of increasing interest to medicine and biology. The fractal dimension is one possible parameter that characterizes chaotic systems, and the analysis of time series is one of the most common means to find the fractal dimension from observables [28].

In fractal geometry, the fractal dimension, D, is a statistical quantity that gives an indication of how completely a fractal appears to fill space, as one zooms down to finer and finer scales. There are many specific definitions of fractal dimension and none of them should be treated as the universal one. From the theoretical point of view the most important are the Hausdorff dimension, the packing dimension and, more generally, the Rényi dimensions. On the other hand the box-counting dimension and correlation dimension are widely used in practice, partly due to their ease of implementation.

3.3.6.1 Box counting dimension

Box counting dimension is method very similar to Hausdorff dimension [9]. Instead open balls of a radius ε needed to cover the set, there are used boxes of edge length ε needed to cover the set. To calculate this dimension for a fractal S, imagine this fractal lying on an evenly-spaced grid, and count how many boxes are

required to cover the set. The box-counting dimension is calculated by seeing how this number changes as we make the grid finer.

Suppose that $N(\varepsilon)$ is the number of boxes of side length ε required to cover the set. Then the box-counting dimension is defined as:

$$D(S) = \lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln(\varepsilon)} \ [-; -, -] \tag{3.40}$$

Box count dimension can be explained on the describe of Matlab code 1. First input data are prepared. Data are normalized in the x coordinate (see Eq. 3.43) and y coordinate (see Eq. 3.44). Then we compute the count of iterations. In every iteration is reduced the size of box edge. It starts with two boxes with size of box edge equal to the length of data divided the count of boxes. In each iteration number of boxes is increased quadratically and the size of box is reduced proportionally with number of boxes. Also number of boxes that cover the data is counted in each iteration.

Finally the box count dimension is computed as the direction of straight line that approximates curve created from the logarithms of box sizes and box count.

```
for numBox = 1:p-1
    NumberOfBoxes = 2^numBox;
    sizeBox = floor(length(c)/2^numBox);
    addBox = 0;
    for i = 1:NumberOfBoxes
        boxStart = (i-1)*sizeBox + 1;
        boxEnd = i*sizeBox;
        sizeBoxNormalized = xStar(boxEnd) - xStar(boxStart);
        % peak to peak of signal in the current box
        PPinBox = max(c(boxStart:boxEnd)) - min(c(boxStart:boxEnd));
        % none integer part of box leave for another box
        numberOfBoxesTemp(i) = fix(PPinBox/sizeBoxNormalized);
        addBox = addBox + mod(PPinBox,sizeBoxNormalized); %correction of halfBoxes
    end
    boxCount(numBox) = sum(numberOfBoxesTemp) + floor(addBox);
    numberOfBoxesTemp = [];
end
```

Source code 1: Calculation of Box counting dimension
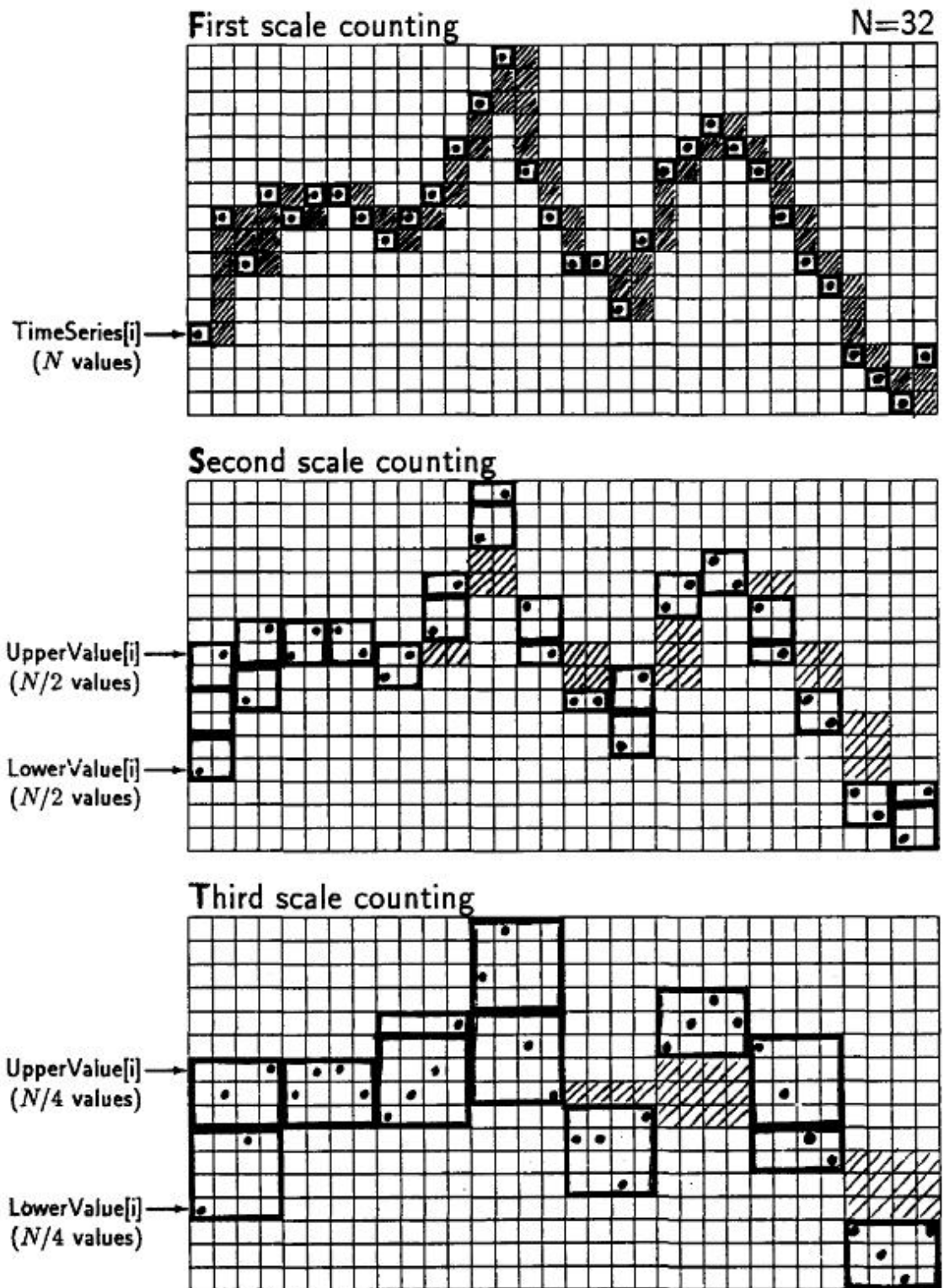
Figure 3.5: Adaptation of box counting: First three steps.

Demonstration of algorithm is shown on Fig. 3.5 used from [3].

3.3.6.2 Sevcik estimate for fractal dimension

This method go out from the Hausdorff dimension [9] but it computes only the approximate fractal dimension. The Hausdorff dimension of a set in a matric space may be expressed as:

$$D_h = \lim_{\varepsilon \to 0} \frac{\ln[N(\varepsilon)]}{\ln(\varepsilon)} \ [-;-,-] \tag{3.41}$$

where $N(\varepsilon)$ is the number of open balls of a radius ε needed to cover the set. In a metric space, given any point P, an open ball of center P and radius ε is a set of all points x for which $dist(P, x) < e$. A line of length L may be divided into $N(\varepsilon) = L/(2 \times \varepsilon)$ segments of length $2 \times \varepsilon$, and may be covered by $N(\varepsilon)$ open balls of radius ε. Thus, equation 3.41 may be rewritten as:

$$D_h = \lim_{\varepsilon \to 0}[\frac{\ln(L) + \ln(2 \cdot \varepsilon)}{\ln(\varepsilon)}] = \lim_{\varepsilon \to 0}[1 - \frac{\ln(L) - \ln(2)}{\ln(\varepsilon)}] = \lim_{\varepsilon \to 0}[1 - \frac{\ln(L)}{\ln(\varepsilon)}] \tag{3.42}$$

Waveforms are planar curves in a space with coordinates usually having different units. Since the topology of a metric space does not change under linear transformation, it is convenient linearly to transform a waveform into another in a normalized space, where all axes are equal. I propose to use two linear transformations that map the original waveform into another embedded in an equivalent metric space. The first transformation, normalizes every point in the abscissa as:

$$x_i^* = \frac{x_i}{x_{max}} \tag{3.43}$$

Where x_i are the original values of the abscissa, and $x_m ax$ is the maximum x_i. The second transformation normalizes the ordinate as follows:

$$y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \tag{3.44}$$

where y_i are the original values of the ordinate, and $y_m in$ and $y_m ax$ are the minimum and maximum y_i, respectively. These two linear transformations map the

N points of the waveform into another that belongs to a unit square. This unit square may be visualized as covered by a grid of $N \times N$ cells. N of them containing one point of the transformed waveform. Calculating L of the transformed waveform and taking $\varepsilon = 1/(2 \times N')$ equation 3.45 becomes:

$$D_h \approx D = \lim_{N' \to \infty} 1 + \frac{\ln(L) - \ln(2)}{\ln(2 \cdot N')}] = 1 + \frac{\ln(L)}{\ln(2 \cdot N')} \qquad (3.45)$$

where L is the length of the curve in the unit square and $N' = N - 1$ [28].

Chapter 4

Implementation of methods

In this chapter I describe the algorithms for preprocessing, implementation of methods for HRV analysis which I theoretically explained in the chapter 3 and the graphical user interface (GUI) for easier usage of the toolbox. I also made batch computation for easier computation of a large number of data. The section 4.1 describes data, the section 4.2 describes data preprocessing necessary before starting HRV analysis, the section 4.3 describes implementation of all HRV analysis methods, the section 4.4 describes GUI and its functions and the section 4.5 describes batch computation. All source codes and applications were made in Matlab version 2007b [19].

4.1   Data

As input I used time series of RR intervals extracted from the data from two Physionet ECG database [23]. The MIT-BIH Atrial Fibrillation Database (MIT-BIH AF DB) includes 25 long-term ECG recordings of human subjects with atrial fibrillation (mostly paroxysmal). The individual recordings are each 10 hours in duration, and contain two ECG signals each sampled at 250 samples per second with 12-bit resolution over a range of ±10 millivolts. The original analog recordings were made at Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center) using ambulatory ECG recorders with a typical recording bandwidth of

approximately 0.1 Hz to 40 Hz. The rhythm annotation files (with the suffix *.atr*) were prepared manually; these contain rhythm annotations of types (AFIB (atrial fibrillation), (AFL (atrial flutter), (J (AV junctional rhythm), and (N (used to indicate all other rhythms). Beat annotation files (with the suffix *.qrs*) were prepared using an automated detector and have not been corrected manually. The MIT-BIH Normal Sinus Rhythm Database (MIT-BIH NSR DB) includes 18 long-term ECG recordings of subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital (now the Beth Israel Deaconess Medical Center). Subjects included in this database were found to have had no significant arrhythmias; they include 5 men, aged 26 to 45, and 13 women, aged 20 to 50. Each record contain two ECG signals sampled at 128 samples per second and is long about 24 hours.

4.2 Data preprocessing

For my purpose I selected ten subjects from each database as the adequate sample for extraction features from two case, AF and NSR. Names of files are: 04015.*dat*, 04043.*dat*, 04048.*dat*, 04126.*dat*, 04746.*dat*, 04908.*dat*, 04936.*dat*, 05091.*dat*, 05121.*dat* and 05261.*dat* from MIT-AF database and from MIT-NSR database names of files are: 16265.*dat*, 16272.*dat*, 16273.*dat*, 16420.*dat*, 16483.*dat*, 16539.*dat*, 16773.*dat*, 16786.*dat*, 16795.*dat* and 17052.*dat*. For computation of HRV analysis features it was necessary to compute RR intervals.

Firstly, I detected R peaks by using Pan Tompkins QRS detector [22] because annotation contains only position of QRS complexes but not R peaks. QRS complexes in annotation are marked on the basis second ECG signals in data files where are significant S peaks. The detector of R peak was processed by me in data blocks of length 120 minutes. The R peak detector is implemented in function *pan_tompkins_qrs.m* by Xaviert Aubert [20]. This algorithm resample input data to 200 Hz for preprocessing, where noise is removed. Then data are normalized and R peak detection is based on the adaptive threshold. Returned train of R peaks is corrected to the right position of input signal. Obtained R peaks were almost mutually correlated with

annotation of data. Exceptions were created by artefacts in the first ECG signal of selected subjects where R peaks couldn't be detected (see Fig. 4.1).
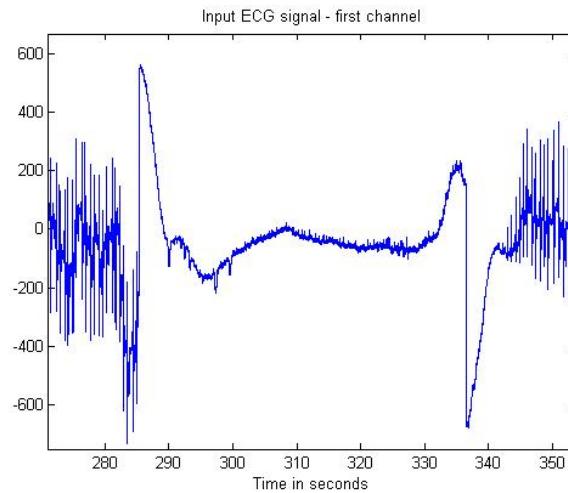
Figure 4.1: Example of artefact from input data.

In next step I computed RR intervals as differences between two adjacent R peaks which are explained in section 3.1 by mathematical definition 3.1. This way gained RR intervals could contain some outliers. These outliers were created by artefact in input ECG signals. It means that RR interval was too long, because on passage with artefact couldn't be detected R peaks. It is shown on Fig. 4.2.
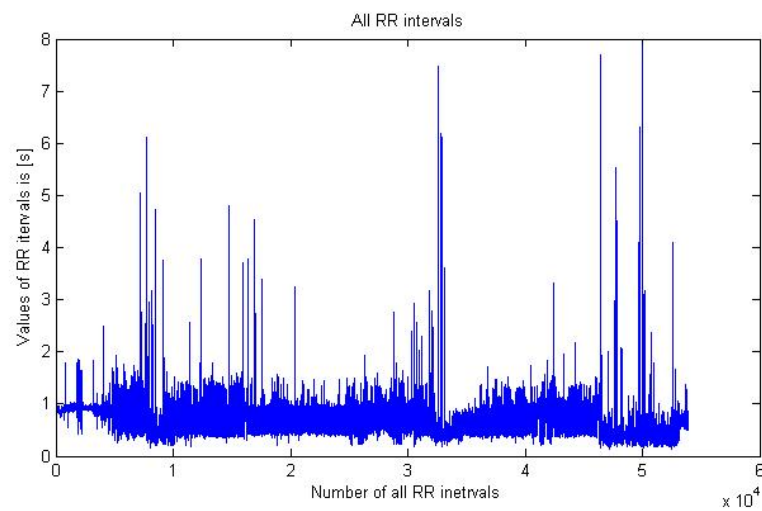
Figure 4.2: RR intervals without removing of outliers.
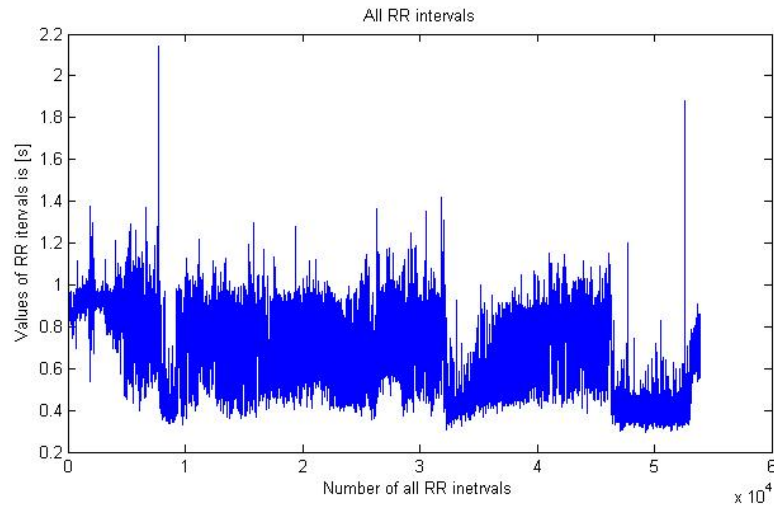
Figure 4.3: RR intervals with removing of outliers.

I created function called *removeOutlier.m* for removing of these outliers. This function was initialized with mean of first one hundred of samples. In each iteration I took five RR intervals. I calculated mean from first two and last two samples called *meanNN*. If the middle one was out of bounds (from $0,75 \cdot meanNN$ to $1,25 \cdot meanNN$) I replaced it by mean of previous successful pass (see source code 2). It means that the value of middle sample lay inside of bounds. The result of this procedure is shown on Fig. 4.3.

```
meanLast = mean(data(1:100));
for i = N_samp+1:N_RR-N_samp-1
    meanNN(i) = sum([data(i-N_samp:i-1) data(i+1:i+N_samp)])/(N_INT-1);
    if ~((1-Perc)*meanNN(i)<data(i) && data(i)<(1+Perc)*meanNN(i))
        rr_interval_all_s_outlier(i) = 1;
        rr_interval_all_s_outlier_Remove(i) = meanLast;
    else
        meanLast = meanNN(i);
    end;
end;
```

Source code 2: Removing of outliers and creation of signal smoother

Final preprocess procedure was resampling RR intervals obtained by previous described operations. I used function called *hrtach.m* which using Berger tachome-

ter [2] for constructing evenly sampled RR intervals. I sampled the signal by four Hertz. The algorithm is very simply. In order to produce the evenly sampled signal, it is to anti-alias filter the hypothetical analog signal. It is done by integrating the hypothetical analog signal over a boxcar.

After that I stored computed data in appropriate named Matlab file (*RRint_xxx.mat*, the *xxx* means the original name of input file) for faster computation by next using.

4.3 HRV analysis

In order to make the work with HRV analysis easier, I designed two instances GUI (see section 4.4), and batch computation (see section 4.5). Linear methods are generally designed for stationary signals. Biological signals do not fulfil this condition. For this purpose I decided to divide the input signal into segments. I set to five, ten, twenty and thirty minute length of segments. This is the way how to get length of data of the same size, therefore it is necessary to calculate right results. Comparison of results between selected segmentation is summarized in the chapter 5.

4.3.1 Linear methods

All linear methods explained in section 3.2 was implemented to Matlab by their mathematical definition in relevant subsections. I used combination of standard Matlab function for computation these methods, example is *std* - standard deviation, *mean* - mean of row of data, *min* - find minimum in a row of data and *max* - find maximum in a row of data. I chose only few functions to demonstrate implementation.

STV and II

The *STV* mentioned in subsection 3.2.11 defined by 3.15, and *II* mentioned in subsection 3.2.13 defined by 3.18 were implemented in the function *stv_and_ii.m*, see source code 3.

```
number_of_sHR = floor(length(data)/10);
STV_temp = zeros(1,number_of_sHR);
sHR = zeros(1,number_of_sHR);
for i = 1:number_of_sHR-1
    STV_temp(i) = abs(data(10*i+1)-data(10*(i-1)+1));
    sHR(i) = data(10*(i-1)+1);
end
STV = mean(STV_temp);
II = STV/std(sHR);
```

Source code 3: Calculation of STV and II

4.3.2 Nonlinear methods

When processing nonlinear methods, I concentrated to basic entropies and fractal dimensions. For their calculation I had to implement some methods and others I used from Matlab.

Approximate entropy

ApEn is computed with the help of function *apen.m*. This method is described in subsection 3.3.4, see Eq. 3.38 and 3.39. Inputs of this method are two parameters: r - the threshold, m - the size of vectors for comparison. I chose the parameter $r = 0.2$ and parameter $m = 2$. The reasons for choosing these values are also described in section 3.3.4.

Sample entropy

SampEn is method very similar to ApEn. The differences between these methods are explained in subsection 3.3.5. To calculation SampEn I used function *sampen.m* with the same initialization as ApEn. This function was downloaded from Physionet [23].

Shannon entropy

I calculated this feature with help of the function *entropy*. The description of this function is in subsection 3.3.3.

Box counting dimension

Box counting dimension is fractal dimension which is described in subsection 3.3.6.1. I programmed this algorithm according to article [3] and I called this function *boxcount.m* (see source code 1).

Sevcik estimate of fractal dimension

Here, I briefly describe the algorithm for the Sevcik estimate of fractal dimension. Description of this method is in subsection 3.3.6.2. Sevcik is done by using function *sevcik.m* which I wrote with help of article [28]. The input data are normalized. The array x_i is normalized by dividing of its maximum. Array of x_i is created by values from 1 to length of data. y_i contain all values of input data which it is normalized too. Maximum of y_i is discounted from y_i and then resulting values are divided by difference between maximum of y_i and minimum of y_i (see source code 4). The final estimation of fractal dimension (see source code 5).

```
xStar = x/max(x);
yStar = (y - min(y))./(max(y)-min(y));
```

Source code 4: Normalization of input for computing of Sevcik estimate
of fractal dimension

```
for i = 2:N
    L(i) = sqrt((xStar(i) - xStar(i-1))^2 + (yStar(i)-yStar(i-1))^2);
end
```

Source code 5: Sevcik estimate of fractal dimension

4.4 Graphical User Interface (GUI)

GUI was made with help of the toolbox designed by Daniel Novák [20] from the department of cybernetics, CTU in Prague. GUI main file is called *HRV_gui.m*. After startup of this file the main window shows on the screen (Fig. 4.4). The biggest area of window is filled in three plot windows. The first one shows all RR intervals, the second one shows RR intervals of selected segment and the third one shows HR of selected segment.

On the right side from these plots there are three buttons, two popup menus, two input fields and checkbox. The first button *LoadData* is given for read the data, the second button *HRV Analysis* is given for computing HRV analysis. The first popup menu shows length of segments and the second popup menu shows the starts and the ends of segments. They are placed in area called *Prepared segments*. Under this there is area called *Manual selection*. It contains two input fields, the start of segment and the end of segment in minutes, and the third button which executes the selection of segment. The last element is check box. If it's checked, outliers are removed.

HRV analysis starts after clicking on the button *HRV Analysis*. After few seconds, it depends on the computers speed, appears the new window with result of analysis for selected segment (Fig. 4.5).

4.5 Batch computation

This version of program is assigned for easier computation of a large number of data files. RR intervals and HRV analysis are computed from all data files placed in the folder which is specified at the beginning of file named *ExtractFeatures_batch.m*. This file is the main file which can be run for computation all analysis.

The results are saved in file *HRV_an_res_xxx_all_yy_min_segs.mat* which it is placed in the folder called *results* in the folder of input data. The name of file contains the name of original file (*xxx* means the name) and the length of segments (*yy*).
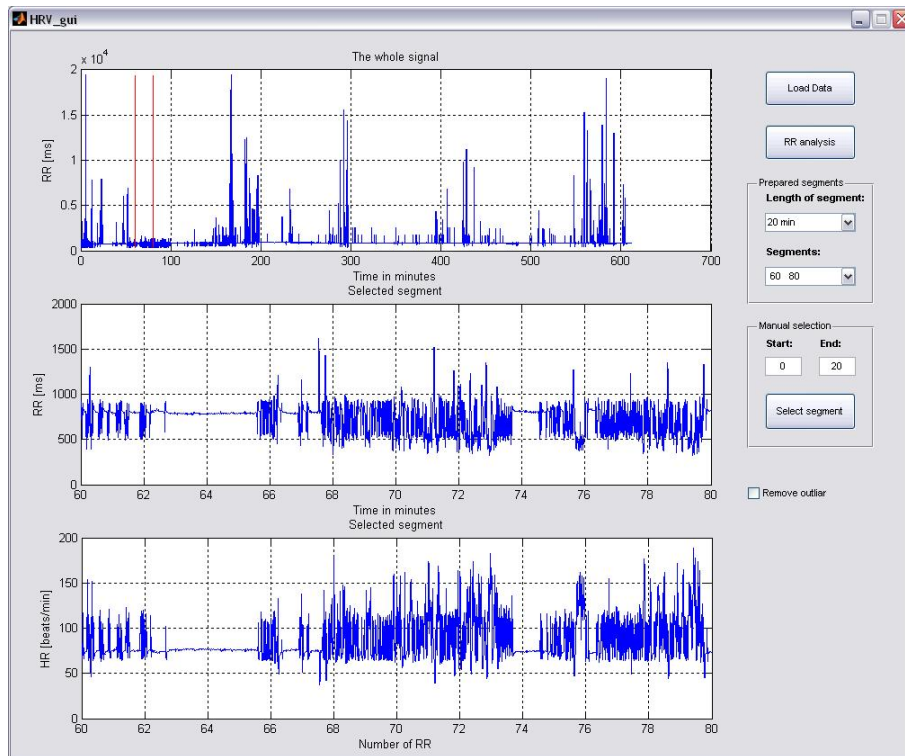
Figure 4.4: Starting window of GUI.
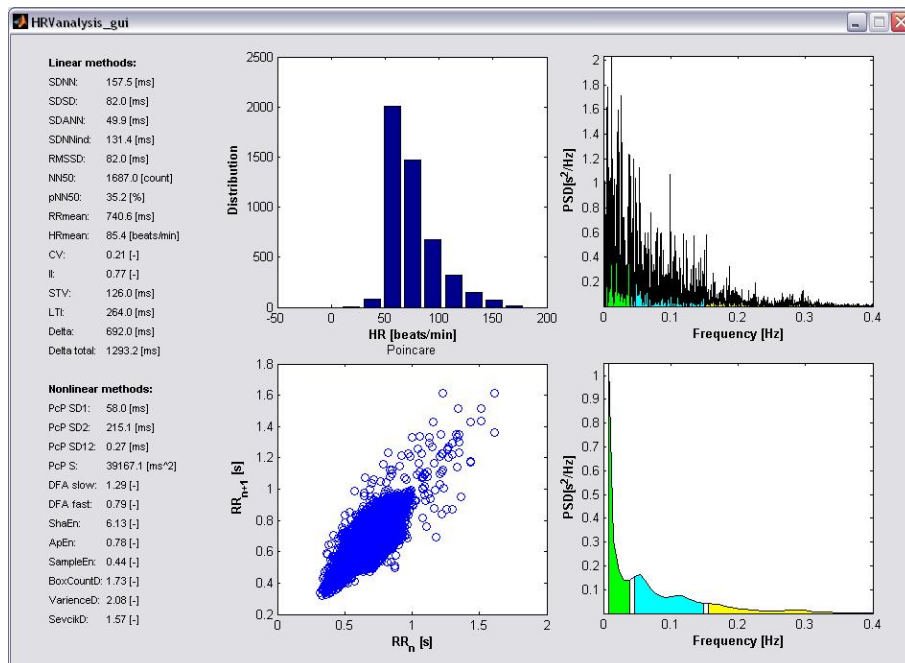
Figure 4.5: GUI with results of HRV analysis.

Chapter 5

Results

The task was to assess the features for distinguishing between subject with AF and subject with NSR. Real data were used from Physinet databases. Description of databases is in chapter 4. Total number of 10 ten-hour samples with AF and 10 twentyfour-hour samples with NSR were collected (see Fig. 5.1), and spanned into segments (see Fig. 5.2). The data were divided into segments of four lengths for reasons explained in section 4.3. The aim was to find significant features for recognition of AF.

All methods explained in chapter 3 were used as a features for characterization of heart rate. Features were computed for all segments. Total number of five-minute segments was 3529 from that 1139 for AF and 2390 for NSR, total number of ten-minute segments was 1757 from that 567 for AF and 1190 for NSR, total number of twenty-minute segments was 871 from that 281 for AF and 590 for NSR and total number of thirty-minute segments was 578 from that 186 for AF and 390 for NSR. Mean values and standard deviations were computed from extracted features for each segmentation and each feature.

For signification of results was used Wilcoxon rank sum test also known as Mann-Whitney U-test. This test is implemented in Matlab by function *runksum*. It performs a two-sided rank sum test of the null hypothesis that data in the vectors x (group of features from AF subjects) and y (group of features from NSR subjects) are independent samples from identical continuous distributions with equal medians,

(a) Subject with NSR   (b) Subject with AF
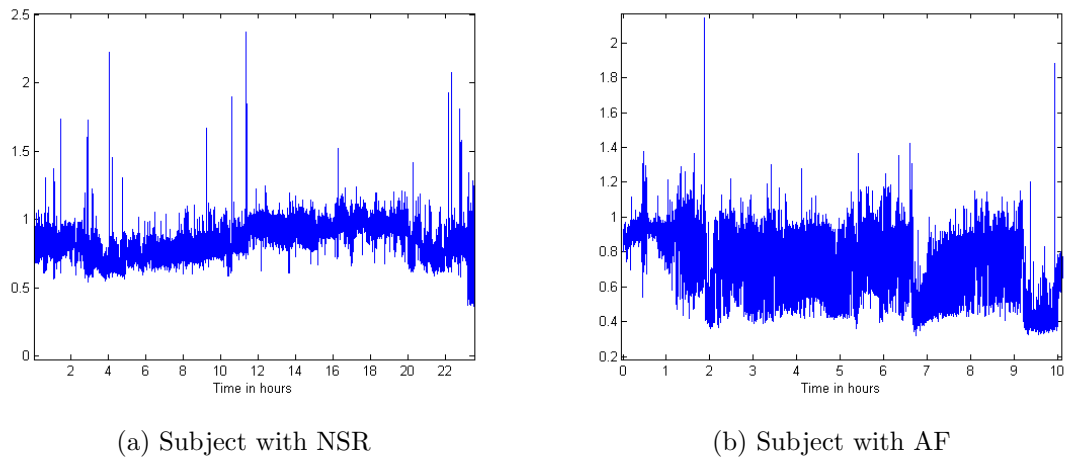
Figure 5.1: All RR intervals.

(a) Subject with NSR   (b) Subject with AF
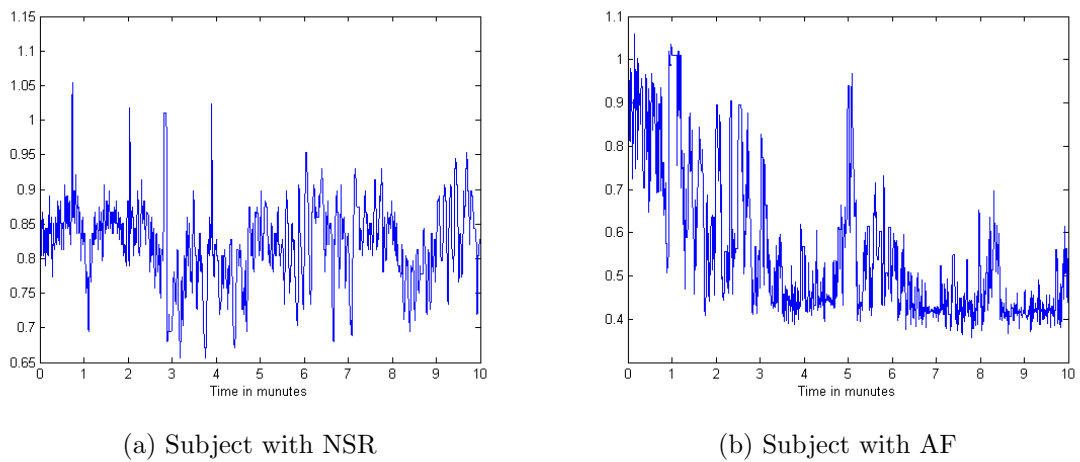
Figure 5.2: Ten-minute segment of RR intervals.

against the alternative that they do not have equal medians. x and y can have different lengths. The result of the test is P-value. If P-value is $p < 0,05$, it means the null hypothesis is rejected at the 5% significance level thus the feature is significant. In other case the feature isn't significant.

The results for five-minute segmentation shows table 5.1. They are in the form $mean \pm SD$. From all features only two features, coefficient of variation (CV) with P-value 0,1941 and $Delta$ with P-value $0,1323$ weren't significant. The best results were achieved by $DFAslow$ and $PcPSD12$. $DFAslow$ with values $1,46 \pm 0,18$ for atrial fibrilation and $1,61 \pm 0,14$ for sinus rhythm could be used for identification of atrial fibrilation based on setting the threshold of mean value. P-value for this parameter is 8,53e-144. Analogous to the second best parameter $PcPSD12$ with values $0,17 \pm 0,07$ for AF and $0,12 \pm 0,04$ for NSR. Also very good results had features: $NN50$, $pNN50$, LTI and $ShaEn$. All other features had also good results but the significance level was less.

For ten-minute segmentation was majority of features also significant (see table 5.2), but this segmentation had worse results. The only exception makes features: $SDNN$ with P-value 0,2096, CV with P-value 0,1112, $Delta$ with P-value 0,1570, $PcPSD2$ with P-value 0,1564 and $SevcikD$ with P-value 0,6870. $DFAslow$ appears the best parameter again with P-value 1,87e-73.

The results for twenty-minute segmentation shows the table 5.3. As you can see the results for twenty-minute segmentation are worse than in both previous segmentations. Of 11 designed nonlinear methods are 6 significant. Designed features from fractal analysis completely fail, P-values are greater than 0,3. Linear methods have slightly better results. Of 15 designed linear methods are 9 significant. The best feature was $DFAslow$ again with P-value 2,14e-36.

Results for thirty-minute segments turned out to be the worst. Comparison of results is presented in table 5.4. Of all twenty-six features that they've passed the test were only fifteen with P-value smaller then $0,05$.

Achieved results indicates the five-minute segmentation to be the best choice. The reason could be greater stationarity stability of signal for shorter segments.

Table 5.1: The results for five-minute segmentation

Features	Atrial fibrillation	Normal sinus rhythm	P-value
Linear methods			
$NN50$ (-)	$82{,}53 \pm 87{,}46$	$41{,}43 \pm 51{,}88$	5,39e-31
$pNN50$ (%)	$6{,}88 \pm 7{,}29$	$3{,}46 \pm 4{,}33$	5,39e-31
$RMSSD$ (ms)	$31{,}84 \pm 39{,}37$	$20{,}88 \pm 17{,}44$	5,06e-17
$SDNN$ (ms)	$207{,}26 \pm 653{,}90$	$104{,}70 \pm 188{,}77$	8,56e-03
$SDSD$ (ms)	$31{,}85 \pm 39{,}37$	$20{,}89 \pm 17{,}45$	5,06e-17
$SDANN$ (ms)	$121{,}89 \pm 549{,}61$	$47{,}40 \pm 126{,}26$	5,54e-06
$SDNNind$ (ms)	$119{,}29 \pm 264{,}01$	$78{,}66 \pm 99{,}69$	9,08e-04
$RRmean$ (ms)	$894{,}76 \pm 642{,}80$	$885{,}88 \pm 232{,}47$	1,12e-13
$HRmean$ (bpm)	$78{,}44 \pm 18{,}95$	$71{,}61 \pm 12{,}67$	5,01e-17
CV (-)	$0{,}15 \pm 0{,}22$	$0{,}11 \pm 0{,}11$	1,94e-01
II (-)	$0{,}39 \pm 0{,}22$	$0{,}43 \pm 0{,}18$	1,30e-18
STV (ms)	$56{,}09 \pm 46{,}82$	$55{,}78 \pm 35{,}43$	3,22e-05
LTI (ms)	$282{,}46 \pm 1424{,}61$	$155{,}37 \pm 416{,}64$	3,33e-31
$Delta$ (ms)	$451{,}11 \pm 653{,}72$	$331{,}07 \pm 289{,}71$	1,32e-01
$Delta\_total$ (ms)	$918{,}43 \pm 1857{,}99$	$569{,}67 \pm 670{,}99$	5,26e-05
Nonlinear methods			
$PcPSD1$ (ms)	$22{,}52 \pm 27{,}84$	$14{,}77 \pm 12{,}34$	5,06e-17
$PcPSD2$ (ms)	$291{,}59 \pm 924{,}22$	$147{,}18 \pm 266{,}69$	3,82e-03
$PcPSD12$ (-)	$0{,}17 \pm 0{,}07$	$0{,}12 \pm 0{,}04$	1,15e-101
$PcPS$ (ms^2)	$93932{,}64 \pm 590227{,}78$	$16030{,}57 \pm 134365{,}32$	1,07e-02
$DFAslow$ (-)	$1{,}46 \pm 0{,}18$	$1{,}61 \pm 0{,}14$	8,53e-144
$DFAfast$ (-)	$1{,}02 \pm 0{,}35$	$1{,}16 \pm 0{,}24$	5,22e-08
$ShaEn$ (-)	$5{,}16 \pm 0{,}90$	$5{,}47 \pm 0{,}53$	3,36e-20
$ApEn$ (-)	$0{,}65 \pm 0{,}32$	$0{,}61 \pm 0{,}22$	5,22e-08
$SampEn$ (-)	$0{,}64 \pm 0{,}40$	$0{,}57 \pm 0{,}27$	2,85e-06
$BoxCountD$ (-)	$1{,}57 \pm 0{,}14$	$1{,}56 \pm 0{,}09$	1,57e-13
$SevcikD$ (-)	$1{,}42 \pm 0{,}10$	$1{,}42 \pm 0{,}06$	1,97e-02

Table 5.2: The results for ten-minute segmentation

Features	Atrial fibrillation	Normal sinus rhythm	P-value
Linear methods			
$NN50$ (-)	$165{,}17 \pm 171{,}41$	$82{,}72 \pm 100{,}25$	9,50e-18
$pNN50$ (%)	$6{,}88 \pm 7{,}15$	$3{,}45 \pm 4{,}18$	9,50e-18
$RMSSD$ (ms)	$33{,}35 \pm 38{,}21$	$21{,}47 \pm 16{,}74$	6,94e-12
$SDNN$ (ms)	$245{,}52 \pm 700{,}13$	$116{,}59 \pm 204{,}59$	2,10e-01
$SDSD$ (ms)	$33{,}35 \pm 38{,}21$	$21{,}47 \pm 16{,}74$	6,95e-12
$SDANN$ (ms)	$159{,}05 \pm 570{,}14$	$59{,}92 \pm 147{,}77$	9,27e-04
$SDNNind$ (ms)	$122{,}26 \pm 231{,}48$	$79{,}38 \pm 90{,}45$	3,49e-02
$RRmean$ (ms)	$895{,}59 \pm 580{,}88$	$885{,}64 \pm 213{,}20$	2,77e-06
$HRmean$ (bpm)	$78{,}39 \pm 18{,}54$	$71{,}64 \pm 12{,}43$	4,42e-09
CV (-)	$0{,}18 \pm 0{,}28$	$0{,}12 \pm 0{,}12$	1,11e-01
II (-)	$0{,}42 \pm 0{,}23$	$0{,}49 \pm 0{,}20$	1,04e-10
STV (ms)	$56{,}36 \pm 43{,}25$	$55{,}99 \pm 34{,}02$	8,66e-03
LTI (ms)	$208{,}04 \pm 676{,}08$	$165{,}60 \pm 421{,}28$	1,06e-15
$Delta$ (ms)	$451{,}84 \pm 573{,}24$	$331{,}21 \pm 257{,}33$	1,57e-01
$Deltatotal$ (ms)	$1230{,}34 \pm 2275{,}07$	$729{,}24 \pm 856{,}81$	8,83e-05
Nonlinear methods			
$PcPSD1$ (ms)	$23{,}58 \pm 27{,}02$	$15{,}18 \pm 11{,}84$	6,95e-12
$PcPSD2$ (ms)	$345{,}80 \pm 989{,}80$	$164{,}03 \pm 289{,}07$	1,56e-01
$PcPSD12$ (-)	$0{,}16 \pm 0{,}06$	$0{,}11 \pm 0{,}04$	3,33e-44
$PcPS$ (ms^2)	$102554{,}44 \pm 556775{,}30$	$17314{,}52 \pm 120073{,}74$	1,10e-02
$DFAslow$ (-)	$1{,}47 \pm 0{,}18$	$1{,}61 \pm 0{,}13$	1,87e-73
$DFAfast$ (-)	$1{,}06 \pm 0{,}32$	$1{,}18 \pm 0{,}23$	9,17e-20
$ShaEn$ (-)	$5{,}28 \pm 0{,}87$	$5{,}56 \pm 0{,}51$	3,87e-10
$ApEn$ (-)	$0{,}63 \pm 0{,}33$	$0{,}58 \pm 0{,}22$	6,37e-04
$SampEn$ (-)	$0{,}58 \pm 0{,}38$	$0{,}52 \pm 0{,}26$	3,23e-02
$BoxCountD$ (-)	$1{,}58 \pm 0{,}13$	$1{,}58 \pm 0{,}08$	2,3e-02
$SevcikD$ (-)	$1{,}43 \pm 0{,}10$	$1{,}44 \pm 0{,}07$	6,87e-01

Table 5.3: The results for twenty-minute segmentation

Features	Atrial fibrillation	Normal sinus rhythm	P-value
Linear methods			
$NN50$ (-)	$331{,}80 \pm 337{,}23$	$164{,}32 \pm 193{,}12$	1,10e-10
$pNN50$ (%)	$6{,}91 \pm 7{,}03$	$3{,}42 \pm 4{,}02$	1,10e-10
$RMSSD$ (ms)	$35{,}24 \pm 36{,}71$	$22{,}02 \pm 16{,}09$	4,55e-08
$SDNN$ (ms)	$305{,}77 \pm 775{,}69$	$130{,}43 \pm 214{,}77$	7,46e-01
$SDSD$ (ms)	$35{,}25 \pm 36{,}71$	$22{,}03 \pm 16{,}09$	4,55e-08
$SDANN$ (ms)	$217{,}04 \pm 661{,}48$	$73{,}74 \pm 159{,}24$	9,43e-02
$SDNNind$ (ms)	$121{,}10 \pm 188{,}48$	$79{,}09 \pm 77{,}62$	4,04e-01
$RRmean$ (ms)	$896{,}07 \pm 448{,}79$	$885{,}02 \pm 195{,}40$	3,54e-03
$HRmean$ (bpm)	$78{,}42 \pm 18{,}17$	$71{,}68 \pm 12{,}08$	5,90e-05
CV (-)	$0{,}23 \pm 0{,}35$	$0{,}14 \pm 0{,}14$	5,43e-02
II (-)	$0{,}44 \pm 0{,}24$	$0{,}51 \pm 0{,}22$	4,67e-05
STV (ms)	$56{,}69 \pm 40{,}26$	$55{,}98 \pm 33{,}82$	1,44e-01
LTI (ms)	$244{,}97 \pm 977{,}57$	$163{,}71 \pm 210{,}40$	3,90e-08
$Delta$ (ms)	$452{,}94 \pm 499{,}54$	$330{,}70 \pm 232{,}76$	2,03e-01
$Deltatotal$ (ms)	$1714{,}06 \pm 2982{,}12$	$948{,}34 \pm 1110{,}66$	1,28e-03
Nonlinear methods			
$PcPSD1$ (ms)	$24{,}92 \pm 25{,}96$	$15{,}57 \pm 11{,}38$	4,55e-08
$PcPSD2$ (ms)	$431{,}06 \pm 1096{,}74$	$183{,}64 \pm 303{,}48$	6,78e-01
$PcPSD12$ (-)	$0{,}14 \pm 0{,}06$	$0{,}11 \pm 0{,}03$	7,07e-19
$PcPS$ (ms^2)	$114027{,}25 \pm 462114{,}85$	$18608{,}52 \pm 99019{,}14$	1,18e-02
$DFAslow$ (-)	$1{,}48 \pm 0{,}17$	$1{,}61 \pm 0{,}12$	2,14e-36
$DFAfast$ (-)	$1{,}11 \pm 0{,}29$	$1{,}20 \pm 0{,}22$	1,75e-07
$ShaEn$ (-)	$5{,}39 \pm 0{,}84$	$5{,}64 \pm 0{,}50$	3,56e-05
$ApEn$ (-)	$0{,}59 \pm 0{,}34$	$0{,}55 \pm 0{,}22$	6,38e-02
$SampEn$ (-)	$0{,}52 \pm 0{,}37$	$0{,}48 \pm 0{,}24$	6,60e-01
$BoxCountD$ (-)	$1{,}58 \pm 0{,}13$	$1{,}60 \pm 0{,}08$	9,27e-01
$SevcikD$ (-)	$1{,}44 \pm 0{,}10$	$1{,}45 \pm 0{,}07$	3,75e-01

Table 5.4: The results for thirty-minute segmentation

Features	Atrial fibrillation	Normal sinus rhythm	P-value
Linear methods			
$NN50$ (-)	$499{,}16 \pm 500{,}50$	$244{,}86 \pm 279{,}39$	4,46e-08
$pNN50$ (%)	$6{,}93 \pm 6{,}95$	$3{,}40 \pm 3{,}88$	4,46e-08
$RMSSD$ (ms)	$35{,}38 \pm 35{,}59$	$22{,}29 \pm 15{,}75$	5,97e-06
$SDNN$ (ms)	$323{,}16 \pm 783{,}78$	$136{,}62 \pm 215{,}03$	8,43e-01
$SDSD$ (ms)	$35{,}38 \pm 35{,}59$	$22{,}29 \pm 15{,}75$	5,97e-06
$SDANN$ (ms)	$234{,}31 \pm 670{,}56$	$80{,}40 \pm 158{,}72$	2,25e-01
$SDNNind$ (ms)	$119{,}97 \pm 184{,}34$	$79{,}36 \pm 75{,}98$	3,10e-01
$RRmean$ (ms)	$894{,}59 \pm 407{,}85$	$884{,}28 \pm 192{,}32$	1,50e-02
$HRmean$ (bpm)	$78{,}44 \pm 17{,}87$	$71{,}75 \pm 11{,}93$	1,19e-03
CV (-)	$0{,}24 \pm 0{,}36$	$0{,}14 \pm 0{,}14$	1,03e-01
II (-)	$0{,}44 \pm 0{,}24$	$0{,}50 \pm 0{,}22$	2,49e-03
STV (ms)	$56{,}57 \pm 39{,}55$	$55{,}92 \pm 32{,}16$	1,97e-01
LTI (ms)	$241{,}92 \pm 870{,}35$	$168{,}00 \pm 227{,}80$	4,89e-07
$Delta$ (ms)	$448{,}90 \pm 482{,}51$	$330{,}74 \pm 227{,}83$	3,52e-01
$Deltatotal$ (ms)	$1996{,}10 \pm 3380{,}70$	$1100{,}85 \pm 1239{,}92$	5,36e-03
Nonlinear methods			
$PcPSD1$ (ms)	$25{,}02 \pm 25{,}17$	$15{,}76 \pm 11{,}14$	5,97e-06
$PcPSD2$ (ms)	$455{,}72 \pm 1108{,}27$	$192{,}44 \pm 303{,}90$	7,73e-01
$PcPSD12$ (-)	$0{,}13 \pm 0{,}06$	$0{,}10 \pm 0{,}03$	1,04e-11
$PcPS$ (ms^2)	$114036{,}63 \pm 414317{,}77$	$18978{,}64 \pm 90801{,}49$	3,05e-02
$DFAslow$ (-)	$1{,}49 \pm 0{,}16$	$1{,}62 \pm 0{,}11$	2,02e-27
$DFAfast$ (-)	$1{,}13 \pm 0{,}27$	$1{,}22 \pm 0{,}21$	3,94e-06
$ShaEn$ (-)	$5{,}45 \pm 0{,}81$	$5{,}69 \pm 0{,}49$	4,03e-04
$ApEn$ (-)	$0{,}57 \pm 0{,}33$	$0{,}53 \pm 0{,}21$	1,28e-01
$SampEn$ (-)	$0{,}48 \pm 0{,}35$	$0{,}45 \pm 0{,}23$	9,20e-01
$BoxCountD$ (-)	$1{,}60 \pm 0{,}13$	$1{,}63 \pm 0{,}08$	1,19e-01
$SevcikD$ (-)	$1{,}45 \pm 0{,}10$	$1{,}46 \pm 0{,}07$	4,21e-01

The best feature from all is $DFAslow$ for all segmentation. It could be used for automated detection of AF, based on the choice of right threshold of mean value. In comparison with the results from the article [7] I have reached better results. At least suitable methods for detection AF are CV, $Delta$ and $SevcikD$.

Finally, the simple classification of data was undertaken for evaluation of the practical usability of the features. As a result classification using arbitrary decision tree in WEKA brought us results of 92,1 % specificity and 92 % sensitivity in distinguishing between atrial fibrillation and normal sinus rhythm. In comparison with results from article [15] was reached similar results.

Chapter 6

Conclusion

After familiarizing with the state of the art papers different linear and nonlinear methods for heart rate variability analysis were implemented. From large group of linear methods fifteen most used features were selected and from nonlinear methods eleven most used features were computed. They are described in a chapter 3. The aim of this work was not only theoretical analysis of the features but also part of the goal was to create tool for HRV analysis and to apply it for assessment of features which could be used for recognition of atrial fibrilation.

Before extraction of features preprocessing of the data was necessary such as detection of R peaks, RR intervals computation, removing outliar and resampling, the details can be found in section 4.2. Graphical user interface was created for easier operating with data and computing features of HRV analysis. The GUI was supplemented with the batch computation for large amount of data. Both programmed application are available on the enclosed CD. CD also contains HTML-based manual.

The main aim of this work was to assess features for recognition atrial fibrilation from normal sinus rhythm. The features were extracted from real data downloaded from website of Physionet. Before execution of HRV analysis the data was segmented into five, ten, twenty and thirty minute segments. The features were calculated for all types of segmentation. After execution of Mann-Whitney statistical test we got very interesting results. The five-minute segmentation shows to be the best. For

that segmentation the significance of the features was the biggest. From all features the best were slow detrended fluctuation with P-value $8, 53e - 144$ and ratio of $SD1$ to $SD2$ of Poincaré plot called $PcPSD12$ with P-value $1, 15e - 101$. Other features had also good results of significance. For example $NN50$ wit P-value $5, 39e - 31$, pNN50 with the same P-value, LTI with P-value $3, 33e - 31$ and Shannon entropy with P-value $3, 36e - 20$. The exceptions were created by nonsignificant features CV with P-value $0, 194$ and delta with P-value $0, 132$. Better results were reached then in article [7].

Finally the classification of data was undertaken for evaluation of the practical usability of the features. The classification using arbitrary decision tree in WEKA brought us results of atrial fibrilation correct classification with the specificity 92,1 % and the sensitivity 92 % which is very similar with results reached in e.g. article [15].

Future work

In the future there is possibility to upgrade the application with addition of another features to create more complex tool. To reach better results it might be possible to try the adaptive segmentation approach or better preprocessing.

Bibliography

[1] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. C. Barger, and R. J. Cohen. Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat to beat cardiovascular control. *Science*, 213:220–222, 1981.

[2] R. Berger, S. Akselrod, D. Gordon, and R. Cohen. An efficient algorithm for spectral analysis of heart rate variability. *IEEE Trans Biomed Eng.*, 33:900–904, 1986.

[3] H.F.V. Boshoff. A fast box counting algorithm for determining the fractal dimension of sampled continuous functions. *IEEE*, 92:43–48, 1992.

[4] Peng C-K., Havlin S., Stanley H. E., and Goldberger A. L. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5:82–87, 1995.

[5] Peng C-K., Buldyrev S. V., Havlin S., Simons M., Stanley H. E., and Goldberger A. L. Mosaic organization of dna nucleotides. *Phys Rev E*, 49:1685–1689, 1994.

[6] S. Castiglioni and M. Di Rienzo. How the threshold "r" influences approximate entropy analysis of heart-rate variability. *Computers in Cardiology*, 35:561–564, 2008.

[7] D. Chamchad, G. Djaiani, H. J. Jung, L. Nakhamchik, J. Carroll, and J. C. Horrow. Nonlinear heart rate variability analysis may predict atrial fibrillation after coronary artery bypass grafting. *International Anesthesia Research Society*, 103(5):1109–1112, November 2006.

[8] J. P. Marques de Sá. Characterization of fetal heart rate using approximate entropy. *Computers in Cardiology*, 32:671–673, 2005.

[9] M. M. Dodson and S. Kristensen. Hausdorff dimension and diophantine approximation. *American Mathematical Society*, pages 1–43, June 2003.

[10] EMIS and PiP. Atrial fibrillation, 2009. Patient UK, http://www.patient.co.uk.

[11] D. J. Ewing, C. N. Martin, R. J. Young, and B. F. Clarke. The value of cardiovascular autonomic function tests: 10 years experience in diabetes. *Diabetic Care*, 8:491–198, 1985.

[12] G. Georgoulas, Ch. D. Stylios, and P. P. Groumpos. Predicting the risk of metabolic acidosis for newborns based on fetal heart signal classification using support vector machines. *IEEE Transaction on Biomedical Engineering*, 53(5), May 2006.

[13] H. Gonçalves, A. P. Rocha, D. Ayeres de Campos, and J. Bernardes. Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes precending delivery. *Med Bio Eng Comput*, 44:847–855, 2006.

[14] E.H Hon and S.T. Lee. Electronic evaluations of the fetal heart rate patterns preceding fetal death, further observations. *Am J Obstet Gynec*, 87:814–826, 1965.

[15] N. Kikillus, G. Hammer, N. Lentz, F. Stockwald, and A. Bolz. Three different algorithms for identifying patients suffering from atrial fibrillation during atrial fibrillation free phases of the ecg. *Computers in Cardiology*, 34:801–804, 2007.

[16] R. E. Kleiger, J. P. Miller, J. T. Bigger, A. J. Moss, and the Multicenter Post Infarction Research. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am J Cardiol*, 59:256–262, 1987.

[17] Goldberger A. L., Amaral L. A. N., Glass L., Hausdorff J. M., Ivanov P. Ch., Mark R. G., Mietus J. E., Moody G. B., and Peng C-K. Components of a

new research resource for complex physiologic signals, June 2000. Circulation Electronic Pages, http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[18] C. Lerma, O. Infante, H. Pérez-Grovas, and M. V. José. Poincaré plot indexes of heart rate variability capture dynamic adaptations after haemodialysis in chronic renal failure patients. *Clin Physiol and Func Im*, 23:72–80, 2003.

[19] The MathWorks. Matlab and simulink version 2007b, 2009. MathWorks, http://www.mathworks.com.

[20] E. Naujokat, M. Arndt, D. Novak, and CH. Norra. A new data acquisition system for monitoring circadian variations of activity and ecg: Clinical application. *World Association of Sleep Medicine Congress, http://www.wasm2005.org/*, October 15-18, 2005. Berlin, Germany.

[21] Task Force of The European Society of Cardiology, The North American Society of Pacing, and Electrophysiology. Heart rate variability. *European Heart Journal*, 17:354–381, 1996.

[22] J. Pan and W. Tompkins. A real-time qrs detection algorithm. *IEEE Trans Biom Eng*, 32(3):230–236, 1985.

[23] MA Physionet. Cambridge. Physionet [home page on the internet], updated 2009. PhysioNet, http://www.physionet.org.

[24] S. Pincus. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A*, 88:2297–2301, 1991.

[25] J. Piskorski and P. Guzik. Filtering poincaré plots. *Computation Methods in Science and Technology*, 11(1):39–48, June 2005.

[26] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Physiological Society and Heart and Circulatory Physiology*, 278:H2039–H2049, 2000.

[27] B. M. Sayers. Analysis of heart rate variability. *Ergonomics*, 16:17–32, 1973.

[28] C. Sevcik. A procedure to estimate the fractal dimension of waveforms. *Complexity International*, 5:1–16, 1998.

[29] C. E. Shannon. Prediction and entropy of printed english. *The Bell System Technical Journal*, 30, January 1951.

[30] P. van Leeuwen and H. Bettermann. The status of nonlinear dynamics in the analysis of heart rate variability. *Herzschr Elektrophys*, 10:127–130, 1999.

[31] M. M. Wolf, G. A. Varigos, D. Hunt, and J. G. Sloman. Sinus arrhythmia in acute myocardial infarction. *Med J Australia*, 2:52–53, 1978.

Appendix A

Contents of enclosed CD

To this work is available on enclosed CD, which contains source codes and electronic appearance of this work.

- **Diploma Thesis:** Diploma thesis in electronic appearance

- **Documentation:** Documentation of application

- **HRVanalysis:** The main application