**TECHNISCHE
UNIVERSITÄT
WIEN**
Vienna University of Technology

Diplomarbeit

Determination and Evaluation of Heart Rate Variability Parameters with Focus on Nonlinear Methods

ausgeführt am Institut für

Analysis and Scientific Computing

der Technischen Universität Wien

unter Anleitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.tech. Felix Breitenecker

durch

Stefan Hagmair
Rauscherstr. 10/22
1200 Wien

Wien, September 2015	
Ort, Datum	Unterschrift

ABSTRACT

More than 50 years ago, during the investigation of fetal distress, it was realized that the variation of the heart rate, i.e., Heart Rate Variability (HRV), is a marker of the health status, since it unveils changes in beat-to-beat variation of the heart, even before there was a remarkable change in heart rate itself. HRV reflects the balance between the sympathetic and the parasympathetic nervous system. Furthermore, several physiological effects, which influence the normal rhythm of the heart, are manifested by HRV. Since the heart rate itself is nonstationary and the structure generating the signal involves nonlinear contributions, nonlinear methods to quantify the variability of the heart rate gained interest over the last years. Cardiovascular Diseases (CVD) are more common and their occurrences are increasing since centuries. HRV analysis is a useful noninvasive tool for early detection and the prevention of CVD.

The goal of this thesis is to implement numerous indices to quantify HRV derived from mathematical models and compare them to each other in different test cases. Most of them belong to the section of nonlinear methods, though some other standard measures, as statistical parameters and one index of the time-frequency domain, are calculated. The implemented methods are tested on their ability to differentiate between healthy and pathological subjects. Furthermore, their sensitivity to a varying data length is investigated. In addition, the HRV measures are tested if there is a difference between young and elderly people. The last test case examines subjects with ventricular arrhythmias. The models are applied to baseline data and on-therapy data, i.e., during medication, of the same subject in order to detect effects of antiarrhythmic treatment.

The results show that all of the implemented indices are able to differentiate between nonpathological and pathological subjects. Most of them show a significant difference before and after antiarrhythmic treatment, though no index is sensitive to age. Their robustness to varying lengths of recordings is formidable.

The well-trodden statistical indices justified their existence with significant differences in all, except the age-dependency, test cases. However, they are strongly correlated to each other. Apart from the age-dependency test case, all of the fractal indices show thoroughly remarkable results, too. Just one of them found no significant differences before and after anti-

arrhythmic treatment. Two of them are independent for all the applied data.

In summary, this thesis shows that fractal descriptors are an appropriate support for analyzing the HRV, and therefore to prevent or detect cardiovascular diseases. Especially the Hurst exponent, well established in the financial community, should get more attention in analyzing biomedical signals, such as HRV.

KURZFASSUNG

Bereits vor 50 Jahren wurde beobachtet, dass Schwankungen der Herzrate, die sogenannte Herzratenvariabilität (HRV), ein Indikator für den Gesundheitszustand ist. Diese Beobachtung wurde bei der Untersuchung von Komplikationen in der Schwangerschaft gemacht, als eine zeitliche Veränderung zwischen zwei Grundschlägen (RR-Intervall) gefunden wurde, bevor man eine auffällige Veränderung der Herzschlagfrequenz selbst wahrnehmen konnte. Das Gleichgewicht zwischen dem sympathischen und parasympathischen Nervensystem wird durch die HRV widergespiegelt. Darüber hinaus manifestieren sich noch weitere physiologische Effekte in der HRV, die den normalen Herzrhythmus beeinflussen. Der vermehrte Einsatz von nichtlinearen Methoden zur Quantifizierung der HRV in den letzten Jahren ist darauf zurückzuführen, dass die Herzschlagfrequenz selbst ein nichtstationärer Prozess ist und komplexe, nichtlineare Einflüsse den Herzschlag regeln. Herz-Kreislauf-Erkrankung gehören zu den häufigsten Todesursachen und die Anzahl von Erkrankungen steigt seit Jahrzehnten an. Die HRV Analyse ist ein wichtiges, nichtinvasives Hilfsmittel zur Früherkennung und Vorbeugung von Herz-Kreislauf Erkrankungen.

Das Ziel dieser Arbeit ist die Implementierung von zahlreichen Indizes zur Qunatifizierung der HRV basierend auf mathematischen Modellen. Danach werden diese in verschiedenen Testszenarien miteinander verglichen. Die meisten der implementierten Indizes kommen aus dem Bereich der nichtlinearen Methoden. Es wurden jedoch auch Standardmethoden, wie beispielsweise statistische Parameter, sowie ein Index aus dem Zeit-Frequenz Bereich berechnet. Die implementierten Parameter wurden auf ihre Fähigkeit untersucht, gesunde Daten von pathologischen Daten zu unterscheiden. Weiters wurden die Modelle auf ihre Abhängigkeit von der Datenlänge, welche der EKG-Dauer entspricht, getestet. Ein weiteres Testszenario stellte die Unterscheidung von gesunden jungen und alten Probanden dar. Der letzte Test befasste sich mit Probanden, welche an ventrikulären Herzrhythmusstörungen leiden. Die Modelle errechnen dabei die Parameter aus den Daten der Probanden, einerseits vor dem Beginn einer medikamentösen Behandlung, andererseits nach der Behandlung. Die Modelle wurden also darauf getestet, ob sie einen Behandlungseffekt widerspiegeln können.

Die Testergebnisse zeigen, dass alle implementierten Parameter signifikant zwischen gesunden und pathologischen Probanden unterscheiden können. Die meisten Modelle haben auch signifikante Unterschiede bei Probanden in der Herzrythmusstörungsbehandlung aufgezeigt. Kein einziger Parameter konnte nach dem Alter differenzieren. Die Robustheit der Modelle gegenüber unterschiedlicher Datenlängen ist hingegen beeindruckend.

Die altbekannten statistischen Parameter haben mit signifikanten Ergebnissen in allen Testszenarien, außer dem Alterstest, ihr Dasein gerechtfertigt. Diese korrelieren alle sehr stark miteinander. Abgesehen vom Alterstest haben die fraktalen Parameter auch durch die Bank hervorragende Ergebnisse geliefert. Nur ein Index aus jener Gruppe fand keine signifikanten Unterschiede bei Probanden in Herzrythmusstörungsbehandlung. Zwei der fraktalen Parameter sind unabhängig voneinander.

Zusammenfassend zeigt diese Arbeit, dass fraktale Parameter eine geeignete Unterstützung für die Analyse der HRV sind und sich somit Herz-Kreislauf Erkrankungen früh erkennen lassen. Vor allem der Hurst Exponent, welcher in der Finanzwelt schon sehr etabliert ist, sollte mehr Aufmerksamkeit bei der Analyse von biomedizinischen Signalen, wie HRV, erhalten.

ACKNOWLEDGEMENTS

First of all, I want to thank Prof. Felix Breitenecker, whose lively lectures got me connected to the Austrian Institute of Technology where this thesis was carried out. Furthermore, I enjoyed being part of the organized summerschools, conferences and social events. It is a pleasure to join the ARGESIM group.

I am grateful to my colleagues at the AIT. I almost always came with a smile and even on days I appeared a bit frustrated, I left the office with a smile on my face. Above all, "hearty" thanks to my supervisors, Christopher Mayer and Martin Bachler for their excellent support. Without them, this thesis would not have been done in time.

I am deeply grateful to all of my family. Especially my parents Brigitte and Franz, who gave me the opportunity to do my studies and even more, to focus on my studies. I can always rely on their support in all circumstances.

Of course, I did not master my years of study alone. In the first term a group of adorable members, the "Mathe-Homies" arose. With advanced time, our "quality time" became a bit less, but it is still beautiful every time we convene.
People of the "MittwochsFilm-FilmMittwoch" are the source of fruitful discussions, nice movies and delicious food.
Beside all the brain work, time in the mountains had been very important for me. Thanks Claudia and Kathi for the nice "Almsommer" in 2012 and the "MTB-crew" for the best holidays in the Soča valley and Val Gardena. It should be mentioned, that profound thoughts have had space as well.

At least two more people are not covered in all the called groups above. Thanks to Jos and Maria. Jos for being a unconventional thinker and encouraging me every time we meet up and Maria for being a friend in every condition of life.

Last but not least, thanks to the large number of notes, taking care of that everything is included in my thesis.

Sorry for being stressed and chaotic over the last weeks,

Merci,   *Stefan*
        Indi/Este/Hacke

CONTENTS

1 | INTRODUCTION

The "European Society of Cardiology" reports in the European Cardiovascular Diseases (CVD) Statistics 2012 that cardiovascular disease causes 47% of all deaths in Europe. Furthermore, CVD is the main cause of death in women in all countries of Europe and is the main cause of death in men in all but six countries [61].

Several papers showed the significant relationship between Autonomic Nervous System (ANS) and cardiovascular mortality [6]. Heart Rate Variability (HRV) analysis is a useful tool for understanding the status of the ANS. Furthermore, HRV is an early predictor of cardiac failures [23]. A plurality of HRV measures in different domains, e. g., time domain, frequency domain, and nonlinear techniques have been proposed. In order to standardize initial normative values and standard methods of measurement, the "European Society of Cardiology" and the "North American Society of Pacing and Electrophysiology" formed a Task Force and released "Guidelines - Heart Rate Variability" [86].

The complex origin of biomedical signal limits this traditional linear approaches [82]. Hence, nonlinear methods gained recent interest, in order to reveal more information embedded in the Heart Rate (HR) [9]. Thus, most of the proposed indices in this thesis are nonlinear.

1.1 SCOPE OF WORK

In the article "Heart Rate Variability: A Review" [6] different linear, frequency domain, and nonlinear techniques used for the analysis of the HRV are presented and discussed. The scope of this thesis is to implement existing indices of different domains in MATLAB®. After implementation, the indices are compared to each other in different test cases to answer the following research questions:

- How sensitive are the indices regarding the length of the data and what is the best data length region for each index?

- Are there indices, which can distinguish between nonpathological and pathological subjects?

- Are any of the indices able to differentiate between younger and elderly healthy subjects?

- Do some indices show a variation before and after antiarrhythmic treatment?

- Which indices are interdependent?

1

1.2   METHODICAL APPROACH

At the beginning, a literature research was carried out. Although "Heart Rate Variability: A Review" [6] provides a great overview of all the indices and furthermore useful citations, for some of the implemented algorithms a more detailed description was essential. One has to keep in mind that implementation is one thing, but (a range of) suitable input parameters are needed too. Especially, since some of the implemented indices are not just limited to HRV data, but can be applied to different fields (e.g., the Hurst exponent or Detrended Fluctuation Analysis (DFA) to financial data).

The data is obtained from PhysioNet [34]. This source provides already annotated Electrocardiogram (ECG) recordings. Nevertheless, preprocessing of the data is needed in order to ensure that the calculation of the indices is done only on sinus beats, i.e., normal heartbeats. At the end statistical tests were applied to analyze the indices with respect to the research questions.

The flow chart in figure 1 sketches the steps needed, in order to calculate an index either in the time domain or in the frequency domain or by using nonlinear methods. Most of the nonlinear methods can be applied directly on the NN data sequence, i.e., the sequence of normal beats. For calculation of the Largest Lyapunov Exponent (LLE), resampling is needed [88].

The implementation process in this thesis starts with filtering, since PhysioNet provides annotated RR-intervals.
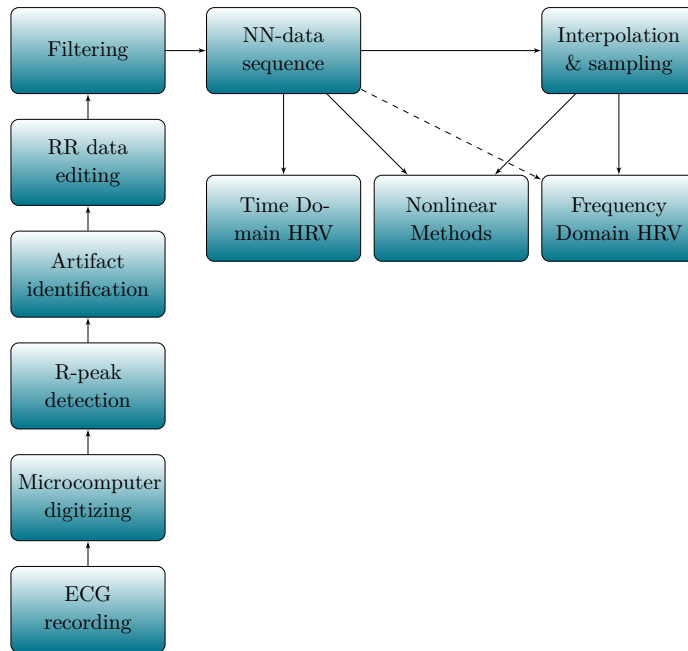
Figure 1: Steps used beginning with an ECG signal in order to calculate HRV indices in different domains (adapted from [86]). The dashed line from the NN-data sequence to the frequency domain indicates the use of the Lomb-Scargle Periodogram, which allows the calculation of frequency parameters directly from the unsampled NN-intervals (see section 5.2).

1.3 STRUCTURE OF THESIS

This thesis is organized as follows:

Chapter 2: First, the physiological background of HRV is described in chapter 2. Afterward methods of measuring the HRV are presented. The ECG is described in detail, since it defines the gold standard of measurements. This leads directly to the definition of HRV, completing chapter 2.

Chapter 3: In chapter 3, the used data is specified. Furthermore, this chapter deals with preprocessing. Hence, one filtering algorithm and a clustering method are presented.

Chapter 4: The test cases are presented in chapter 4. This chapter is completed with the applied statistical methods.

Chapter 5: Chapter 5, i. e., methods, describes and defines the implemented indices. The methods are sectioned into three parts: time domain methods, frequency domain methods and nonlinear methods. After the definition of the indices and their typical applications, chapter 5 lists the implementation steps performed for this thesis.

Chapter 6: The results of the implemented indices, with respect to the test cases are presented in chapter 6.

Chapter 7: In chapter 7, the discussion of indices with respect to the research questions and the literature can be found.

Chapter 8: Chapter 8 concludes the thesis with summarized answers to the research questions, as well as further suggestions of the use of HRV indices.

2 | BACKGROUND

> *The goal of life is to make your heartbeat match the beat of the universe, to match your nature with Nature.*
>
> *Joseph Campbell*

The clinical relevance of Heart Rate Variability (HRV) was first appreciated by Hon and Lee in 1965. They noted that fetal distress was preceded by alterations in interbeat intervals before any appreciable change occurred in the heart rate itself [43]. A substantial relationship between the Autonomic Nervous System (ANS) and cardiovascular mortality was found [86]. The variation of the heart rate reflects many physiological factors modulating the normal rhythm of the heart. HRV is a noninvasive Electrocardiogram (ECG) marker reflecting the activity of the ANS on the Sinoatrial (SA) node.

2.1 PHYSIOLOGICAL BACKGROUND

The following section is based on chapters 9-10 of [39], if not further specified.

One complete cardiac cycle consists of two stages. The active phase is called "systole" and is followed by a period of relaxation called "diastole". Each cardiac cycle represents one heartbeat.
Figure 2, a Wiggers[1] diagram, visualizes the different events occurring in the cardiac cycle for the left side of the heart. Pressure curves are given for the aortic pressure, atrial pressure and the ventricular pressure. The solid red curve describes changes in ventricular volume. The lower dark blue line illustrates the ECG recording and the bottom line shows a Phonocardiogram.
The normal rhythmical impulse of the heart is generated in the SA node, which is located in the right atrium. Hence, the SA node is commonly called "pacemaker". Via internodal pathways this impulse is conducted to the second node, the Atrioventricular (AV) node. This node delays the impulse, before passing it into the ventricles.
There are two parts of the ANS, the parasympathetic (vagus) and sympathetic part modi-

---

1 Carl J. Wiggers (May 28, 1883 – April 28, 1963)

fying and regulating the heart rate (see figure 3). Sympathetic nervous system stimulation increases the heart rate by the release of neuropeptide as well as norepinephrine [94].
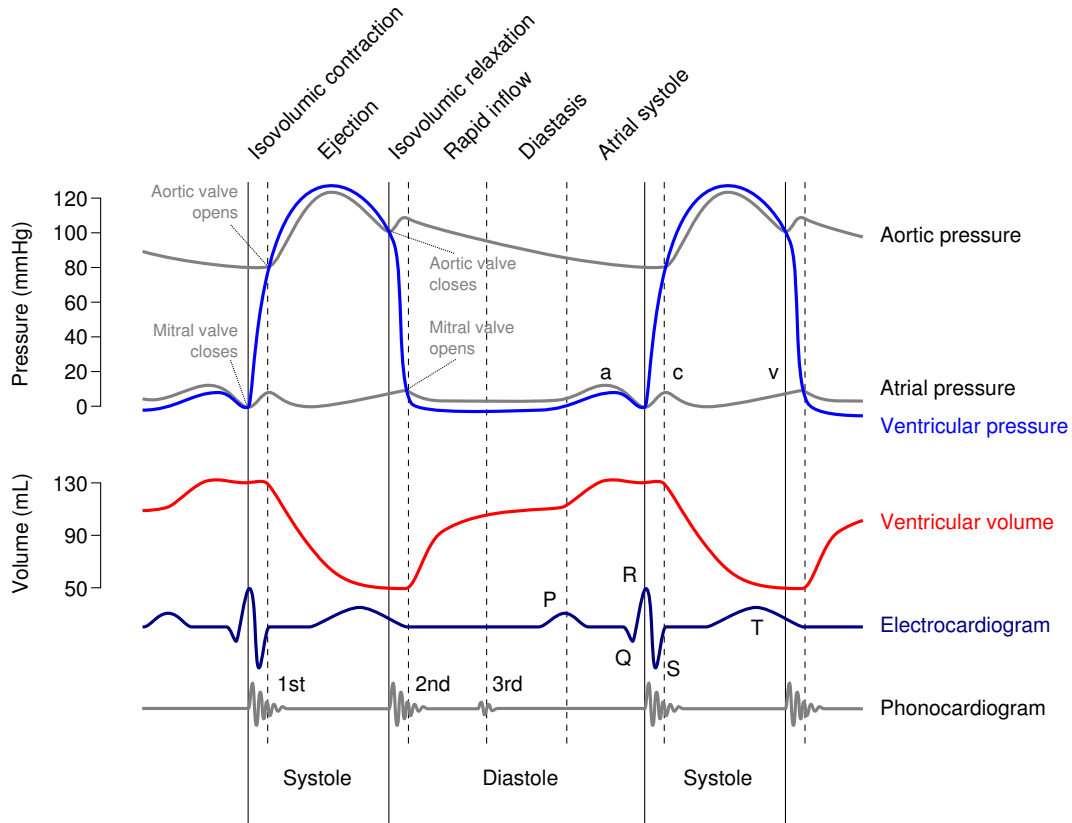
Figure 2: Events for the left side of the heart during the cardiac cycle [2].

In case of a young adult, the normal Heart Rate (HR) of 70 beats per minute (bpm) raises to 180-200 bpm [39]. In contrast, parasympathetic stimulation reduces the HR via the release of acetylcholine by inhibition of the sympathetic nervous system and by direct hyperpolarization of sinus nodal cells [67]. The beats generated form the SA node are called sinus beats. Occasionally beats are generated from other sources. These are termed ectopic beats. Healthy subjects exhibit a small number of ectopic beats, though. Up to 10 ectopic beats per hour were reported in a study of healthy men, aged 40-66 years [68].

The regulation system of the HR is one of the most complex systems in humans [91]. In [22], Costa et al. proposed that pathologic dynamics increase the regulatory and thus lead to less variability. As a rough guide, one can say:

"Healthy subjects have a higher HRV compared to unhealthy ones."

Figure 3: Sympathetic (in red) and parasympathetic nervous system (in blue) control the heart rhythm (adapted from [1]).

2.2 HRV MEASUREMENT AND THE ECG

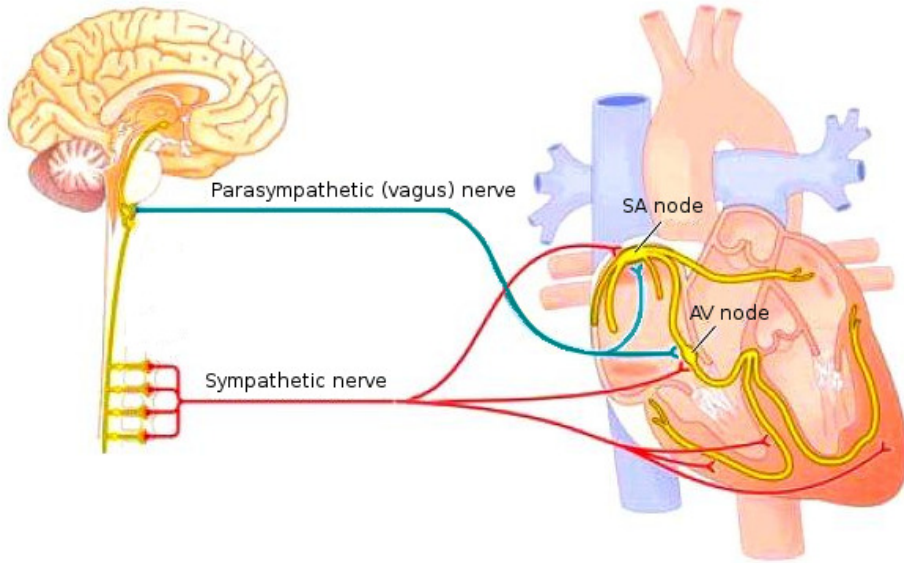There are different methods measuring the HR. These are based on two principles: The first one is based on the pumping activity of the heart, leading to blood flow and generation of a pulse wave. The second one, and also gold standard for HRV measurements, involves the use of an ECG. By placing surface electrodes on the body skin either at the thorax or at the extremities electrical potentials are recorded. If the heart muscle depolarizes it causes tiny electrical changes and the resulting potential differences are measured. In order to measure these differences two measuring points are needed. For a conventional 12-lead[2] ECG ten electrodes are used and the whole cardiac cycle is recorded.

Figure 4 illustrates an ECG signal with the P, Q, R, S and T waves. Several algorithms for detection already exists [11, 12].

The P wave is caused by spreading of the depolarization through the atria and is followed by the QRS complex. This sharp complex appears as a result of depolarization of the ventricles. Since the ventricles have a larger muscle mass compared to the atria, the QRS complex usually has a larger amplitude than the P wave. Finally the ventricular T wave represents the repolarization of the ventricles.

---

2 A lead represents the electrical potential difference between two measuring points.
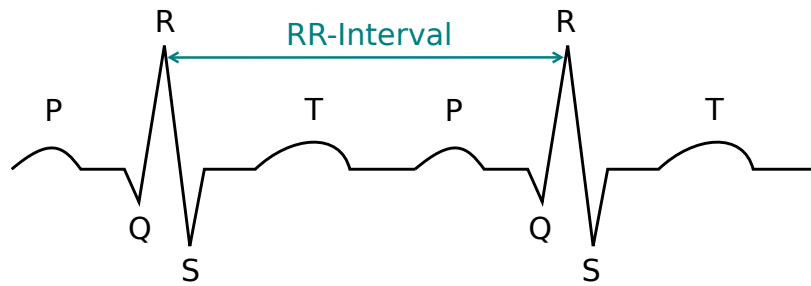
Figure 4: Illustration of an ECG-signal. The beat-to-beat interval (i.e., RR-interval) is defined between two consecutive R-peaks.

2.3 DEFINITION OF HRV

The term HRV refers to the variation of beat-to-beat intervals. Usually, HRV is studied in an ECG by considering the time duration between two R-peaks. Therefore, the beat-to-beat intervals are often named RR-intervals. As the x-axis represents time, the unit of RR-intervals is given in seconds (s) or milliseconds (ms). A schematic representation of an ECG-signal can be seen in figure 4.

3 | DATA

All data used for the tests have been taken from PhysioNet [34], a free-access, online archive of physiological signals. PhysioNet guarantees that all data have been fully deidentified (anonymized), and may be used without further institutional review board approval.

In the following three sections the used databases are described in detail. This chapter concludes with section 3.4, treating preprocessing.

3.1 NONPATHOLOGICAL DATA

The nonpathological dataset is a composition of the three PhysioNet databases: `fantasia`, `nsrdb` and `nsr2db`.

- *Fantasia Database* [47]:
  This database consists of forty 120 minutes long recordings of healthy subjects. It is divided in two cohorts: twenty young and twenty elderly subjects. Each cohort includes equal numbers of men and women and the Electrocardiogram (ECG) is digitized at 250 Hz.

- *Massachusetts Institute of Technology (MIT)-Boston's Beth Israel Hospital (BIH) Normal Sinus Rhythm Database* [34]:
  The `nsrdb` includes 18 long-term ECG recordings of subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital. It includes 5 men, aged 26 to 45, and 13 women, aged 20 to 50. Data is sampled at 128 Hz.

- *Normal Sinus Rhythm RR-Interval Database* [34]:
  Furthermore, the `nsr2db`, a composition of 54 approximately 24 hours long ECG recordings, is included. The 30 healthy men are 28.5 to 76 years old and the 24 women range from 58-73 years. Sample frequency of the signal is 128 Hz.

A histogram of the age distribution of the entire nonpathological data is plotted in figure 5. Table 1 shows the number of young and elderly subjects and their range for each database.
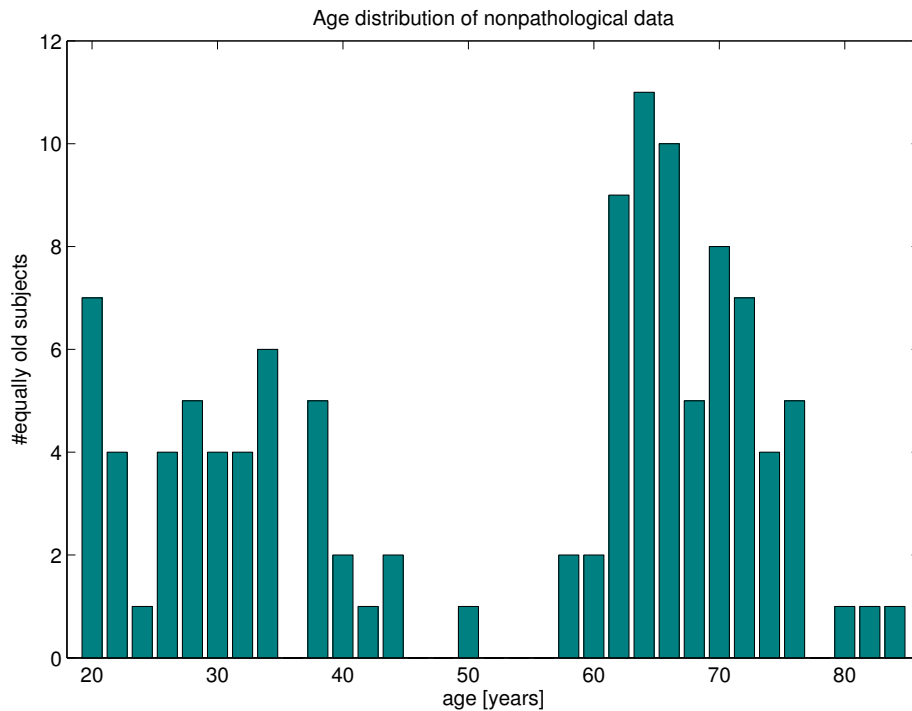
Figure 5: Histogram of the age distribution of the total 112 healthy subjects. Their mean age is 53 years.

	young	old
	no. subjects / range[years]	no. subjects / range[years]
`fantasia`	20 / 21-34	20 / 68-85
`nsrdb`	18 / 20-50	- / -
`nsr2db`	8 / 28.5-40	46 / 58-76
total	**46 / 20-50**	**66 / 58-85**

Table 1: Age distribution of nonpathological data.

3.2 PATHOLOGICAL DATA

The pathological dataset is also a collection of three PhysioNet databases: `chf2db`, `mitdb` and `svdb`.

- *Congestive Heart Failure RR-Interval Database* [13]:
  This database consists of 29 long-term (each approximately 24 hours long) recordings

of subjects with congestive heart failure. The age of the subjects ranges from 34-79 years. Sample frequency of the signals is 128 Hz. Furthermore, the New York Heart Association (NYHA) classes (I., II. and III.) for this database are denoted [10].

- *MIT-BIH Arrhythmia Database* [65]:
  The `mitdb` is a junction of 48 half-hour recordings from 47 subjects (25 men aged 32 to 89 years and 22 women aged 23 to 89 years. It includes 25 examples of uncommon but clinically important arrhythmias and the remaining recordings are randomly chosen. The digitization rate is 360 Hz.

- *MIT-BIH Supraventricular Arrhythmia Database* [38]:
  The `svdb` includes 78 half-hour ECG recordings of subjects with supraventricular arrhythmia, sampled at 128 Hz. Further descriptions, i. e., age or gender are not available.

Table 2 lists the number of recordings and their recording lengths (number of RR-intervals) of the pathological database, in comparison to the nonpathological database, since the nonpathological database defines the control group. The shortest recording of both databases consists of 1431 RR-intervals.

	nonpath			path		
db name	fantasia	nsrdb	nsr2db	chf2db	mitdb	svdb
records (female)	40 (20)	18 (13)	54 (24)	29 (2^1)	48 (22)	78 ($-^2$)
$N_{\min}$	4924	7983	77373	90850	1512	1431

Table 2: PhysioNet databases. $N_{\min}$ is the minimal number of RR-intervals per database.

From `mitdb`, the last 4 recordings, i. e., 231-234, were excluded, since after clustering, less than 1100 RR-intervals were available. This threshold defines the minimal required length in order to run Test Case I (see section 6.1). The same reason holds for record 860 from `svdb`. Table 3 clarifies and summarizes the used data of the nonpathological and the pathological databases.

	nonpath	path
records (female)	112 (57)	150 (21^3)

Table 3: Used databases after exclusion.

---

1 gender only known for 10 recordings: 8 men and 2 women
2 gender not specified
3 gender is provided just for 57 recordings

3.3 CRIS DATA

The Cardiac Arrhythmia Suppression Trial (CAST) RR Interval Sub-Study Database [85] was initiated to test the hypothesis that suppression of ventricular arrhythmias by antiarrhythmic drugs after myocardial infarction improves survival [25]. Subjects of the CAST RR Interval Sub-Study Database were randomly assigned to receive different antiarrhythmic drugs. The database consists of 1543 24-hour RR-interval records from 809 subjects.
For 734 subjects the following criteria are fulfilled:

- had usable pre-treatment and on-therapy recordings,

- received a randomly-assigned antiarrhythmic treatment that successfully reduced their PVC rates by at least 80% on the first attempt,

- continued on that treatment.

In order to avoid daytime-dependent variations, the window for all the data samples was defined around 6 p.m. Three more recordings have been excluded, since there was no data in the chosen time window. The database is divided in two parts, i. e., crisa and crisb, where crisa contains the baseline data and crisb contains the on-therapy recordings.

	crisa	crisb
records (female)	731 (132)	731 (132)

Table 4: CRIS database after exclusion.

3.4 PREPROCESSING

PhysioNet provides annotated beats as **.txt*-files. In order to obtain RR-intervals, the time differences between normal beats have to be computed. A full list of the PhysioNet annotations[1] is available.
If a sequence of not-normal beats interjacent two normal beats, their time difference is immoderate long. Hence, the first step of preprocessing is to remove all intervals longer than 2.5 seconds. In [14] deletion of beats longer than 2 seconds is reported.
One clustering method and one filtering method are presented in the following sections. The clustering method was applied to data, in order to perform the calculations only on the sinus beat cluster.

---

1 http://physionet.org/physiobank/annotations.shtml

3.4.1 Clustering

Clustering (or cluster analysis) is the technique for developing meaningful subgroups of objects [40]. To be more precise, to classify objects into smaller groups based on the similarities among the entities [40].

In this thesis a density based clustering algorithm is chosen, in order to group data points of different heart rhythms. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN)-algorithm, shows - as the name suggests - a high robustness against noise [29]. Furthermore, DBSCAN does not require an a-priori number of clusters. The two input parameters needed to define a cluster are the neighborhood threshold ε and the minimum number of points ($MinPts$).

Ensemble Density-Based Spatial Clustering of Applications with Noise (EDBSCAN) is an improvement of the DBSCAN algorithm, since the choice of the neighborhood ε is one of the difficulties of the DBSCAN procedure. EDBSCAN varies the threshold parameter for the neighborhood ε, r-times equidistantly from $\varepsilon_{\min}$ to $\varepsilon_{\max}$.

The ε ranges are picked automatically by the algorithm depending on the input data. A detailed description of the clustering is published in [93]. The remaining input parameters are the minimum number of points in one cluster ($MinPts$) and the number of iterations (r), to obtain ε. For all data filtered with the EDBSCAN $r = 20$ and $MinPts = 4$ as in [44] are chosen.

Figure 6 shows a clustered Poincaré plot[2] of a subject with arrhythmia. Since only the sinus beat cluster is considered for indices calculations, the number of RR-intervals is reduced.

---

2 The definition of the Poincaré plot is provided in the appendix (A.2)
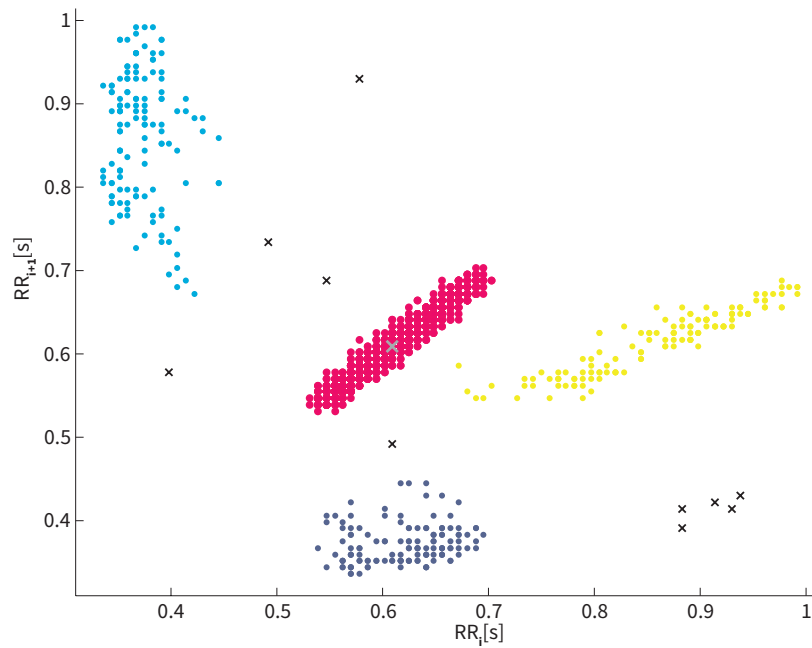
Figure 6: Clustered poincaré plot of a subject with arrhythmia. Noise (points in no cluster) is marked as black crosses. The red cluster defines the sinus beat cluster, where HRV Parameters are calculated. Other clusters are also refused.

One detriment of filtering by clusters is to get rid of data points (obviously recordings with the same data length before filtering do not have to be equally long after filtering). Furthermore, the clustering procedure is computationally expensive.

3.4.2   Impulse Rejection Filter

The Impulse Rejection Filter (IRF) is a median filter and was introduced by McNames et al. [60]. Median filters replace the incorrect, or detected as incorrect, RR-interval with the median of a specified window. IRF detects erroneous beats on its own. In order to identify the ectopic beats, a test statistic is calculated (see equation (1)).
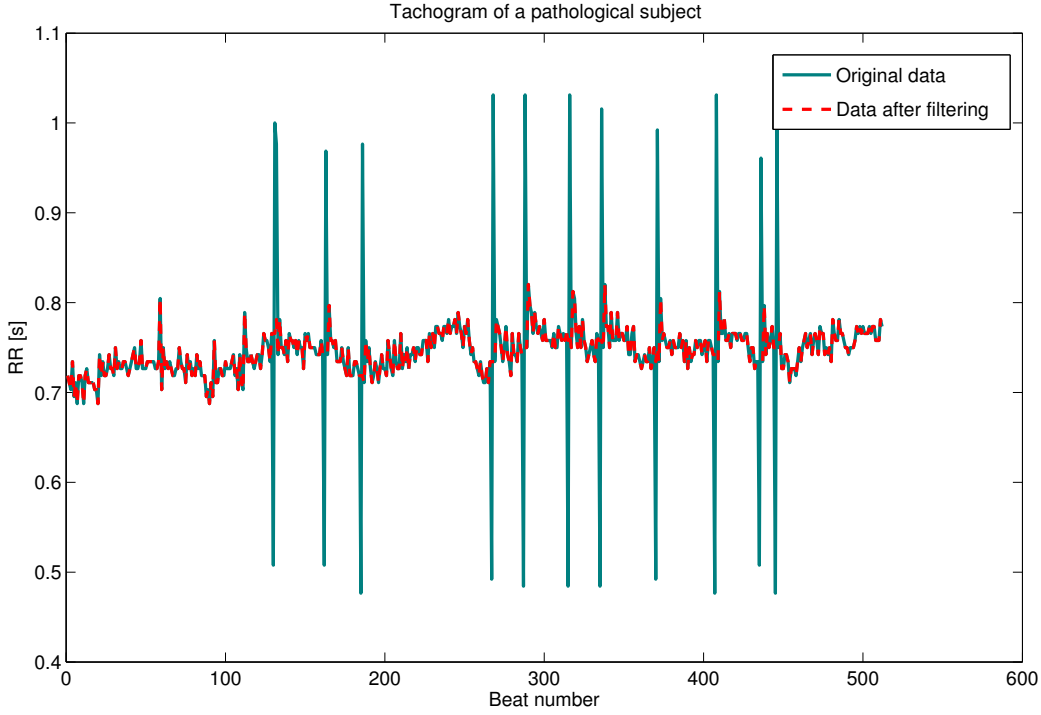
Figure 7: Tachogram of a subject with CHF. The teal[1] line represents the original data before filtering and the filtered data is colored red. Ectopic beats are replaced by median values of specified windows.

Figure 7 shows an approximately six minutes long recording of a subject with CHF and some sharp impulses (ectopic beats). The IRF replaces these, as sharp detected peaks, by the median value of the neighborhood (figure 7b). To be more precise, for every beat number a test statistic is calculated:

$$D(j) := \frac{|x(j) - X_M|}{1.483 \cdot \text{median}(|X - X_M|)} \quad \text{with} \quad j = 1, \dots, N \quad , \tag{1}$$

where X_M defines the median value of the entire recording, i. e.,

$$X_M := \text{median}(X) = \text{median}\left(\left\{x_1, \dots, x_N\right\}\right). \tag{2}$$

The filtered signal $\tilde{x}(i)$ is defined as:

$$\tilde{x}(j) = \begin{cases} x(j) & D(j) < \tau \\ r(j) & D(j) \geq \tau. \end{cases} \tag{3}$$

---

1 Teal is a deep blue-green color; a dark cyan. Teal gets its name from the colored area around the eyes of the common teal, a member of the duck family - Wikipedia, Sept. 2015.

The replaced beat value $r(j)$ depends on a user specified window length w_m:

$$r(j) := \text{median}\left(\left\{x(j+m) : |m| \leq \frac{w_m - 1}{2}\right\}\right). \tag{4}$$

For computation, $w_m = 5$ and $\tau = 4$ are chosen (same as in [60]).

4 | TEST CASES AND STATISTICS

At the beginning of this chapter, the different test cases are described. In order to test for the significance of index differences between different test groups, we us statistical methods that are described at the end of this chapter in section 4.5.

4.1 TEST I: DATA LENGTH SENSITIVITY

The first test is to figure out how sensitive the implemented indices react to different recording lengths. At the beginning, recordings of 1100 RR-intervals of the nonpathological dataset are tested against pathological recordings of the same length. Then the data length is reduced stepwise by 50 RR-intervals from the end. The initial data length of 1100 intervals was chosen since the shortest unfiltered signal consists of 1431 data points and after clustering 1100 points were available for each recording. Furthermore, 1100 RR-intervals is the upper length limit for short-term recordings [15] (20 minutes recording with an average heart rate of 60 beats per minute (bpm) results in a data length of 1200 intervals).
The starting time of each record is not specified for most of the recordings, so a sequence of 1431 points is taken from the middle of each recording before clustering.

4.2 TEST II: NONPATHOLOGICAL VS. PATHOLOGICAL SUBJECTS

To test if the indices can differentiate between healthy and unhealthy subjects, indices of the pathological database are tested against the calculated indices of the nonpathological database. After running the first test, a data length of 1024 RR-intervals seems to be a reasonable and feasible choice. A length of the power of two, i.e., 1024, was chosen in order to ensure that each index can exploit the same data length for computation. For instance, the implemented algorithm of the Hurst exponent H crops the input data at the maximum feasible value of power of two (see section 5.3.5).
The first 1024 points of the clustered data from Test Case I were used.

4.3   TEST III: YOUNGER VS. OLDER HEALTHY SUBJECTS

For testing the indices on their age-dependency, the nonpathological database was split into two groups. The first group contains subjects of age 53 and less, subjects in the second group are older than 53 years. This threshold was chosen, since it is the mean age of the nonpathological database. For a more detailed description of the two groups and a histogram, see section 3.1. The data length of 1024 points is the same as in the second test case. Furthermore, the same clustering and statistical tests were applied.

4.4   TEST IV: PRE- VS. POST-ANTIARRHYTHMIC TREATMENT

To test if indices can discriminate between data before and after arrhythmia suppressing medication, the CRIS-database was used. Again, the data length is fixed by 1024 RR-intervals. To obtain the designed data length, the first 2000 data points around 6 p.m. from the 24-hour recording were filtered and afterward the sequence was cropped. For each of the 731 subjects the indices before and after treatment were tested for significant differences.

4.5   STATISTICS

The Lilliefors[1] test was used to detect if the results are normally distributed. If they follow a normal distribution, the t-test was used to determine the differences between the data sets. Otherwise, the Wilcoxon rank-sum test was used to obtain the p-value. In addition, a visual inspection of the distribution was made in Q-Q plots. Figure 8 shows the Q-Q plot for the index *pNN50* of the nonpathological data set.

The Wilcoxon rank-sum test is a nonparametric test of the null hypothesis that data x and y are samples from continuous distributions with equal medians, against the alternative that they are not. The test assumes that the two samples are independent but there is no assumption of normality. Furthermore, the two data samples can have different size.

In order to compare all the indices in Test Case I, for varying data lengths, only the Wilcoxon rank sum test was used in this case.

The assumption of independency is not given for test case IV, since the compared data samples (pre- and post-treatment) are of the same subject, and so dependent. In this case and furthermore, if the data was not normal distributed, the Wilcoxon signed-rank test was applied.

A test with a p-value $p < 0.05$ is called significant. If the p-value is smaller than 0.01, the test outcome was declared as very significant. To test if indices correlate with each

---

1 Hubert W. Lilliefors (1928 – February 23, 2008)
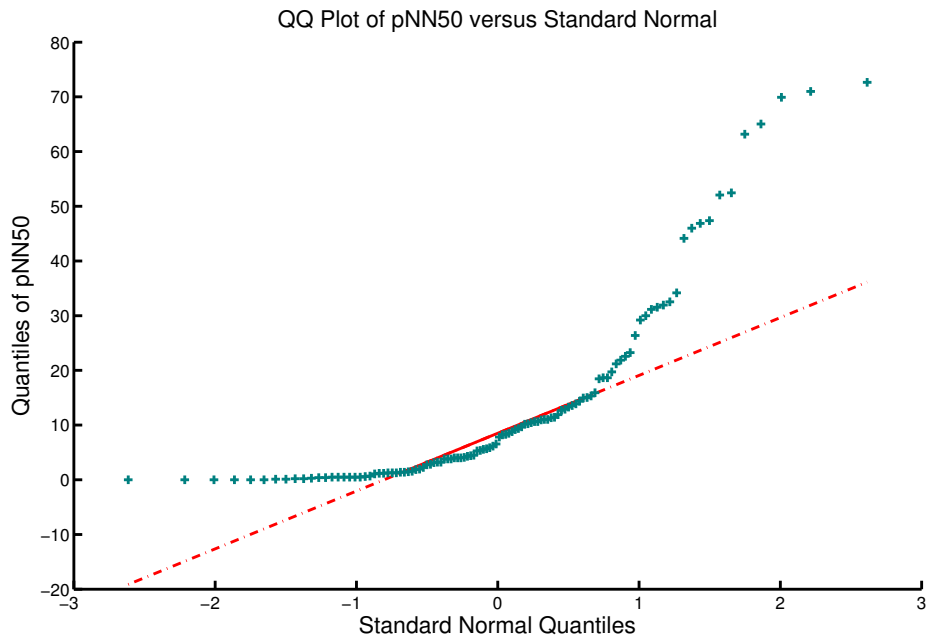
Figure 8: Q-Q plot for the index pNN50 of healthy subjects. The dot-dashed red line represents the normal distribution. Especially at the edges the sample of pNN50 is not normal distributed.

other, Spearman's[2] rank correlation coefficient r was used. According to [66], Spearman's correlation coefficient is more robust against outliers than Pearson's correlations coefficient and should be used if the variables are not normally distributed. It assesses the strength of association between two variables, by using a monotonic function. If one variable is a perfect monotone function of the other $r \in \{-1, 1\}$ holds. The sign of the Spearman's rank correlation coefficient specifies the direction of dependence between the two variables. Strong correlation was defined as $|r| > 0.85$.

All tests were done with MATLAB®built-in functions.

---

2 Charles E. Spearman (September 10, 1863 – September, 17 1945)

5 | METHODS

For all the algorithms in this chapter we consider the set of RR-intervals $\left\{RR_1, \ldots, RR_N\right\}$ of length N and denote it as:

$$X := \left\{x_1, \ldots, x_N\right\} = \left\{x(1), \ldots, x(N)\right\} = \left\{RR_1, \ldots, RR_N\right\}. \tag{5}$$

In order to obtain useful and comparable results the data has to be preprocessed, before calculating the indices. Detailed description of used filter and clustering algorithms is provided in section 3.4.

The methods can be divided into three main sections, where most of the implemented algorithms belong to the Nonlinear Methods.

5.1 TIME DOMAIN ANALYSIS

Due to completeness and also comparison reasons, the most important statistical methods are briefly described. Since geometric methods are relatively insensitive to the analytical quality of the set of RR-intervals [55], we will also consider two of them.

5.1.1 Statistical Methods

There are two classes of parameters, those obtained directly from the RR-series and those considering the differences of the time series. In literature RR-intervals are sometimes denoted as NN-intervals, in order to emphasize that this is the duration between "normal" beats. The Standard Deviation of the NN intervals (SDNN) is defined as:

$$SDNN := \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \text{mean}\,(X))^2}. \tag{6}$$

The total variance of Heart Rate Variability (HRV) and therefore the SDNN increases with the length of the recording. Hence, same data lengths have to be considered in order to compare SDNN measures [86]. For the following measures, we consider the differences in the RR-series. Precisely, the Standard Deviation of the Successive Differences (SDSD) is calculated as:

$$SDSD := \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} \left(|x_{i+1} - x_i| - \text{mean}\,(|\Delta X|)\right)}, \tag{7}$$

where ΔX is the set of successive differences of X of length $N - 1$, i. e.,

$$\Delta X := \left\{ x_2 - x_1, \ldots, x_N - x_{N-1} \right\}. \tag{8}$$

The Number of interval differences of successive NN intervals greater than 50 ms (NN50) and the percentage of interval differences of successive NN intervals greater than 50 ms (pNN50) have the following definitions:

$$NN50 := \#(|\Delta X| > 50 \text{ ms}), \tag{9}$$

$$pNN50 := \frac{NN50}{N} \cdot 100. \tag{10}$$

5.1.2   Geometrical methods

In this section, two measures, i. e., the HRV index (HRV_{Idx}) and the Triangular Interpolation of NN interval histogram (TINN) are presented. The calculation of the index TINN follows [86].

Figure 9 illustrates the calculation of the TINN of a healthy subject. On the x-axis the duration of RR-intervals and on the y-axis the number of each equally long RR-interval are plotted, respectively. $D(j)$ is the approximated discrete density distribution function. The function D is constructed as follows:

- "bin" the x-values, i. e., divide entire range of RR-intervals into sequences of intervals.

- $D(j)$ assigns the number of equally long NN-intervals at each bin.

If the sample frequency is given, its inverse should be used as the bin width. D_{max} is the count of the modal bin and is used in the formula of the HRV_{Idx}. Precisely, the geometric indices are defined as:

$$HRV_{Idx} := \frac{N}{D_{\max}}, \tag{11}$$

and the index TINN is defined as the baseline width of the fitted triangular:

$$TINN := B - A. \tag{12}$$

In order to obtain the two points A and B on the x-axis, a triangle with the upper edge lying in $D_{\max}$ is fitted to the distribution function, in a least-squares sense. The edges of the base of the triangle are calculated by means of a piecewise continuous linear function $q(t)$ defined as:

$$q(t) := \begin{cases} 0, & t \leq A \\ q_1, & A \leq t \leq C \\ q_2, & C \leq t \leq B \\ 0, & t \geq B. \end{cases} \tag{13}$$
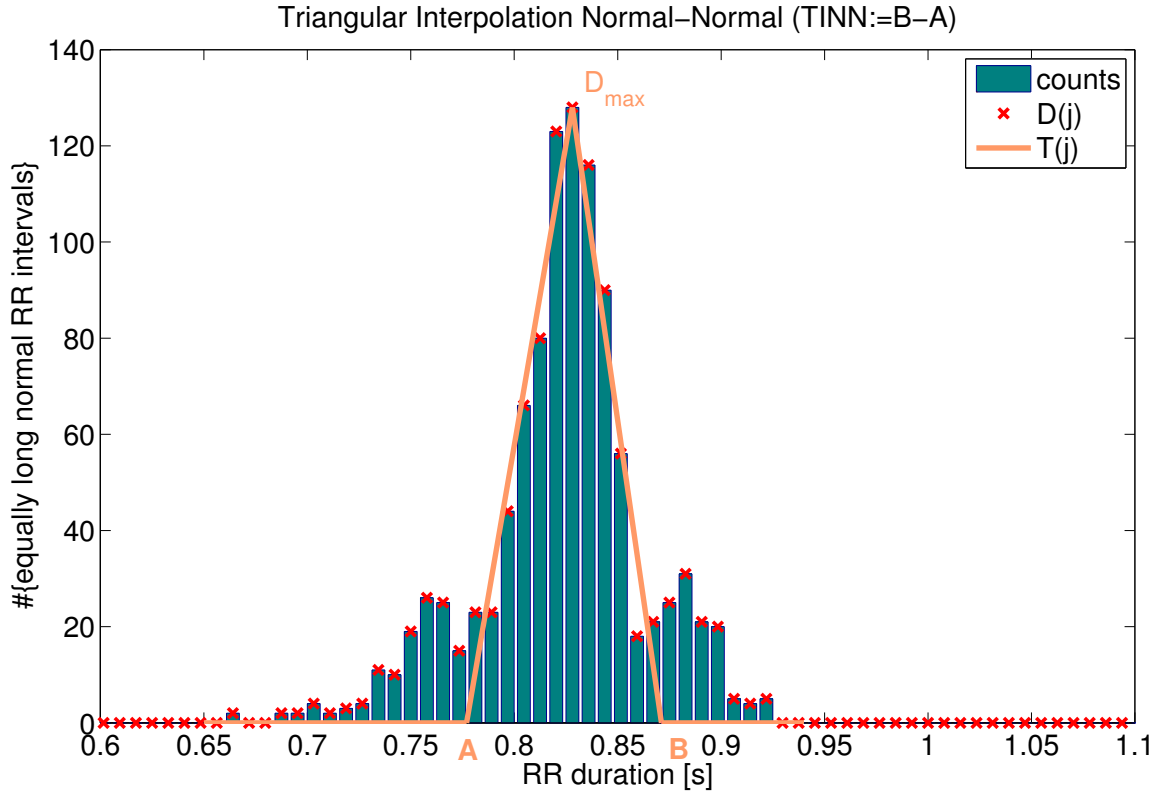
Figure 9: Calculation of the TINN of a nonpathological subject. TINN is defined as the baseline width: B-A.

If the integral over all sections of the squared difference reaches the minimum, i.e.,

$$\int_0^\infty \left(D(t) - q(t)\right)^2 \mathrm{d}t \to 0, \tag{14}$$

the values for A and B are fixed.

It should be noted, that the number of bins strongly influences the HRV_{Idx}, since if the bin width is larger, the denominator is larger and thus HRV_{Idx} smaller. The sample frequencies in the composite nonpathological database are different and the same is true for the pathological database. In order to make the results comparable, a fixed sample frequency of 128 Hz was chosen for the calculation of both geometrical indices. According to [86], 128 Hz is also the standard scale for the geometrical measures, otherwise the size of bins should be quoted.

5.2   FREQUENCY DOMAIN ANALYSIS

This section covers the second category of HRV analysis, namely the frequency domain. First, common HRV frequency domain measures, which are not implemented in this thesis, are described. Subsequently, the Continuous Wavelet Transform with the implemented index decay scaling exponent $\tau(q)$ is presented.

5.2.1   Common HRV Frequency Domain Measures

The simple calculated time methods can not distinguish between sympathetic and parasympathetic contributions of HRV [6]. Hence, the Power Spectral Density (PSD) analysis has been employed. PSD is the power distribution as a function of frequency. The frequency band is sectioned into four bands, namely ultra low frequency (ULF, <0.003 Hz), very low frequency (VLF, 0.003-0.04 Hz), low frequency (LF, 0.04-0.15 Hz) and high frequency (HF, 0.15-0.4 Hz). The last three spectral components are distinguished in short-term recordings [86].

One commonly used technique to estimate the PSD is the Fourier transform, i. e., Fast Fourier Transform (FFT). FFT requires an evenly sampled time series, hence, the first step is resampling (see figure 1). This is generally done by interpolation (e. g., linear or cubic splines) of the RR-series and a common samplerate is 4 Hz (see figure 10).

In addition to classical FFT based methods, used for the calculation of frequency parameters in HRV, the Lomb-Scargle periodogram allows to estimate the PSD directly from the unevenly sampled RR-series [20]. The direct path from unevenly sampled RR-series to the calculation of frequency parameters is also sketched in figure 1.

The recommended standard frequency measures (according to "Guidelines", [86]) for short-term recordings are the powers in the three main ranges: VLF, LF and HF. Their unit is given in ms^2 or the powers are normalized. The formula of the normalized power in LF is given by:

$$\text{LF}_{\text{norm}} = \text{LF}/(\text{Total Power-VLF}) \cdot 100. \tag{15}$$

For the normalized HF power, the acronym "LF" is just replaced by "HF" in (15).

Moreover, the ratio LF/HF is computed as a standard measure. In figure 11, the PSD of a healthy subject is plotted. Furthermore, the standard frequency measures of this recording are embedded, boxed beside the legend.

5.2.2   Continuous Time Wavelet Transform Analysis

The Continuous Wavelet Transform (CWT) possesses the ability to construct a time-frequency representation of a signal. Hence, time-frequency analysis allows to study a transformed, two-dimensional signal in contrast to the time domain and the ordinary frequency domain

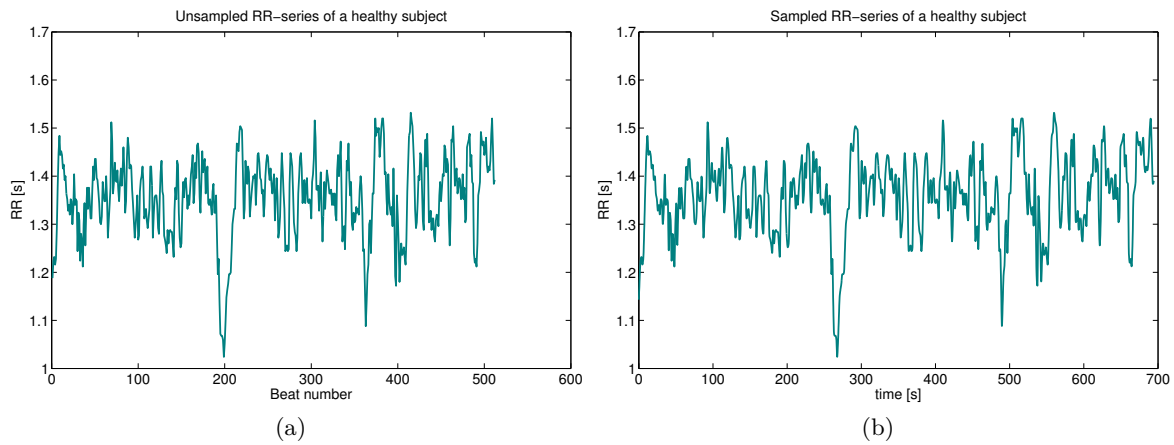Figure 10: Tachogram of a healthy subject. Note the different x-axis. (a) Unsampled RR-intervals. (b) Interpolated and resampled RR-series in order to calculate frequency parameters. Interpolation was done with cubic splines.
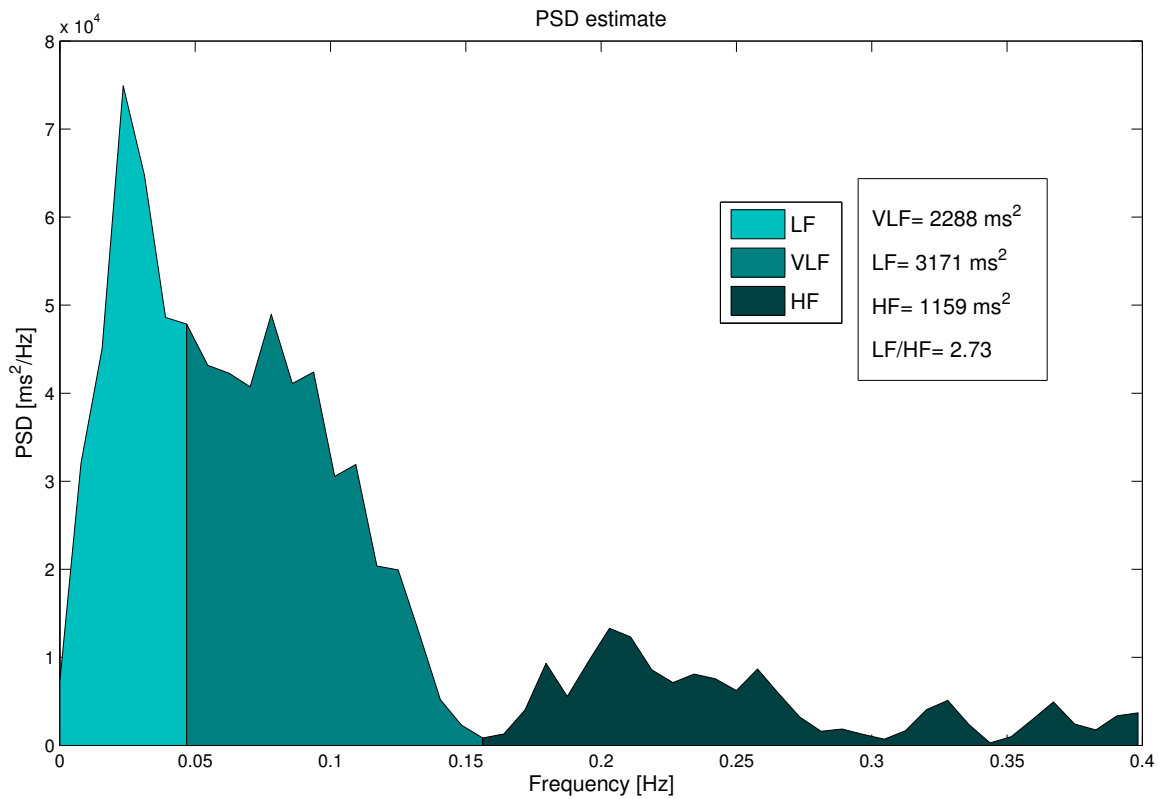

Figure 11: Power Spectral Density of a health subject. The three bands (VLF, LF and HF) are marked with different colors. The standard measurements for short term recordings are boxed beside the legend.

methods.

A wavelet is a mathematical function used to divide a given function or continuous-time signal $f(t)$ into different scale components. Each scale component can then be studied with a resolution that matches its scale. A wavelet transform is the representation of the signal by wavelets. For the CWT arbitrary scale and dilation parameters can vary continuously. In [80], the CWT has been used to evaluate the effect of local anesthesia on HRV parameters. Furthermore, different scalograms, i. e., the visualization of the CWT coefficients, and their visual pattern are described in [7].

The CWT for a function f is defined as:

$$W(a,b) := \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)\, \mathrm{d}t, \tag{16}$$

where the set of wavelets is defined by

$$\psi_{a,b}(t) := \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \tag{17}$$

i. e., by scaling and shifting the so-called "mother wavelet" ψ.

The scalogram is the absolute value squared of the output of the wavelet transform and shows the scale parameter a on the y-axis over the translation parameter b. The intensity of colors in this two-dimensional plane indicates the wavelet energy density function (see figure 12a). In [7], Acharya et al. described that the scalogram for normal cases should be "flowery" and regular, whereas ectopic signals are indicated as a sudden surge of radial lines in the scalogram.

In order to automatically capture and compare the characteristics of the scalogram, the Wavelet Transform Modulus Maxima (WTMM) method was applied in this thesis. According to [35] this method consist of the following steps:

1. Calculation of the CWT according to equation (16).

2. Obtain the local maxima of the modulus of the CWT: The largest wavelet transform coefficients are found at each scale a by:

$$\frac{\partial W(a,b)}{\partial b} = 0. \tag{18}$$

3. Generate a partition function $Z(q,a)$:

$$Z(q,a) = \sum_{a,b} \sup_{a} |W(a,b)|^q. \tag{19}$$

4. Calculation of the decay scaling exponent $\tilde{\tau}(q)$ using linear regression:

$$\log Z(q,a) \approx \tilde{\tau}(q)\log a + C(q). \tag{20}$$

5. Obtain the index τ by a linear fit of the plot $\tilde{\tau}(q)$ over q.

The calculation was done by means of the MATLAB® Wavelet Toolbox and the WaveLab [17] library was used as well. According to [5] the Morlet wavelet is chosen as "mother wavelet". Figure 12b shows a plot of a the Morlet wavelet with support $[-4, 4]$.
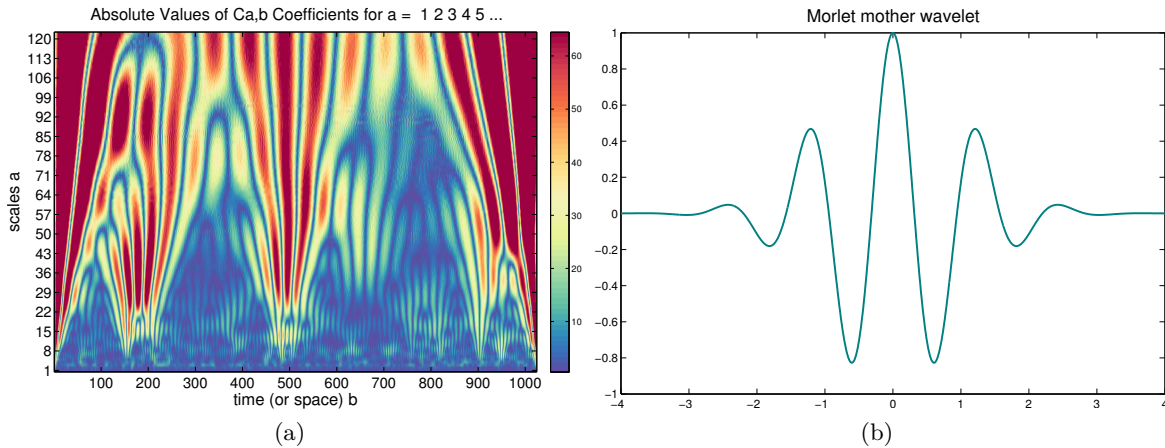
Figure 12: Morlet mother wavelet and scalogram of CWT analysis: (a) Scalogram of CWT coefficients with scales a=1, ..., 122. (b) Morlet mother wavelet with support $[-4, 4]$.

5.3 NONLINEAR METHODS

Already in the 1980s, it was indicated that beat generation in a human heart is a chaotic process [63]. As the RR-time series is not constant over time, and irregular, it can not be completely explained by a linear approach [53]. Thus, recently there is an increasing interest in nonlinear methods and methods from chaos theory [46].

The following nonlinear measures can be divided in two categories, first chaos theory descriptors and second fractal descriptors. Another set of nonlinear descriptors are information descriptors, such as Approximate Entropy (ApEn) and Sample Entropy (SampEn). The application of those parameters in HRV data has been investigated in [59].

5.3.1 Correlation Dimension Analysis

The first nonlinear measure, the Correlation Dimension (CD) assesses the fractal dimension of the system attractor in the reconstructed phase space, i.e., measures the complexity of the chaotic system [54].

CD has been applied to healthy and unhealthy subjects, including different diseases of the myocardium [18]. In 1998, Fojt and Holcik observed a significant decrease of CD in patients

with bigeminy [16].

The CD index is calculated based on an algorithm proposed by Grassberger and Procaccia [37]. In the first step, a phase space reconstruction is created:

$$Y_i = \Big\{ x(i), x(i+\tau), \ldots, x(i+(m-1)\tau) \Big\}, \quad i = 1, \ldots, N - (m-1)\tau. \tag{21}$$

Hereby, m is the embedding dimension and τ is the delay time. So Y is a $M \times m$ matrix with $M := N - (m-1)\tau$ as a composition of row vectors Y_i.

The measure CD is obtained by considering correlations between points on this attractor. Therefore, the correlation integral is defined as:

$$C(r) := \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \Theta \left(r - \| Y_i - Y_j \| \right), \tag{22}$$

where r is the radius and Θ denotes the Heaviside function:

$$\Theta(x) := \begin{cases} 0, & x \leq 0 \\ 1, & x > 0. \end{cases} \tag{23}$$

$\| Y_i - Y_j \|$ is the Euclidean distance between a pair of points within the attractor. Hence, the correlation function counts the number of distances closer than r, over the total number of distances. Obviously $C(r)$ has values between 0 and 1. If M is sufficiently large and r is small enough, the following definition holds:

$$CD := \lim_{r \to 0} \frac{\log C(r)}{\log r}. \tag{24}$$

To obtain an optimal value for the delay time τ the Autocorrelation Function (ACF) method, i. e., finding the place where ACF is below a certain predefined threshold close to zero, is commonly used [8]. Another design criteria to determine the delay time is the Mutual Information (MI) method [31]. In this work, the delay time and the embedding dimension are chosen as $\tau = 1$ and $m = 22$, according to [62]. The last step is the design of the scaling region for the power law behavior. The approach of [62] is the following: Obtain r_{max} at $C(r_{max})) = 0.1$ and $r_{min} = r_0 + 0.25(r_{max} - r_0)$, where r_0 is the smallest radius. Since most of the subjects had nearly the same scaling region, parameters for r_{max} and r_{min} conversely are fixed as $r_{max} = 0.15$ and $r_{min} = 0.06$. In figure 13, the plot of the scaling region and the linear fit for a healthy subject is shown.
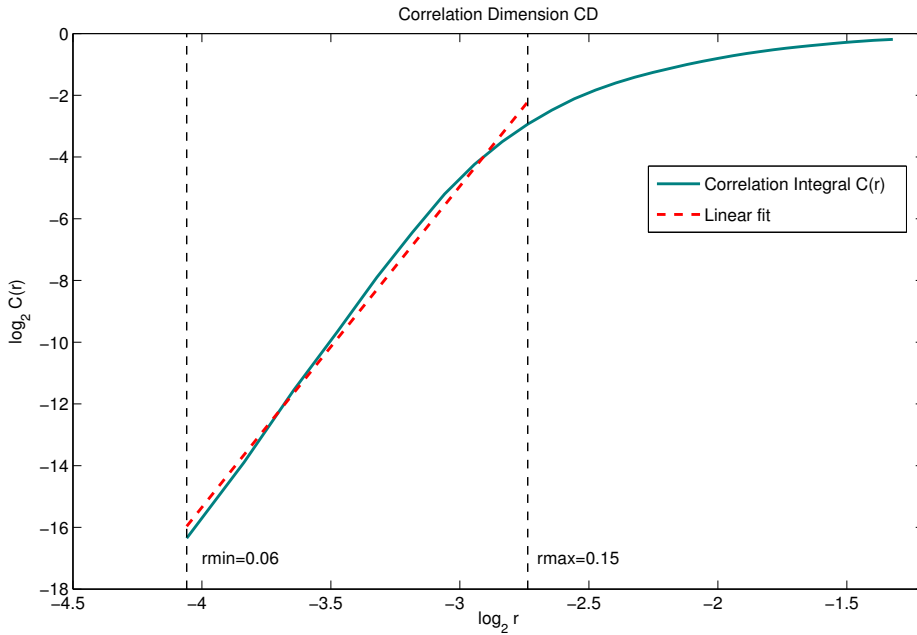
Figure 13: CD of a nonpathological subject. The fitting region lies between the black dashed lines.

5.3.2 Largest Lyapunov exponent

Chaotic systems are sensitive to initial conditions. Small changes in the state variables at one point evolve large differences in the behavior of the system at some future point [88]. The Lyapunov exponent assesses this dependency on initial conditions, by defining an average rate of divergence of two neighboring trajectories [6].

Wolf proposed the first algorithm that allows the estimation of nonnegative Lyapunov exponents from an experimental time series in 1985 [92]. Since the shortest dataset used in this thesis just contains 1431 samples (unfiltered) the algorithm of Rosenstein, which allows the calculation of the Largest Lyapunov Exponent (LLE) for small data sets is chosen [79].

The first step of this approach is again the phase space reconstruction (see equation (21)). The reconstructed phase space vectors Y_i forms a $M \times m$ matrix:

$$Y = (Y_1, Y_2, \ldots, Y_M)^{\mathrm{T}}. \tag{25}$$

By a given embedding dimension m and a delay time τ, M is defined as:

$$M := N - (m-1)\tau. \tag{26}$$

Shalizi proposed the "magic embedding dimension", as most of the time $m = 2d + 1$ is true [81]. d denotes the dimension of the state vector. Other methods for simultaneously determining the embedding dimension and the delay time have been proposed [52].

After the reconstruction, the nearest neighbor of each point on the trajectory has to be found. This is done by searching the nearest neighbor $Y_{\hat{j}}$ to a particular reference point Y_j:

$$d_j(0) = \min_{Y_{\hat{j}}} \left(\left\| Y_j - Y_{\hat{j}} \right\| \right). \tag{27}$$

The Largest Lyapunov exponent is then estimated as the mean rate of separation of the nearest neighbors.

According to [88], the LLE was computed on a resampled RR-series (see section 5.2 for resampling). The delay time was set to approximately one beat after resampling ($\tau = 4$).

5.3.3 Detrended Fluctuation Analysis

The Detrended Fluctuation Analysis is used to quantify the fractal scaling properties of short RR-interval series [45]. It is a modification of the root-mean-square analysis of random walks applied to nonstationary signals [4]. This method has a broad field of applications, such as Deoxyribonucleic Acid (DNA) sequences [70], economics [90] and characterization of HRV, e. g., classification of Congestive Heart Failure (CHF) patients [82].

The Detrended Fluctuation Analysis (DFA) algorithm consists of the following steps [71]:

1. Integrate the signal $y(k) = \sum_{i=1}^{k} x(i) - \text{mean}(X)$.

2. Divide the profile $Y = \left\{ y_1, \ldots, y_N \right\}$ into $N_t := \lfloor N/t \rfloor$ nonoverlapping segments of equal length t.

3. Calculate the local trend for each segment by a least-squares method with a polynomial p_j:

$$F_t(j) := \frac{1}{t} \sum_{i=1}^{t} \left[y((j-1)t + i) - p_j(i) \right]^2. \tag{28}$$

4. Average over all segments:

$$F(t) := \left[\frac{1}{N_t} \sum_{j=1}^{N_t} F_t(j) \right]^{1/2}. \tag{29}$$

By repeating steps 2 to 4 for increasing time scales, the fluctuation function $F(t)$ typically also increases, following a power law:

$$F(t) \sim t^{\alpha}. \tag{30}$$

Even if F is denoted as $F(t)$, it is not a continuous function, as t represents the heart beat number and not the time.

DFA can be performed for different polynomial orders. An illustration of a first-order DFA (DFA1) of a healthy subject is shown in figure 14.
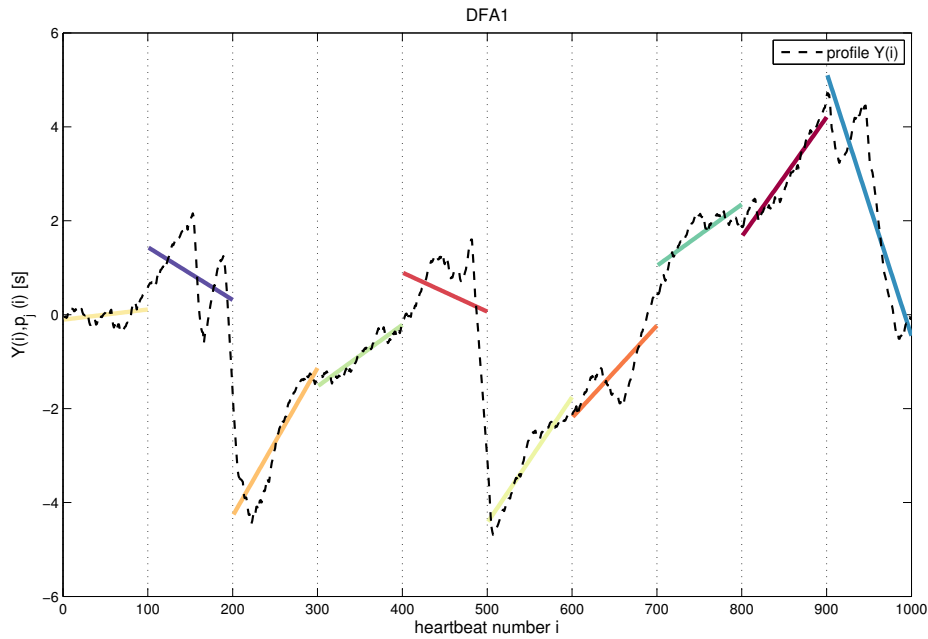
Figure 14: DFA1 of a nonpathological subject. Segment length: $t = 100$. In each segment a polynomial is fitted - in a least-squares sense - to the (black dashed) profile.

The plot of the fluctuation function $F(t)$ over t on a double logarithmic scale is shown in figure 15. The discrepancy of the fluctuation function to the linear fit at the beginning indicates to define two scaling exponents, i.e., α_1 and α_2. The range of the short-term scaling exponent α_1 was set to $4 \leq t \leq 11$ beats, hence, the long-term scaling exponent α_2 has a range of $11 < t < \lfloor N/4 \rfloor$ beats [72].
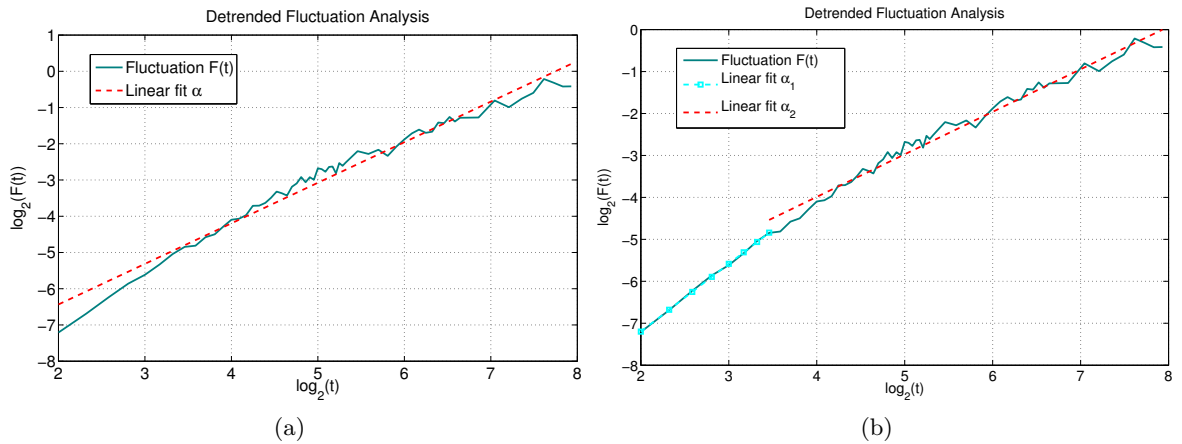
Figure 15: Scaling methods of Detrended Fluctuation Analysis: (a) one scaling exponent α, (b) short-term scaling exponent α_1 and long-term scaling exponent α_2, respectively.

5.3.4   Fractal Dimension

The terms "Fractal" and "Fractal Dimension" were first introduced by Mandelbrot in 1967 in the paper: "How long is the coast of Britain? Statistical Self-Similarity and Fractional Dimension" [57]. Mandelbrot noted that the total length of the coastline changes, with different measuring sticks used. In detail, the length of the coastline increases as the used measuring stick is scaled smaller.

The term fractal is used for objects exhibiting a repeating pattern that displays at every scale. If the subunits exactly resemble the larger scale structure, this property is known as self-similarity [33]. Famous examples of fractals are the Koch[1] curve (also known as Koch snowflake), or the Sierpińksi[2] Triangle.

For ordinary geometric objects, the Fractal Dimension (FD) equals the well known Euclidean or topological dimension, i. e., lines are one-dimensional, surfaces have two dimensions and solids are three-dimensional. But the fractal dimension has not to be an integer, usually it is greater than its topological dimension [30].

To understand the need of the FD, the Koch curve is considered. Its topological dimension is one, but the length of the curve between two arbitrary points is infinite. One can assume that the curve can be explained as a fractal line, but it is to simple to be two-dimensional [41]. The FD of the Koch curve is defined as $FD := \frac{\log(4)}{\log(3)} \approx 1.26$, which lies indeed between one and two. The first two iterations of the Koch curve are plotted in figure 16.

---

1 Niels Fabian Helge von Koch (January 25, 1870 – March 11, 1924)
2 Wacław Franciszek Sierpiński (March 14, 1882 – October 21, 1969)

Figure 16: The first two iterations of the Koch curve. In each iteration the lines are divided into three equal parts. The middle part is duplicated and then arranged like a triangle.

5.3.4.1 *Higuchi's algorithm*

Before introducing Higuchi's method for calculating the Fractal Dimension, several meanings of the term Fractal Dimension have to be clarified. Especially Higuchi's FD should not be confused with the Correlation Dimension, which describes the FD of an attractor in the phase space.
Beginning at this point, the abbreviation FD refers to Higuchi's Fractal Dimension.
FD is calculated directly from the analyzed signal. This marker assesses the self-similarity in signals [82]. Thus, the FD has been applied to different biomedical signals, such as Electroencephalography (EEG) and Electrocardiogram (ECG) recordings [27].
The following algorithm was proposed by Higuchi in 1988 [42].

1. For a fixed $k \in \left\{1, \ldots, k_{\max}\right\}$, k new time series are constructed:

$$x_m^k = \left\{x(m), x(m+k), x(m+2k), \ldots, x(m+\lfloor a \rfloor k)\right\} \quad m = 1, \ldots, k \tag{31}$$

   where $\lfloor a \rfloor := \left\lfloor \frac{N-m}{k} \right\rfloor$.

2. Compute the length of each time series:

$$L_m(k) = \sum_{i=1}^{\lfloor a \rfloor} |x(m+ik) - x(m+(i-1)k)| . \tag{32}$$

3. Normalize the lengths for each k:

$$\widetilde{L_m}(k) = L_m(k) \frac{N-1}{\lfloor a \rfloor k} . \tag{33}$$

4. Calculate the average length:

$$L(k) := \mathrm{mean}_m \left(\widetilde{L_m}(k) \right) . \tag{34}$$

Theses steps are repeated for $k = 1, \ldots, k_{\max}$.

The *FD* is defined as the slope of the linear fit (in a least squares sense) of the plot $1/\log(L(k))$ versus $\log(k)$. (see figure 17)

The implemented method is an adaption of Higuchi's algorithm with a moving time window of fixed window length. Hence, this method is sometimes specified as "running FD" [50]. According to [74], the parameter k_{max} was set to $k_{max} = 10$ and the RR-intervals are divided into windows with a length of 100 samples, shifted by one sample. The FD of one entire record was obtained by averaging over all FDs of all windows.
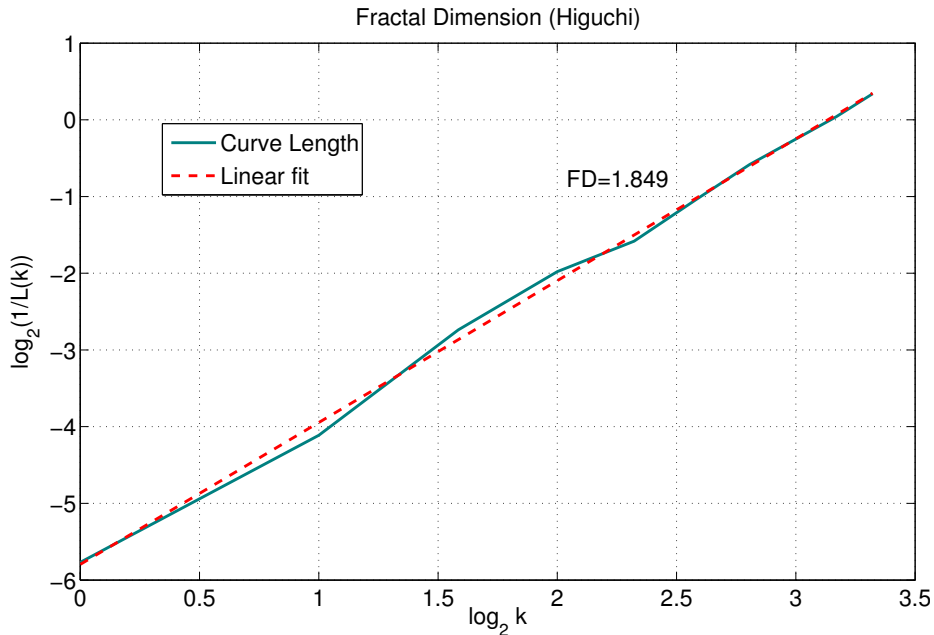
Figure 17: FD calculated on a window of 100 samples of a nonpathological subject. $k_{max} = 10$

5.3.4.2   *Katz's algorithm*

Katz's method gained recent interest in analyzing biomedical signals. According to Google Scholar it has been cited by at least 400 papers. Researchers published the better performance of Katz's algorithm over Higuchi's approach, applying them on EEG data [28]. Katz described the fractal dimension of a waveform, i. e., ordered collections of (x, y) point pairs, where the x-values increase monotonically. The implemented algorithm for the Katz's method is based on [49]. In this paper the fractal dimension D_{Katz} is defined as:

$$D_{Katz} := \frac{\log L}{\log d}, \tag{35}$$

where L is the total length of the curve and d is the diameter (the planar extend) of the curve. For ordered pairs of points $(x_i, y_i)_{i=1}^N$ the distance is calculated as:

$$d_{i,j} := \|(x_j, y_j) - (x_i, y_i)\|. \tag{36}$$

Hence, the diameter is the maximum distance and the total length is the cumulative sum of successive distances, i. e.,

$$d = \max\left(d_{i,j}\right) \quad \text{and} \quad L = \sum_{i=1}^{N-1} d_{i,i+1}. \tag{37}$$

To overcome the problem of different measurement units, Katz introduced a general unit or yardstick. More precisely this yardstick a is defined as average distance between successive points. With this yardstick $a = L/N$, it follows:

$$D_{Katz} := \frac{\log(L/a)}{\log(d/a)} = \frac{\log(N)}{\log(N) + \log(d/L)}. \tag{38}$$

For mono-dimensional sequences, such as HRV time series, Katz generated a bi-dimensional curve, in order to calculate the fractal dimension, i. e., $(t_i, RR_i)_{i=1}^N$. The time-axis can be interpreted as beat number, or continuous time, when the heart beat occurs. Therefore, equation (36) sums together different units and the use of the general unit is obsolete.

According to [19], the algorithm is simplified and even faster. Instead of generating a bi-dimensional curve, the calculations of length and diameter are direct computed using the one-dimensional sequence. The total length L is computed thus, by using:

$$L := \sum_{i=1}^{N-1} |x(i+1) - x(i)| \tag{39}$$

and the maximum extend, i. e., the range of X is defined as:

$$d := \max(X) - \min(X). \tag{40}$$

The definition of the modified Katz's dimension results in the same form, as Katz proposed:

$$D := \frac{\log(L)}{\log(d)}. \tag{41}$$

5.3.5 Hurst Exponent

The Hurst exponent, denoted by H, is a measure for long-term memory and fractality of a time series [6]. In the last decade it became popular in the finance community [21, 75], but the Hurst exponent is also applied to EEG data [24] and HRV sequences [58, 51].

For self-similar models such as fractional Gaussion noise and fractional Brownian motion, the Hurst exponent H is related to the FD and the following equation holds:

$$H = E + 1 - FD, \quad \text{where} \tag{42}$$

E denotes the Euclidean dimension, hence, for the one-dimensional RR-interval series the equation becomes:

$$H = 2 - FD. \tag{43}$$

Values for the Hurst exponent ranges from 0 to 1. Furthermore, the specific value of $H = 0.5$, indicates a random walk behavior. For $H > 0.5$ the time series is characterized by long-time correlations on all time scales [58], i.e., a trend reinforcing series. If $H < 0.5$, it indicates a long term switching in adjacent values. I.e., after a high value in the series there will probably be a low value, and so on. As a result the series indicates no trend or tendency, in contrast there is a return to the average value [24].

In this section, the method to calculate the Hurst exponent by rescaled range analysis is described. The algorithm is the following [75]:

1. For $j = 1, \ldots, N$ calculate the mean adjusted series:

$$y_j = x_j - \text{mean}(X).$$

2. Calculate cumulative deviate series: $Z_t = \sum_{j=1}^{t} y_j$.

3. Calculate range series: $R_t = \max(Z_1, \ldots, Z_t) - \min(Z_1, \ldots, Z_t)$.

4. Calculate standard deviation series: $S_t = \sqrt{\frac{1}{t} \sum_{j=1}^{t} y_j^2}$.

5. Rescaled range series is the fraction: $(R/S)_t := \frac{R_t}{S_t}$.

The implemented algorithm starts with one initial data block of

$$t_0 = 2^{\lfloor \log_2(N) \rfloor} \tag{44}$$

data points. E.g., for a length of N=1024 RR-intervals, the initial data block consits of $t_0 = 2^{10} = 1024$ points. This length is chosen for many reasons. First, it is easy to cut each segment in every step into halves. Second, for Test case I, the initial data length is 1100 points, which is close to 2^{10}. Finally, if HRV data is also applied to the frequency domain, data lengths of the power of two allow directly FFT computation.

After calculating steps 1.-5., the data sequence is divided in half. Steps 2.-5. are calculated for each partition. The rescaled range $(R/S)_1$ is the mean value of both rescale range values. This process continued up to a minimal segment length of $2^3 = 8$ data points. Figure 18 illustrates the first two iterations.
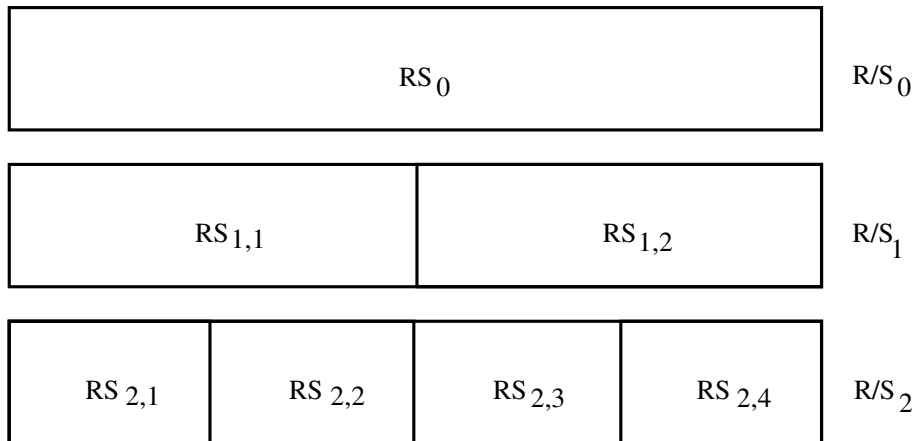
Figure 18: The initial and first two steps of the implemented rescaled range analysis. In each step every partition is divided into two halves. The rescaled range value is the average value of all the partitions in one step.

Hurst found that following the power law holds:

$$(R/S)_t \sim t^H. \tag{45}$$

The Hurst exponent H can be obtained by fitting a straight line in the double logarithmic plot of (R/S) over t.

5.4 IMPLEMENTATION

All calculations were done with MATLAB®version R2013b. The Wavelet toolbox [64] and WaveLab [17], a collection of MATLAB®functions related to wavelet analysis, were employed in order to calculate the scaling exponent τ with the WTMM method.

The calculation of the Largest Lyapunov Exponent is a modification of the MATLAB®function *lyarosenstein.m*, available on MathWorks®file exchange server[3]. The pretty illustrations of the correlations were done with the *schemaball.m* function, also available on MathWorks®file exchange[4].

After importing data from PhysioNet and filtering data with the Ensemble Density-Based Spatial Clustering of Applications with Noise (EDBSCAN) algorithm, several routines are executed, in order to obtain results for the test cases. Figure 19 gives an illustration of the

---

3 http://www.mathworks.com/matlabcentral/fileexchange/
38424-largest-lyapunov-exponent-with-rosenstein-s-algorithm/content/lyarosenstein.m
4 http://www.mathworks.com/matlabcentral/fileexchange/42279-schemaball

three main steps: Importing data and filtering, test case routines and statistics routines. The test case routines are described in detail:

- test_length($n_0, step$):
  The function `test_length.m` has two input parameters. n_0 specifies the initial data length and *step* defines the decrement. In this thesis there are two restrictions. First, $n_0 \leq 1100$ must be true, since this is the minimum data length after clustering. The second restriction is that the shortest data length has to include more than 100 data points, i. e.,

  $$n_0 - k \cdot step \geq 100, \quad k \in \mathbb{N}$$

  is true. E. g., test_length(600, 200), calculates the indices for data lengths: 600, 400, 200.

- test_nonPath(N):
  The test function for test case II has one input argument, i. e., the desired data length N. If the first test case was executed before and indices for the desired data length have already been computed, this function is obsolete. The termination condition is:

```
IF mod(n0-N,step) == 0 THEN
    exit;
ELSE
    run test_nonPath(N)
END
```

- test_crisAB(N):
  This function has one input argument, the desired data length N. `test_crisAB.m` computes the indices for the pre- and post-anti arrhythmic treatment databases, respectively.

There is no routine for test case III, i. e., age-dependency test, since the calculation is included in test_nonPath(N).

5.5 SUMMARY

Before presenting the results table 5 summarizes the implemented indices and their abbreviation used in this thesis.

Figure 19: Schematic representation of the implementation steps. The dotted lines from the routine test_length indicate that test_nonPath has not always to be calculated, in order to reach the section of statistics for test case II-III. $name \in \{\texttt{nonpath}, \texttt{path}, \texttt{crisa}, \texttt{crisb}\}$. nonPath := $\{\texttt{nonpath}, \texttt{path}\}$, crisAB := $\{\texttt{crisa}, \texttt{crisb}\}$.
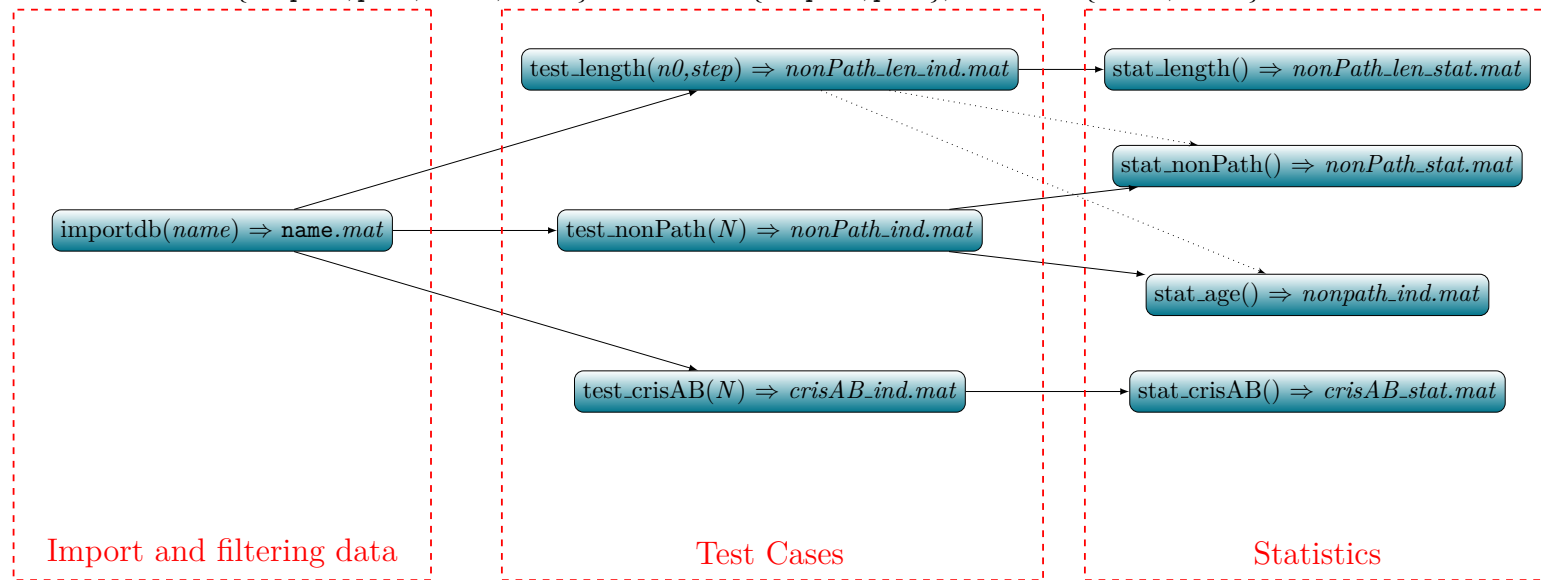
Table 5: Overview of implemented indices and their abbreviation with the unit.

Abbreviation (Unit)	Description
Statistical Indices	
$SDNN$ (s)	Standard Deviation of the NN intervals
$SDSD$ (s)	Standard Deviation of the Successive Differences
$NN50$ (#)	Number of interval differences of successive NN intervals greater than 50 ms
$pNN50$ (%)	percentage of interval differences of successive NN intervals greater than 50 ms
Geometrical Indices	
HRV_{Idx}	Fraction of N over maximum of density distribution function
$TINN$ (s)	Triangular Interpolation of NN interval histogram
Time-Frequency Index	
τ	Slope of the scaling exponent of Continuous Wavelet Transform Analysis
Chaos Descriptors	
CD	Correlation Dimension
LLE	Largest Lyapunov Exponent
Fractal Descriptors	
α_1	Short-time scaling exponent Detrended Fluctuation Analysis
α_2	Long-time scaling exponent Detrended Fluctuation Analysis
FD	Higuchi's Fractal Dimension
D	Katz's Fractal Dimension
H	Hurst exponent

6 | RESULTS

The results are presented for each test case. For Test Case I, the p-values for the different data lengths and all the indices are diagrammed. For the other test cases the p-value is tabulated for each index. Furthermore, the median and 95% central range or, for normally distributed indices, the mean and standard deviation are tabulated.

6.1 TEST CASE I: DATA LENGTH SENSITIVITY

Two of the plotted indices for the length test must be specified in detail. The Hurst exponent H, reduces the data length stepwise in steps of the power of two. Hence, the values of H will not change between 1000 and 550 points. Higuchi's FD is realized with a shifting window of 100 samples, shifted by one beat. As a result the value at $N=100$ is exactly the value of one window. At $N=200$, the value of FD is averaged over ten windows, each consisting of 100 samples.

As shown in figure 20, all of the 14 implemented indices exhibit significant differences for most of the data lengths. The Continuous Wavelet Transform (CWT) index τ is not significant for signals shorter than 400 points except for signals of length 100. Furthermore, all the indices differentiate significantly for data lengths of 1100 points to a lower bound of 400 points.

Figure 21 shows the highly significant differentiation behavior ($p < 0.01$) for all indices, which are significant for all data lengths, i.e., all except the CWT index τ are plotted in this figure.
Between 450 and 200 points, Katz's Dimension D is significant, otherwise very significant too. The geometrical index HRV_{Idx} and the long term scaling exponent α_2 have one outlier above the highly significance level. All the other indices show a very significant differentiation behavior, hardly with respect to length of the signal. The only index with minor fluctuations is the statistical index $SDNN$, but it is still very significant for all data lengths.
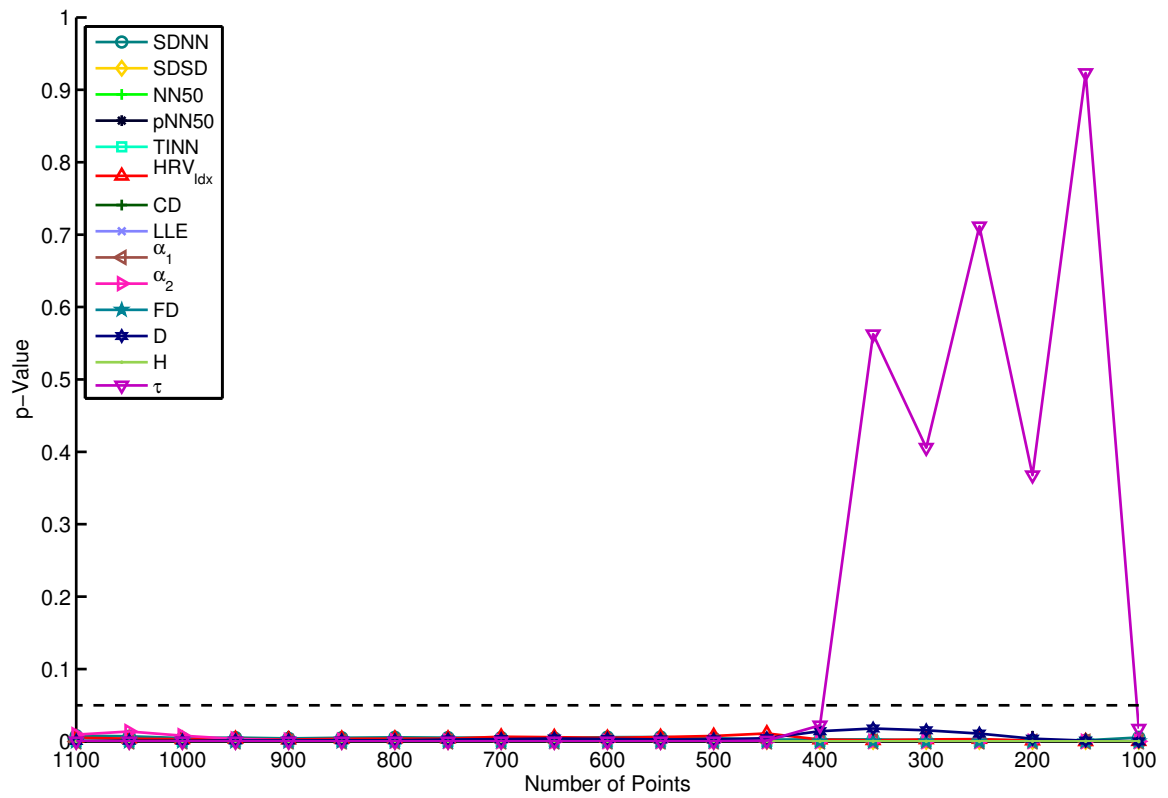
Figure 20: Indices with at least one p-value below the significance threshold of 0.05, marked with the dashed line.
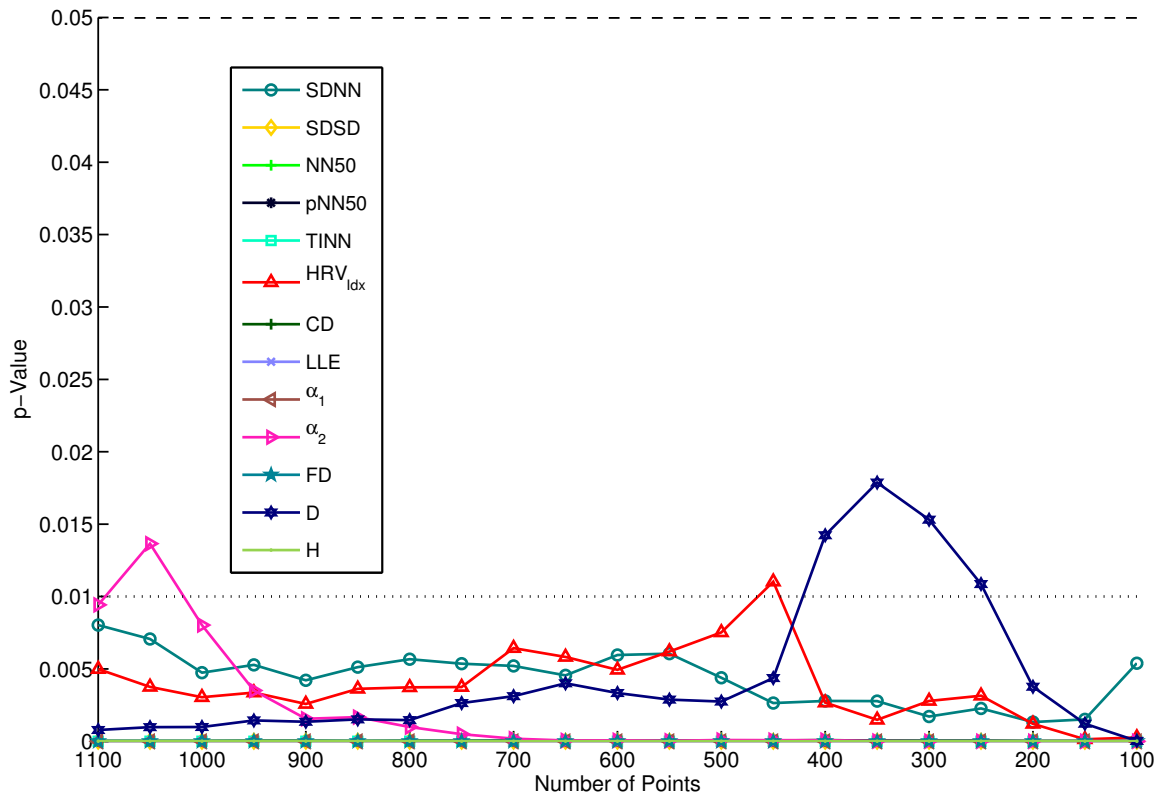
Figure 21: Indices with significant, and for some Number of points, very significant p-value. The dashed line marks the significance threshold of 0.05 and the dotted line represents the very significance threshold of 0.05, respectively.

Table 6 lists the maximum p-values for each index, when partitioning the data lengths in three sections. The section in the middle is found to be the optimal region, since all indices were very significant for data lengths in that range, i. e., from 1000-500 points.

	Test Case I		
	Upper lengths [1100-1050] max. p-Value	Optimal region [1000-500] max. p-Value	Lower lengths [450-100] max. p-Value
Statistical Indices			
SDNN (s)	**< 0.01****	**< 0.01****	**< 0.01****
SDSD (s)	**< 0.01****	**< 0.01****	**< 0.01****
NN50 (#)	**< 0.01****	**< 0.01****	**< 0.01****
pNN50 (%)	**< 0.01****	**< 0.01****	**< 0.01****
Geometrical Indices			
TINN (s)	**< 0.01****	**< 0.01****	**< 0.01****
HRV_{Idx}	< 0.01**	< 0.01**	< 0.05*
Time-Frequency Indices			
τ	< 0.01**	< 0.01**	0.923
Chaos Descriptors			
CD	**< 0.01****	**< 0.01****	**< 0.01****
LLE	**< 0.01****	**< 0.01****	**< 0.01****
Fractal Descriptors			
α_1	**< 0.01****	**< 0.01****	**< 0.01****
α_2	< 0.05*	< 0.01**	< 0.01**
FD	**< 0.01****	**< 0.01****	**< 0.01****
D	< 0.01**	< 0.01**	< 0.05*
H	**< 0.01****	**< 0.01****	**< 0.01****

Table 6: p-values and optimal region for test case I. All significant results ($p < 0.05$) are marked with *. Very significant results ($p < 0.01$) are marked with **. If an index is very significant for all data length it is bold too.

6.2 OVERVIEW OF TEST CASES II-IV

In table 7, the p-values for test cases II-IV are listed. If the data was normally distributed, the p-value was calculated with the t-test. For not normally distributed data, either the Wilcoxon rank-sum test was used for test case II-III, or the Wilcoxon signed rank test was applied for test case IV.

Index	Test Case II: Nonpath./Path. p-Value	Test Case III: Young/Old p-Value	Test Case IV: Pre./Post. p-Value
Statistical Indices			
SDNN	**< 0.01****	0.852	**< 0.01****
SDSD	**< 0.01****	0.593	**< 0.01****
NN50	**< 0.01****	0.946	**< 0.01****
pNN50	**< 0.01****	0.946	**< 0.01****
Geometrical Indices			
TINN	**< 0.01****	0.623	0.075
HRV_{Idx}	**< 0.01****	0.665	**< 0.01****
Time-Frequency Indices			
τ	**< 0.01****	0.890	**< 0.01****
Chaos Descriptors			
CD	**< 0.01****	0.082	0.739
LLE	**< 0.01****	0.755	**< 0.01****
Fractal Descriptors			
α_1	**< 0.01****	0.544	**< 0.01****
α_2	**< 0.01****	0.456	**< 0.01****
FD	**< 0.01****	0.707	**< 0.01****
D	**< 0.01****	0.716	0.907
H	**< 0.01****	0.886	**< 0.01****

Table 7: p-values of test cases II-IV. All significant results ($p < 0.05$) are bold and marked with *. Very significant results ($p < 0.01$) are marked with **

6.3 TEST CASE II: NONPATHOLOGICAL VS. PATHOLOGICAL SUBJECTS

The first column of table 7 shows that all indices can differentiate between nonpathological and pathological data very significantly.

Table 8 shows the distribution parameters of each index for test case II. If the calculated values of the index are normally distributed, the p-value and the Standard Deviation (SD) are

listed. Otherwise, the median and the limits of 95% central range, i.e., the 2.5th percentile and the 97.5th percentile, are given.

| | Test Case II | |
| | median [95% central range] or mean (SD) | |
Index	Nonpathological	Pathological
Statistical Indices		
$SDNN$ (s)	0.0560 [0.0190,0.1801]	0.0784 [0.0151,0.2435]
$SDSD$ (s)	0.0195 [0.0074,0.1408]	0.0689 [0.0083,0.2474]
$NN50$ (#)	73.5 [0,701]	145 [1,835]
$pNN50$ (%)	7.1777 [0,68.4570]	14.1602 [0.0977,81.5430]
Geometrical Indices		
$TINN$ (s)	0.2011 [0.0796,0.4799]	0.1387 [0.0326,0.5048]
HRV_{Idx}	13.3862 [5.5510,29.6157]	10.8936 [3.2753,31.7576]
Time-Frequency Indices		
τ	0.5535 (0.0695 SD)	0.6226 (0.1326 SD)
Chaos Descriptors		
CD	7.4316 (2.9194 SD)	5.4074 (3.7146 SD)
LLE	0.1533 (0.0314 SD)	0.1075 (0.0505 SD)
Fractal Descriptors		
α_1	1.2703 [0.4341,1.6246]	0.5880 [0.1803,1.4238]
α_2	0.9187 (0.1714 SD)	0.8276 (0.2886 SD)
FD	1.7010 (0.1478 SD)	1.9121 (0.1332 SD)
D	1.1066 [0.9076,1.6008]	1.2407 [0.8591,2.6361]
H	0.3029(0.0696 SD)	0.2356 (0.1004 SD)

Table 8: Statistical parameters for test case II. For normally distributed indices mean and SD are listed, otherwise median and 95% central range.

6.4 TEST CASE III: YOUNG VS. OLD SUBJECTS

For the third test case none of the indices was able to obtain significant differences between young and elderly people. The p-values are listed in table 7.

Table 9 shows the distribution parameters of all indices.

| | Test Case III | |
| | median [95% central range] or mean (SD) | |
Index	Young	Old
Statistical Indices		
$SDNN$ (s)	0.0537 [0.0279, 0.1558]	0.0565 [0.0178, 0.1919]
$SDSD$ (s)	0.0184 [0.0067, 0.1887]	0.0203 [0.0077, 0.1086]
$NN50$ (#)	74 [0, 582.15]	73.5 [0, 725.35]
$pNN50$ (%)	7.2266 [0, 56.8506]	7.1777 [0, 70.8350]
Geometrical Indices		
$TINN$ (s)	0.2310 (0.0992 SD)	0.2204 (0.1207 SD)
HRV_{Idx}	15.0990 (6.0606 SD)	14.5495 (6.9383 SD)
Time-Frequency Indices		
τ	0.5524 (0.0702 SD)	0.5542 (0.0695 SD)
Chaos Descriptors		
CD	6.8615 (3.0273 SD)	7.8478 (2.7889 SD)
LLE	0.1522 (0.0312 SD)	0.1541 (0.0317 SD)
Fractal Descriptors		
α_1	1.2206 (0.3189 SD)	1.1852 (0.2915 SD)
α_2	0.9333 (0.1608 SD)	0.9086 (0.1790 SD)
FD	1.6946 (0.1557 SD)	1.7054 (0.1431 SD)
D	1.1060 [0.9065, 1.5856]	1.1122 [0.9078, 1.7349]
H	0.3040 (0.0690 SD)	0.3021 (0.0705 SD)

Table 9: Statistical parameters for test case III. For normally distributed indices mean and SD are listed, otherwise median and 95% central range.

6.5 TEST CASE IV: PRE- VS. POST-ANTIARRHYTHMIC TREATMENT

The last column in table 7 shows that all indices can differentiate between data before and after antiarrhythmic treatment very significantly, except for CD, Katz's Dimension D and the triangular index $TINN$. These three indices show no significance.

Figure 22 shows two boxplots of indices calculated from the pre- and post-antiarrhythmic treatment database. The left boxplot shows one index with no significance in this test case,

i. e., *CD* (see figure 22a). The right boxplot represents the highly significant short-term Detrended Fluctuation Analysis (DFA) exponent α_1.

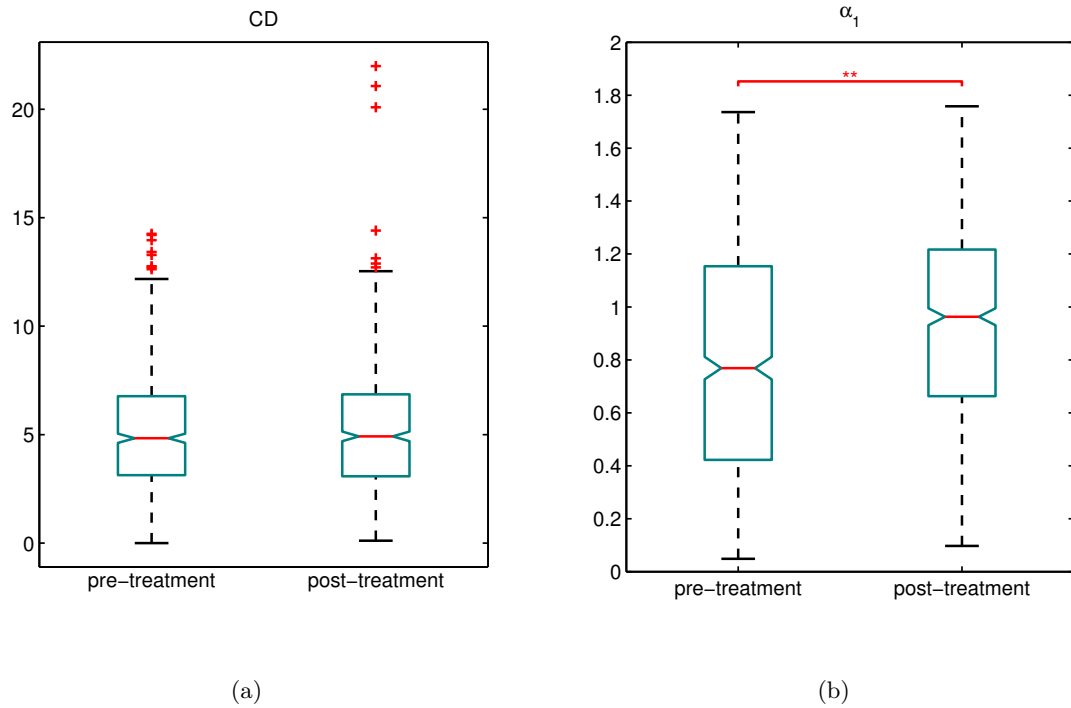
(a)   (b)

Figure 22: Boxplots for test case IV. of the (a) Correlation Dimension index *CD*, (b) short-term Detrended Fluctuation Analysis exponent α_1.

The distribution parameters for the last test case are presented in table 10.

| Index | Test Case IV median [95% central range] or mean (SD) | |
	Pre-Treatment	Post-Treatment
Statistical Indices		
$SDNN$ (s)	0.0492 [0.0123, 0.1998]	0.0377 [0.0105,0.1301]
$SDSD$ (s)	0.0248 [0.0061, 0.2233]	0.0125 [0.0058,0.1321]
$NN50$ (#)	47 [0, 739.15]	12 [0,518.0250]
$pNN50$ (%)	4.5898 [0, 72.1826]	1.1719 [0,50.5884]
Geometrical Indices		
$TINN$ (s)	0.1298 [0.0386, 0.4263]	0.1268 [0.0430, 0.3902]
HRV_{Idx}	9.4815 [2.9777,26.2564]	8.8276 [3.1251,23.2727]
Time-Frequency Indices		
τ	0.6250 (0.0852 SD)	0.6053 (0.0802 SD)
Chaos Descriptors		
CD	4.8334 [0.8016, 10.8110]	4.9201 [0.8277,11.4276]
LLE	0.1120 (0.0427 SD)	0.1030 (0.0391 SD)
Fractal Descriptors		
α_1	0.7689 [0.1720, 1.4956]	0.9629 [0.2798, 1.5197]
α_2	0.9701 [0.3957, 1.3413]	1.0503 [0.5429, 1.3959]
FD	1.8172 [1.5283, 2.0994]	1.7960 [1.5269, 2.0534]
D	1.0910 [0.8049, 2.5588]	1.1090 [0.8477, 2.0787]
H	0.3044 [0.0722,0.4449]	0.3193 [0.1210,0.4424]

Table 10: Statistical parameters for test case IV. For normally distributed indices mean and SD are listed, otherwise median and 95% central range.

6.6 CORRELATION

In figures 23-24 the correlation between indices is shown. For this purpose the Spearman's rank correlation coefficient r was calculated. If $|r| > 0.85$ holds, the correlation is declared as strong. Only correlations of indices, which showed significant differences either in test case II or test case IV were calculated. The correlation matrices, i.e., r-values of tested indices, are given in the appendix (see A.3).

Furthermore, all the strong correlations had a p-value less than 0.01 for testing the hypothesis of no correlation against the alternative that there is a nonzero correlation.

Figure 23 shows the correlations between indices of the nonpathological database and the pathological database, respectively. Since all 14 implemented indices had a very significant difference in test case II, which used data of those two databases, none index had been rejected before calculating the correlations.

Two pairs of indices of the pathological database strongly correlate with each other (see figure 23b). I. e., the statistical indices *pNN50* and *NN50* and both geometrical indices (HRV_{Idx} and *TINN*) have a strong positive correlation. Furthermore, the fractal descriptors Hurst exponent *H* and Katz's dimension *D* have a strong negative correlation.

In figure 23a three of the statistical indices (*pNN50*, *NN50*, *SDSD*) are strongly positively correlated with each other. The geometrical indices HRV_{Idx} and *TINN* are correlated, as for the pathological data (see subfigure b). Furthermore, the HRV_{Idx} correlates with the index *SDNN*. Two of the fractal descriptors, i. e., the short-term scaling exponent α_1 and Higuchi's Fractal Dimension *FD* have a strong negative correlation.

Figure 23: Strong correlations ($|r| > 0.85$) between indices. Teal lines mark positive and red lines negative correlation of: (a) Nonpathological database. (b) Pathological database.

Figure 24 shows the correlations between indices of the pre- and post-antiarrhythmic treatment databases. The three statistical indices (*pNN50*, *NN50*, *SDSD*) are strongly positively correlated with each other, as already seen for the nonpathological database (see figure 23a). The fractal descriptors short-term scaling exponent α_1 and Higuchi's *FD* have a strong negative correlation.

The indices *TINN*, chaos descriptor *CD* and Katz's dimension *D*, are missing in figure 24,

since their p-values in test case IV (calculated with the databases `crisa` and `crisb`) were not significant.
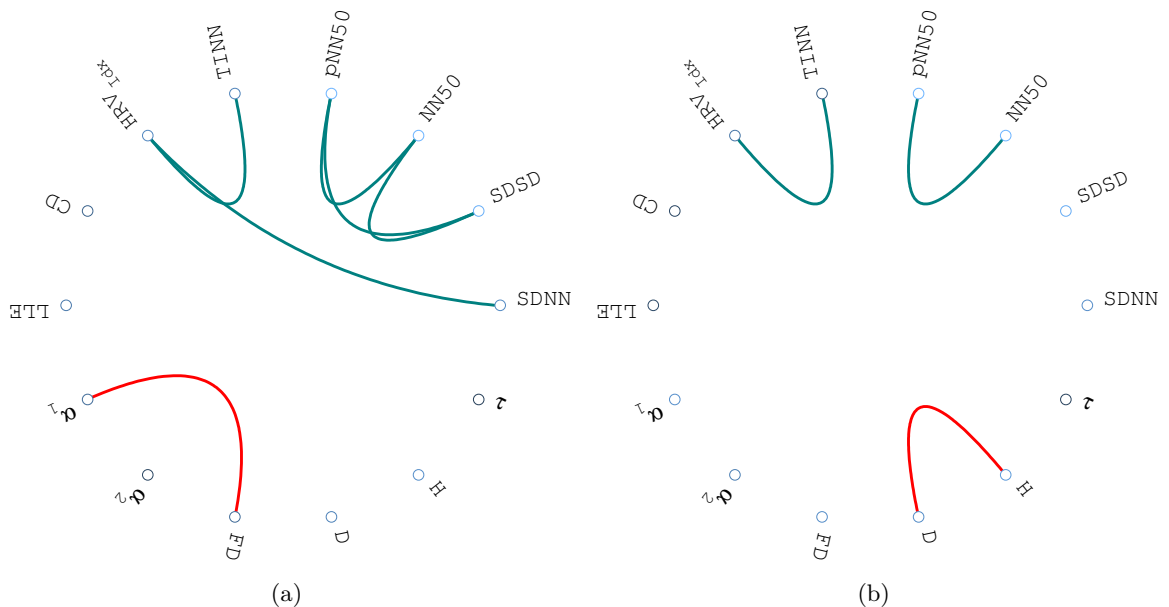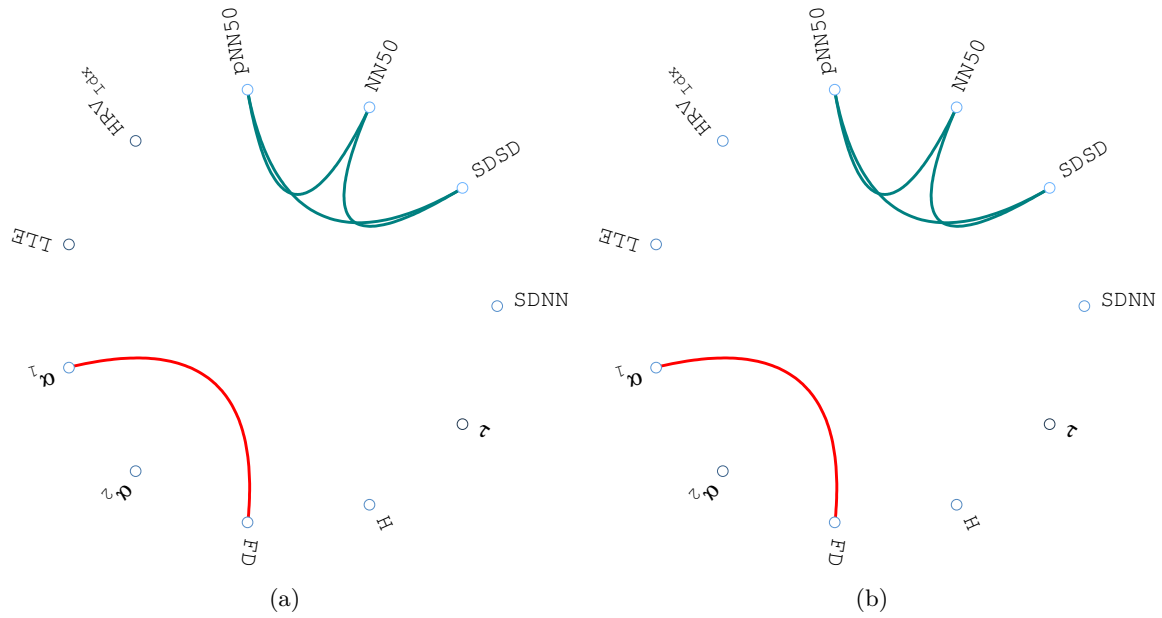

Figure 24: Strong correlations ($|r| > 0.85$) between indices. Teal lines mark positive and red lines negative correlation of: (a) pre-antiarrhythmic treatment database. (b) post-antiarrhythmic treatment database.

7 | DISCUSSION

In this chapter the groups of indices are discussed in detail. One section is dedicated to nonlinear methods and their sensitivity to age. In general, age-dependency of Heart Rate Variability (HRV) is contradictory at all, since for most of the indices studies with significant differences in age and studies which report no differences in age, can be found. Furthermore, all of the subjects for the age-dependency test, i. e., test case III are healthy.

Chapter 7 concludes with limitations of this thesis.

7.1 CRIS RESULTS

Cardiac Arrhythmia Suppression Trial (CAST) was a study to test the hypothesis that suppression of ventricular arrhythmias by antiarrhythmic drugs after myocardial infarction improves survival.

Before considering the CAST RR Interval Sub-Study database (CRIS) results, it can be observed that table 8 shows healthy subjects in the left column, while the right column lists the indices of the pathological subjects. Hence, the left values are the "healthy ones" and with significant differences the right column represents the "unhealthy values". In test case IV, the left column represents indices calculated for the subjects before treatment and the right column lists the indices after antiarrhythmic treatment. Most of the indices show a significant difference in test case IV, too. One would assume that the values after treatment should be closer to the "healthy" values, e. g., a higher Detrended Fluctuation Analysis (DFA) scaling exponent, but the opposite is true. The ranges of the indices after treatment are even worse.

NOTE:

> CAST was stopped early because the first 14-day period of treatment with antiarrhythmic drugs after a myocardial infarction was associated with excess mortality [87].

7.2 STATISTICAL INDICES

The statistical indices *pNN50* and *NN50* show a perfect Spearman correlation of $r=1$ for all test cases (see appendix A.3). This unsurprising observation is owed to the fact, that

NN50 is a multiple of *pNN50* for a fixed data length.

The index *SDSD* correlates strongly with the last two indices mentioned, except for the pathological data with a correlation of $r = 0.8$.

The age independency of the statistical indices was also reported by Malpas et al. in [56], where the statistical index *SDSD* showed no significant differences in age for alcohol addicted people. Stein et al. published in [84] that time domain indices were significantly lower for older men compared to younger ones. In this study each older man was matched with a younger one, approximately 35 years younger. The assertion that older age is associated with a global reduction in HRV can not be verified in this thesis. Just two of the statistical indices, i.e., *NN50* and hence, *pNN50* too, have lower values among the older than the younger subjects. But there was no significant difference between these two groups as well.

7.3   GEOMETRICAL INDICES

Both geometrical indices, i.e., *TINN* and HRV_{Idx}, are strongly positively correlated (see figure 23). *TINN* is robuster against varying data lengths, while HRV_{Idx} is very significant in test case IV, i.e., differentiating antiarrhythmic treatment. The *TINN* ranges of healthy subjects can be found in [4]. Acharya et al. reported that HRV decreases with aging [4]. The reduced values can be observed in table 9, though the p-value of $p = 0.075$ is barely above the level of significance. However, one has to note that age groups in [4] ranged from 5-70 years and for test case III the age of the test groups in this thesis ranged from 20-50 years and 58-85 years, respectively (see table 1 and figure 5).

In [77], Redwood et al. proposed that HRV_{Idx} is an important marker predicting arrhythmic events and mortality following myocardial infarction. Subjects who developed arrhythmic events had a significant lower HRV_{Idx}. Cripps et al. reported that the relative risk of sudden death is significantly higher for lower values of HRV_{Idx} [23]. These results can be observed in table 10, since after antiarrhythmic treatment there was an increase of deaths and among the subjects after treatment the HRV_{Idx} is significantly lower.

7.4   TIME-FREQUENCY INDEX

The time frequency index, i.e., the scaling exponent τ is the only parameter which was not significant for all data lengths. To be more precise, τ left the level of significance for less than 200 data points. This is hardly surprising, intrinsically the Wavelet Transform Modulus Maxima (WTMM) method would require data lengths starting with $N = 2^{13} = 8192$ data points [48]. As Kantelhardt et al. stated in [48]: "For the WTMM method and short series, one has to be very careful in order not to draw false conclusions from results". Hence, the significant outcome for test cases II and IV are looking good at first glance, but their

physiological relevance is nonexistent. E. g., among the pathological subjects the index τ is higher compared to healthy ones (see table 8).

The time-frequency index was implemented nevertheless, due to the fact, that it assesses the long-time correlations as well as the Hurst exponent does. Furthermore, future work with longer data can be analyzed with the implemented algorithm.

7.5   NONLINEAR METHODS AND AGE

Before discussing each nonlinear index in detail, some statements for their sensitivity to age have to be clarified.

In [89], Vandeput et al. discussed the age dependency in HRV for several nonlinear methods. The following indices: *FD*, *CD*, α_1, α_2 and *LLE*, calculated by Vandeput et al. were significantly correlated with age. All, except one index (α_2) were significantly lower with increasing age. For α_2 the opposite was true. The lower values reflect the decreasing nonlinear behavior [89]. Their models were calculated on 300 seconds windows, sliding the window every 30 seconds. Furthermore, Vandeput et al. differentiated between night hours and day time and moreover sex of each subject. Differences in age were especially prominent during daytime. This result is also reported in [62].

In this thesis, just the two Detrended Fluctuation Analysis scaling exponents α_1 and α_2 show lower values for older subjects (see table 9). However, the varying recording times among subjects in one database, may have strongly influenced their mean value. According to [62, 89], this could be especially true for the nonpathological data.

7.6   CHAOS DESCRIPTORS

The computed values of the Correlation Dimension *CD* are similar to the reported ones in [62]. Mia et al. characterized the HRV by means of the *CD* for two groups, i. e., healthy subjects and hypertension patients and distinguish between values during the day and at night. Their outcome was that healthy subjects had a lower *CD* in daytime than at night. This relationship was not seen in hypertensive patients [62]. Since the recording time was not specified for the pathological data and especially for the nonpathological database, used in this thesis (see limitations, section 7.8), this may have affected the results among the healthy people in this thesis.

Acharya et al. reported in [3] that the index *CD* for high variability classes (such as atrial fibrillation) is rather low compared to healthy subjects. This can be verified by the results of test case II (see figure 8).

To determine the *LLE* of RR-time series, usually longer signals than in this thesis are considered (e. g., $20000 - 100000$ intervals as in [78, 83]). The implemented algorithm is

based on "Estimating the Lyapunov-Exponent Spectrum from Short Time Series of Low Precision", where reasonable results for data series of $N = 5000$ points are presented [97]. Nevertheless, significant decreases in pathological subjects (as in [83]) are observable in test case *II*.

7.7 FRACTAL DESCRIPTORS

Between two of the fractal descriptors, i.e., short-term scaling exponent α_1 and Higuchi's *FD*, a strong negative correlation exists for almost all tested databases.

Higuchi's Fractal dimension is very close to 2 for the pathological dataset. These values are also reported in [73]. The *FD* of healthy subjects is significantly smaller. Pierzchalski reported that *FD* reflects the effect of antiarrhythmic drug intake, as it decreases significantly [73]. This is not verified in test case IV, where people after treatment had a significantly higher *FD*.

One would assume that Higuchi's *FD* and Katz's *D* are correlated, which is not obtained by correlation calculation. This could be due to the fact that the implemented Higuchi's *FD* was realized by sliding windows of 100 samples, whereas the index of Katz's Dimension (*D*) was computed directly on all RR-intervals. This may have also effected some outliers in test case I, which *D* shows between 400-250 data points (see figure 21).

The values of the DFA scaling exponents (α_1 and α_2) for healthy subjects are exactly in the same range, as reported in [72]. Yeh et al. proposed in [95] that a record length of ten minutes may be the minimum to assess healthy subjects by the *DFA* method. With a recording length of 1024 RR-intervals, this condition is generally fulfilled (10min $= 600s$, hence, the average RR-interval of 1024 samples within $600s$ is $\text{mean}_i\left(RR\right) = \frac{600}{1024} \approx 0.6s$). Yeh et al. compared the scaling exponents of young and old adults for a duration of ten minutes. The short term scaling exponent α_1 among the young adults was higher compared to old subjects. For the long-term scaling exponent α_2 the opposite was true. Exactly the same results can be found in table 9. Though the differences between the two groups are not significant in this thesis. Significant differences between pathological subjects (with lower scaling exponents) and healthy subjects can be observed. Since the pathological database is a composition of several different pathologies (e.g., such as Congestive Heart Failure (CHF) and Atrial Fibrillation (AF)) their mean value is hard to interpret, but it is in the range of adults with AF and adults with CHF as reported in [95]. To be more precise, the value of α_1 corresponds to AF patients, while α_2 corresponds with CHF subjects.

The Hurst exponent H is significantly lower for pathological data. This effect is also reported in [36], where healthy subjects are compared to CHF patients. The higher value of the Hurst exponent in healthy subjects reflects the higher RR variations, whereas for pathological subjects the variation is low [36]. In [69], Park et al. reported that the Hurst exponent is lower with aging. The study population were children during sevoflurane anesthesia. However, it

is necessary to note that they were aged from 2-12 years.

Equation (43), i. e., the linear relation between *FD* and *H*: *H=2-FD*, can be observed by examining the medians in tables 8-10. It is all the more surprising that their Spearmans's correlations coefficient is not close to ±1. This result can be verified due to the reason, that generally *H* and the *FD* are independent from each other. *FD* is a local property and *H* reflects long-memory dependence, which is a global characteristic [32]. For ideal self-similar processes, the local properties are reflected in the global ones and equation (43) holds [32]. Hence, for Brownian motion *H=0.5* and *FD=1.5* is true.

7.8 LIMITATIONS

The outcome of this thesis is bound by several limitations. The first limitation concerns the composition of the pathological and nonpathological database. They are not as homogenous, as they should be. Precisely, the pathological database includes subjects with different pathologies, including people with and without arrhythmias. Furthermore, the starting time of the recordings of both databases was not specified. Hence, it could have happened that one series of 1100 RR-intervals is cropped from night hours and another series of 1100 RR-intervals is cropped during day time. In [76], Ramaekers et al. reported significant differences for the statistical methods between different times of the day. The same observations can be found in [62, 89].

Furthermore, the separated test groups for testing the indices for age dependency have a wide range of ages. The threshold of less than 53 years to declare a subject as "young" is rather high.

Moreover, the minimum data length after filtering was 1100 RR-intervals. Hence, the shortest recording is "svdb865" of the supraventricular arrhythmia database with 453 seconds, which corresponds to approximately 7.5 minutes. A wide range of indices requires longer data for their calculation (e. g., the index τ, obtained by the WTMM method and several papers calculated the Hurst exponent and Detrended Fluctuation Analysis of longer RR-seqences, e. g., [89]).

Last but not least, since an automatic filtering method, i. e., the Ensemble Density-Based Spatial Clustering of Applications with Noise (EDBSCAN) algorithm was used in this thesis to remove the ectopic beats before calculating the HRV parameters, there is no guarantee that all the "missed and extra beats" are rejected.

8 | CONCLUSION

At the beginning of this thesis, in section 1.1, the research questions were declared. After executing the test cases the answers are the following:

- Most of the implemented indices are insensitive to varying data lengths. Only the Continuous Wavelet Transform index τ needs more than 400 RR-intervals in order to obtain significant differences.

- All indices are able to differentiate between nonpathological subjects and pathological subjects. Furthermore, all but one indices showed very significant differences between those focus groups.

- None of the implemented indices could distinguish between young and elderly healthy subjects, significantly.

- Most of the indices have a significant difference between data before and post antiarrhythmia therapy. In fact, all except the geometrical index *TINN*, the chaotic descriptor *CD* and Katz's Dimension *D*.

- The statistical indices *pNN50*, *NN50*, and *SDSD* are strong positive correlated for all data tested, except the pathological database. The same is true for both geometrical indices, i. e., HRV_{Idx} and *TINN*. Furthermore, Higuchi's *FD* and the short time-scaling exponent α_1 show a strong negative correlation.

Some of the indices had excellent results over all test cases, except of test case III, where not a single index could differentiate between **healthy** young and elderly people. These are:

- Statistical Indices: *SDSD*, *NN50* and *pNN50*.

- Chaos Descriptors: *LLE*.

- Fractal Descriptors: α_1, *FD* and *H*.

A few methods, usually used for calculating Heart Rate Variability (HRV) parameters of 24-hour recordings, were tested on their ability to provide consistent results, even with less than 1100 data points. This is especially the case for the Detrended Fluctuation Analysis (DFA), which was developed as long-term correlation dimension using 24-hour recordings [71]. Since long-term recordings require more effort compared to short-term recordings, short-term indices would be more clinically practical.

Yeh et al. started this "trimming"-process in 2009 with $N = 660$ RR-intervals [96]. One year later, Yeh et al. proposed that their results offered additional evidence that DFA from long-length might be "extended" to ten minutes short-length series and furthermore, that the length dependency of DFA exponents would be an interesting and useful topic for further study [95].

Indeed, most of the implementation and interpretation process was interesting and I can propose:

> Based on our results, the short-term scaling exponent α_1 might even be trimmed from data lengths of ten minutes short-length series, to standard short-term recordings of five minutes duration.

A | APPENDIX

A.1 CORRELATION BETWEEN TINN AND HRV$_{idx}$

Cripps et al. reported in [23] the steps needed, to calculate the HRV_{Idx}:
The first step is to assign the area A of the triangle as the total number of NN-intervals, i.e., $A := N$. The index $\widetilde{HRV}_{Idx}$ is defined as the baseline width of the triangle, hence:

$$\widetilde{HRV}_{Idx} = \frac{2 \cdot A}{D_{\max}} = \frac{2 \cdot N}{D_{\max}} = 2 \cdot HRV_{Idx}.$$

Since this easy calculation is done with an equilateral triangle, this is a good approximation for data with non-zero skewness distribution. Probably, the $\widetilde{HRV}_{Idx}$ with the inability to calculate definite values of data with skewed distributions was responsible for an adaption, the Triangular Interpolation of NN interval histogram (TINN). TINN calculates both points of the width separately and hence, can handle skewed distributions.

A.2 POINCARÉ PLOT

For a given data set $X := \left\{ x_1, \ldots, x_N \right\}$ the Poincaré plot of lag m is obtained by plotting $X_1 := \left\{ x_1, \ldots, x_{N-m} \right\}$ against $X_{1+m} := \left\{ x_{1+m}, \ldots, x_N \right\}$.

The standard Poincaré plot uses a lag of 1, hence:

$$(x_1, x_2), \ (x_2, x_3), \ \ldots$$

defines the data points of the Poincaré plot. A typical unfiltered Poincaré plot of a healthy subject is shown in figure 25. For healthy subjects these plots are typically shaped like a comet, torpedo or a cigar [26]. If there is no variation in the Heart Rate (HR), the Poincaré plot just consists of a single point.
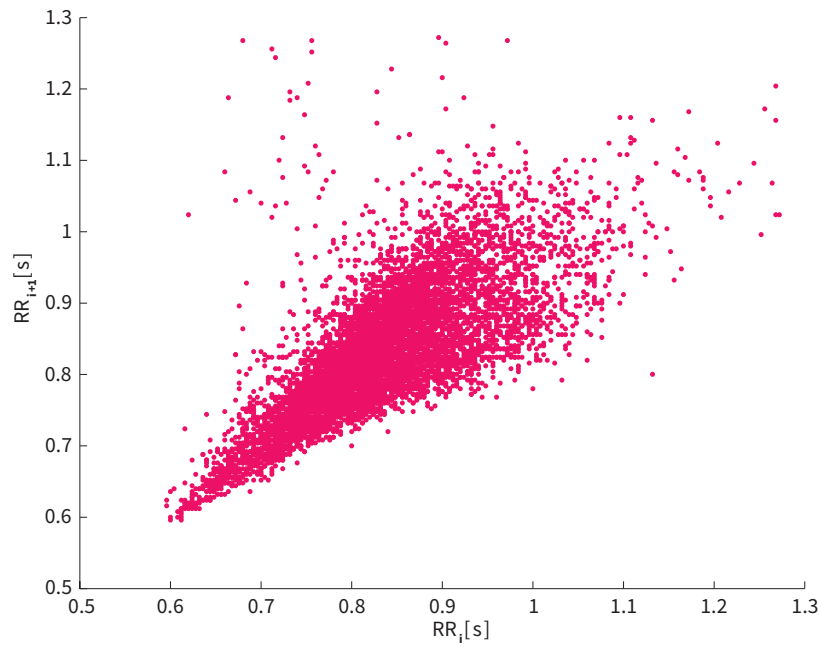
Figure 25: Standard Poincaré plot (lag 1) of a nonpathological subject [44]. Each point in the figure has coordinates (RR_i, RR_{i+1}).

A.3   CORRELATION

Tables 11-12 show the correlation matrices of all the indices, since all of them were signifcant in test case II. The correlation matrices in table 13 are given for the significant indices in test case IV. These are all except of the Triangular Interpolation of NN interval histogram *TINN*, the Correlation Dimension *CD* and Katz's Dimension *D*.

Correlation was declared strong for $|r| > 0.85$, where r denotes the sample correlation coefficient.

Table 11: The correlation coefficients for the nonpathological database.

	SDNN	SDSD	NN50	pNN50	TINN	HRV_{Idx}	CD	LLE	α_1	α_2	FD	D	H	τ
SDNN	1	0.825	0.781	0.781	0.833	0.914	0.168	0.373	−0.115	0.143	0.052	0.096	−0.205	0.193
SDSD	0.825	1	0.951	0.951	0.675	0.739	0.378	0.367	−0.45	−0.149	0.393	0.442	−0.538	0.079
NN50	0.781	0.951	1	1	0.704	0.749	0.415	0.413	−0.443	−0.121	0.427	0.557	−0.559	0.031
pNN50	0.781	0.951	1	1	0.704	0.749	0.415	0.413	−0.443	−0.121	0.427	0.557	−0.559	0.031
TINN	0.833	0.675	0.704	0.704	1	0.974	0.187	0.51	0.037	0.068	−0.066	0.181	−0.264	0.025
HRV_{Idx}	0.914	0.739	0.749	0.749	0.974	1	0.166	0.485	0.003	0.085	−0.047	0.156	−0.248	0.057
CD	0.168	0.378	0.415	0.415	0.187	0.166	1	0.472	−0.209	−0.276	0.226	0.344	−0.304	−0.361
LLE	0.373	0.367	0.413	0.413	0.51	0.485	0.472	1	0.219	−0.377	−0.227	0.226	−0.317	−0.386
α_1	−0.115	−0.45	−0.443	−0.443	0.037	0.003	−0.209	0.219	1	−0.016	−0.923	−0.668	0.568	−0.214
α_2	0.143	−0.149	−0.121	−0.121	0.068	0.085	−0.276	−0.377	−0.016	1	0.044	−0.328	0.472	0.562
FD	0.052	0.393	0.427	0.427	−0.066	−0.047	0.226	−0.227	−0.923	0.044	1	0.672	−0.546	0.208
D	0.096	0.442	0.557	0.557	0.181	0.156	0.344	0.226	−0.668	−0.328	0.672	1	−0.843	−0.123
H	−0.205	−0.538	−0.559	−0.559	−0.264	−0.248	−0.304	−0.317	0.568	0.472	−0.546	−0.843	1	0.21
τ	0.193	0.079	0.031	0.031	0.025	0.057	−0.361	−0.386	−0.214	0.562	0.208	−0.123	0.21	1

Table 12: The correlation coefficients for the pathological database.

	$SDNN$	$SDSD$	$NN50$	$pNN50$	$TINN$	HRV_{Idx}	CD	LLE	α_1	α_2	FD	D	H	τ
$SDNN$	1	0.805	0.707	0.707	0.424	0.704	0.076	0.417	-0.232	-0.185	0.158	0.236	-0.383	0.196
$SDSD$	0.805	1	0.803	0.803	0.166	0.39	0.006	0.311	-0.632	-0.564	0.543	0.432	-0.65	0.255
$NN50$	0.707	0.803	1	1	0.29	0.497	0.168	0.157	-0.534	-0.492	0.597	0.684	-0.717	0.006
$pNN50$	0.707	0.803	1	1	0.29	0.497	0.168	0.157	-0.534	-0.492	0.597	0.684	-0.717	0.006
$TINN$	0.424	0.166	0.29	0.29	1	0.87	0.482	0.225	0.16	0.133	-0.157	0.062	-0.032	-0.224
HRV_{Idx}	0.704	0.39	0.497	0.497	0.87	1	0.392	0.291	0.062	0.09	-0.072	0.189	-0.181	-0.096
CD	0.076	0.006	0.168	0.168	0.482	0.392	1	0.191	0.216	-0.055	-0.03	0.163	-0.078	-0.552
LLE	0.417	0.311	0.157	0.157	0.225	0.291	0.191	1	0.163	-0.235	-0.277	-0.059	-0.034	-0.141
α_1	-0.232	-0.632	-0.534	-0.534	0.16	0.062	0.216	0.163	1	0.459	-0.793	-0.505	0.602	-0.373
α_2	-0.185	-0.564	-0.492	-0.492	0.133	0.09	-0.055	-0.235	0.459	1	-0.624	-0.479	0.618	0.25
FD	0.158	0.543	0.597	0.597	-0.157	-0.072	-0.03	-0.277	-0.793	-0.624	1	0.648	-0.712	0.013
D	0.236	0.432	0.684	0.684	0.062	0.189	0.163	-0.059	-0.505	-0.479	0.648	1	-0.859	-0.19
H	-0.383	-0.65	-0.717	-0.717	-0.032	-0.181	-0.078	-0.034	0.602	0.618	-0.712	-0.859	1	0.099
τ	0.196	0.255	0.006	0.006	-0.224	-0.096	-0.552	-0.141	-0.373	0.25	0.013	-0.19	0.099	1

Table 13: The correlation coefficients for the antiarrhythmic treatment databases, i.e., (a) pre-treatment (`crisa`) database and (b) post-treatment (`crisb`) database.

	SDNN	SDSD	NN50	pNN50	HRV_{Idx}	LLE	α_1	α_2	FD	H	τ
SDNN	1	0.759	0.74	0.74	0.677	0.528	−0.287	−0.287	0.208	−0.388	0.191
SDSD	0.759	1	0.865	0.865	0.313	0.338	−0.715	−0.634	0.575	−0.643	0.333
NN50	0.74	0.865	1	1	0.457	0.255	−0.579	−0.539	0.565	−0.696	0.172
pNN50	0.74	0.865	1	1	0.457	0.255	−0.579	−0.539	0.565	−0.696	0.172
HRV_{Idx}	0.677	0.313	0.457	0.457	1	0.471	0.166	0.093	−0.204	−0.13	−0.127
LLE	0.528	0.338	0.255	0.255	0.471	1	0.218	−0.276	−0.321	−0.134	−0.122
α_1	−0.287	−0.715	−0.579	−0.579	0.166	0.218	1	0.517	−0.884	0.625	−0.461
α_2	−0.287	−0.634	−0.539	−0.539	0.093	−0.276	0.517	1	−0.492	0.66	0.01
FD	0.208	0.575	0.565	0.565	−0.204	−0.321	−0.884	−0.492	1	−0.64	0.287
H	−0.388	−0.643	−0.696	−0.696	−0.13	−0.134	0.625	0.66	−0.64	1	−0.048
τ	0.191	0.333	0.172	0.172	−0.127	−0.122	−0.461	0.01	0.287	−0.048	1

(a)

	SDNN	SDSD	NN50	pNN50	HRV_{Idx}	LLE	α_1	α_2	FD	H	τ
SDNN	1	0.652	0.625	0.625	0.836	0.556	0.148	0.141	−0.207	0.157	0.167
SDSD	0.652	1	0.935	0.935	0.489	0.283	−0.373	−0.304	0.301	−0.36	0.15
NN50	0.625	0.935	1	1	0.542	0.27	−0.304	−0.246	0.27	−0.379	0.091
pNN50	0.625	0.935	1	1	0.542	0.27	−0.304	−0.246	0.27	−0.379	0.091
HRV_{Idx}	0.836	0.489	0.542	0.542	1	0.566	0.321	0.208	−0.346	0.153	−0.003
LLE	0.556	0.283	0.27	0.27	0.566	1	0.549	−0.182	−0.57	0.192	−0.214
α_1	0.148	−0.373	−0.304	−0.304	0.321	0.549	1	0.192	−0.934	0.629	−0.26
α_2	0.141	−0.304	−0.246	−0.246	0.208	−0.182	0.192	1	−0.213	0.537	0.314
FD	−0.207	0.301	0.27	0.27	−0.346	−0.57	−0.934	−0.213	1	−0.677	0.152
H	0.157	−0.36	−0.379	−0.379	0.153	0.192	0.629	0.537	−0.677	1	0.129
τ	0.167	0.15	0.091	0.091	−0.003	−0.214	−0.26	0.314	0.152	0.129	1

(b)

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ACF   Autocorrelation Function

AF   Atrial Fibrillation

ANS   Autonomic Nervous System

ApEn   Approximate Entropy

AV   Atrioventricular

BIH   Boston's Beth Israel Hospital

bpm   beats per minute

CAST   Cardiac Arrhythmia Suppression Trial

CRIS   CAST RR Interval Sub-Study database

CD   Correlation Dimension

CHF   Congestive Heart Failure

CVD   Cardiovascular Diseases

CWT   Continuous Wavelet Transform

DNA   Deoxyribonucleic Acid

DBSCAN   Density-Based Spatial Clustering of Applications with Noise

DFA   Detrended Fluctuation Analysis

EEG   Electroencephalography

ECG   Electrocardiogram

EDBSCAN   Ensemble Density-Based Spatial Clustering of Applications with Noise

FD   Fractal Dimension

FFT   Fast Fourier Transform

IRF   Impulse Rejection Filter

HR   Heart Rate

HRV   Heart Rate Variability

LLE   Largest Lyapunov Exponent

MI   Mutual Information

MIT   Massachusetts Institute of Technology

NN50 Number of interval differences of successive NN intervals greater than 50 ms

NYHA New York Heart Association

pNN50 percentage of interval differences of successive NN intervals greater than 50 ms

PSD   Power Spectral Density

SA   Sinoatrial

SampEn Sample Entropy

SD   Standard Deviation

SDNN Standard Deviation of the NN intervals

SDSD Standard Deviation of the Successive Differences

TINN Triangular Interpolation of NN interval histogram

WTMM Wavelet Transform Modulus Maxima

BIBLIOGRAPHY

[1] Function of the Heart. `http://encyclopedia.lubopitko-bg.com/Function_of_the_Heart.html`. [Online; accessed 2015-08-28].

[2] Wiggers diagram — Wikipedia, The Free Encyclopedia. `https://en.wikipedia.org/wiki/Wiggers_diagram`. [Online; accessed 2015-08-28].

[3] R. Acharya, C. Lim, and P. Joseph. Heart rate variability analysis using correlation dimension and detrended fluctuation analysis. *Itmb-Rbm*, 23:333–339, 2002. doi: 10.1016/S1297-9562(02)90002-1.

[4] U. R. Acharya, N. Kannathal, O. W. Sing, L. Y. Ping, and T. Chua. Heart rate analysis in normal subjects of various age groups. *Biomedical engineering online*, 3(24):8, 2004. doi: 10.1186/1475-925X-3-24.

[5] U. R. Acharya, P. S. Bhat, N. Kannathal, A. Rao, and C. M. Lim. Analysis of cardiac health using fractal dimension and wavelet transformation. *Itbm-Rbm*, 26:133–139, 2005. doi: 10.1016/j.rbmret.2005.02.001.

[6] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri. Heart rate variability: a review. *Medical & biological engineering & computing*, 44(12):1031–51, Dec. 2006. doi: 10.1007/s11517-006-0119-0.

[7] R. Acharya U., P. S. Bhat, N. Kannathal, A. Rao, and M. L. Choo. Analysis of cardiac health using fractal dimension and wavelet transformation. *Itbm-Rbm*, 26:133–139, 2005. doi: 10.1016/j.rbmret.2005.02.001.

[8] B. Ahmadi and R. Amirfattahi. Comparison of Correlation Dimension and Fractal Dimension in Estimating BIS index. *Wireless Sensor Network*, 2(1):67–73, 2010. doi: 10.4236/wsn.2010.21010.

[9] M. Akay. *Nonlinear Biomedical Signal Processing, Dynamic Analysis and Modeling*, volume I. Wiley-IEEE Press, New York, NY, USA, 2001. doi: 10.1109/9780470545379.

[10] American Heart Association et al. Revisions to classification of functional capacity and objective assessment of patients with diseases of the heart. *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. New York: Little Brown & Co*, 1994.

[11] M. Bachler. Automatische Detektion von QRS-Komplex, P- und T-Welle im Elektrokardiogramm in Echtzeit. Master's thesis, Technische Universität Wien, 2011.

[12] M. Bachler, C. Mayer, B. Hametner, S. Wassertheurer, and A. Holzinger. Online and offline determination of QT and PR interval and QRS duration in electrocardiography. *Lecture Notes in Computer Science*, 7719 LNCS:1–15, 2013. doi: 10.1007/978-3-642-37015-1_1.

[13] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald. Survival of patients with severe congestive heart failure treated with oral milrinone. *Journal of the American College of Cardiology*, 7(3):661–670, 1986. doi: 10.1016/S0735-1097(86)80478-8.

[14] A. Bauer, M. Malik, G. Schmidt, P. Barthel, H. Bonnemeier, I. Cygankiewicz, P. Guzik, F. Lombardi, A. Müller, A. Oto, R. Schneider, M. Watanabe, D. Wichterle, and W. Zareba. Heart Rate Turbulence: Standards of Measurement, Physiological Interpretation, and Clinical Use. International Society for Holter and Noninvasive Electrophysiology Consensus. *Journal of the American College of Cardiology*, 52(17):1353–1365, 2008. doi: 10.1016/j.jacc.2008.07.041.

[15] J. T. Bigger, J. L. Fleiss, L. M. Rolnitzky, and R. C. Steinman. The ability of several short-term measures of RR variability to predict mortality after myocardial infarction. *Circulation*, 88(3):927–934, 1993.

[16] C. Bogaert, F. Beckers, D. Ramaekers, and a. E. Aubert. Analysis of heart rate variability with correlation dimension method in a normal population and in heart transplant patients. *Autonomic neuroscience : basic & clinical*, 90:142–147, 2001. doi: 10.1016/S1566-0702(01)00280-6.

[17] J. Buckheit, S. Chen, D. Donoho, I. Johnstone, and J. D. Scargle. *About WaveLab*, 20055.

[18] R. Carvajal, N. Wessel, M. Vallverdú, P. Caminal, and A. Voss. Correlation dimension analysis of heart rate variability in patients with dilated cardiomyopathy. *Computer Methods and Programs in Biomedicine*, 78:133–140, 2005. doi: 10.1016/j.cmpb.2005.01.004.

[19] P. Castiglioni. What is wrong in Katz's method? Comments on: "A note on fractal dimensions of biomedical waveforms". *Computers in Biology and Medicine*, 40:950–952, 2010. doi: 10.1016/j.compbiomed.2010.10.001.

[20] G. D. Clifford and L. Tarassenko. Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. *IEEE transactions on bio-medical engineering*, 52 (4):630–8, Apr. 2005. doi: 10.1109/TBME.2005.844028.

[21] M. Corazza. Multi-Fractality in Foreign Currency Markets. *Finance*, 6(2):65–98, 2002. doi: 10.17578/6-2-1.

[22] M. Costa, A. L. Goldberger, and C.-K. Peng. Multiscale entropy analysis of complex physiologic time series. *Physical review letters*, 89(6):068102, 2002. doi: 10.1103/PhysRevLett.92.089803.

[23] T. R. Cripps, M. Malik, and T. G. Farrell. Prognostic value of reduced heart rate variability after myocardial infarction: clinical evaluation of a new analysis method. *British heart journal*, 65(1):14–19, 1991.

[24] M. H. Díaz, F. M. Córdova, L. Cañete, F. Palominos, F. Cifuentes, C. Sánchez, and M. Herrera. Order and Chaos in the Brain: Fractal Time Series Analysis of the EEG Activity During a Cognitive Problem Solving Task. *Procedia Computer Science*, 55 (Itqm):1410–1419, 2015. doi: 10.1016/j.procs.2015.07.135.

[25] A. E. Epstein, J. T. Bigger, D. G. Wyse, D. W. Romhilt, R. A. Reynolds-Haertle, and A. P. Hallstrom. Events in the Cardiac Arrhythmia Suppression Trial (CAST): Mortality in the Entire Population Enrolled. *Journal of the American College of Cardiology*, 18(1):14–19, 1991. doi: 10.1016/S0735-1097(10)80210-4.

[26] H. D. Esperer, C. Esperer, and R. J. Cohen. Cardiac arrhythmias imprint specific signatures on lorenz plots. *Annals of Noninvasive Electrocardiology*, 13(1):44–60, 2008.

[27] R. Esteller, S. Member, G. Vachtsevanos, S. Member, J. Echauz, and B. Litt. A Comparison of Waveform Fractal Dimension Algorithms. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 48(2):177–183, 2001.

[28] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt. A Comparison of waveform fractal dimension algorithms. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 48(2):177–183, 2001.

[29] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[30] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. 2003. doi: 10.1002/0470013850.

[31] A. Fraser and H. Swinney. Independent coordinates for strange attractors from mutual information. *Physical review A*, 33(2):1134–1140, 1986.

[32] T. Gneiting and M. Schlather. Stochastic models which separate fractal dimension and Hurst effect. (1):8, 2001. doi: 10.1137/S0036144501394387.

[33] A. L. Goldberger. Non-linear dynamics for clinicians: Chaos theory, fractals, and complexity at the bedside. *Lancet*, 347(9011):1312–1314, 1996. doi: 10.1016/S0140-6736(96)90948-4.

[34] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215—-e220, 2000.

[35] E. Gospodinova and M. Gospodinov. Wavelet-based Multifractal Analysis of RR Time Series. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(4):1067–1071, 2014.

[36] E. Gospodinova, M. Gospodinov, N. Dey, I. Domuschiev, A. S. Ashour, and D. Sifaki-pistolla. Analysis of Heart Rate Variability by Applying Nonlinear Methods with Different Approaches for Graphical Representation of Results. *International Journal of Advanced Computer Science and Applications*, 6(8):38–45, 2015.

[37] P. Grassberger and I. Procaccia. Characterization of strange attractors. *Physical review letters*, 50(5):346–349, 1983.

[38] S. D. Greenwald. *Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information.* PhD thesis, Massachusetts Institute of Technology, 1990.

[39] J. E. Guyton, Arthur C and Hall. *Textbook of medical physiology.* WB Saunders Company, Philadelphia, 10th edition, 2000.

[40] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. L. Tatham. *Multivariate Data Analysis.* Pearson Prentice Hall Upper Saddle River, NJ, 7th edition, 2010.

[41] D. Harte. *Multifractals, Theory and Applications.* CRC Press, 2001.

[42] T. Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2):277–283, 1988.

[43] E. H. Hon and S. T. Lee. The fetal electrocardiogram. *American Journal of Obstetrics & Gynecology*, 91(1):56–60, 1965.

[44] M. Hörtenhuber. Analysis of Heart Rate Variability via Data Modelling of Poincaré Plots. Master's thesis, Technischen Universität Wien, 2014.

[45] H. V. Huikuri, T. H. Makikallio, C.-K. Peng, A. L. Goldberger, U. Hintze, and M. Moller. Fractal Correlation Properties of R-R Interval Dynamics and Mortality in Patients With Depressed Left Ventricular Function After an Acute Myocardial Infarction. *Circulation*, 101(1):47–53, Jan. 2000. doi: 10.1161/01.CIR.101.1.47.

[46] H. V. Huikuri, T. H. Mäkikallio, and J. Perkiömäki. Measurement of Heart Rate Variability by Methods Based on Nonlinear Dynamics. *Journal of Electrocardiology*, 36 (SUPPL.):95–99, 2003. doi: 10.1016/j.jelectrocard.2003.09.021.

[47] N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *The American journal of physiology*, 271(4 Pt 2):R1078–R1084, 1996.

[48] J. W. Kantelhardt, S. a. Zschiegner, and H. E. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316:87–114, 2002. doi: 10.1016/S0378-4371(02)01383-3.

[49] M. Katz. Fractals and the analysis of waveforms. *Computers in biology and medicine*, 18(3):145–156, 1988.

[50] W. Klonowski. Application of nonlinear dynamics in biosignal analysis. *Proceedings of SPIE - The International Society for Optical Engineering*, 5975, 2005. doi: 10.1117/12. 675580.

[51] G. Krstacic, A. Krstacic, M. Martinis, E. Vargovic, A. Knezevic, A. Smalcelj, M. Jembrek-Gostovic, D. Gamberger, and T. Smuc. Non-linear analysis of heart rate variability in patients with coronary heart disease. *Computers in Cardiology*, (29): 673–675, 2002. doi: 10.1109/CIC.2002.1166862.

[52] W. Liebert, K. Pawelzik, and H. G. Schuster. Optimal Embeddings of Chaotic Attractors from Topological Considerations. *Europhysics Letters (EPL)*, 14(6):521–526, 2007. doi: 10.1209/0295-5075/14/6/004.

[53] F. Lombardi. Chaos theory, heart rate variability, and arrhythmic mortality. *Circulation*, 101(1):8–10, 2000. doi: 10.1161/01.CIR.101.1.8.

[54] F. Lombardi, G. Sandrone, A. Mortara, D. Torzillo, M. T. L. Rovere, M. G. Signorini, S. Cerutti, and A. Malliani. Linear and nonlinear dynamics of heart rate variability after acute myocardial infarction with normal and reduced left ventricular ejection fraction. *The American Journal of Cardiology*, 77(15):1283–1288, 1996. doi: http://dx.doi.org/10.1016/S0002-9149(96)00193-2.

[55] M. Malik, R. Xia, O. Odemuyiwa, A. Staunton, J. Poloniecki, and A. Camm. Influence of the recognition artefact in automatic analysis of long-term electrocardiograms on time-domain measurement of heart rate variability. *Medical and Biological Engineering and Computing*, 31(5):539–544, 1993.

[56] S. C. Malpas, E. A. Whiteside, and T. J. Maling. Heart rate variability and cardiac autonomic function in men with chronic alcohol dependence. *British heart journal*, 65 (2):84–88, 1991.

[57] B. B. Mandelbrot. How long is the coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science*, 156:636–638, 1967.

[58] M. Martinis, a. Knežević, G. Krstačić, and E. Vargović. Changes in the Hurst exponent of heartbeat intervals during physical activity. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 70(1 1):1–9, 2004. doi: 10.1103/PhysRevE.70.012903.

[59] C. C. Mayer, M. Bachler, M. Hörtenhuber, C. Stocker, A. Holzinger, and S. Wassertheurer. Selection of entropy-measure parameters for knowledge discovery in heart rate variability data. *BMC Bioinformatics*, 15(Suppl 6):S2, 2014. doi: 10.1186/1471-2105-15-S6-S2.

[60] J. Mcnames, T. Thong, and M. Aboy. Impulse rejection filter for artifact removal in spectral analysis of biomedical signals. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 1, pages 145–148. IEEE, 2004.

[61] P. S. Melanie Nichols, Nick Townsend and M. Rayner. European Cardiovascular Disease Statistics 2012. Technical report, Brussels, 2012.

[62] X. Miao, W. He, H. Yang, and H.-M. Tai. Heart rate variability characterization using correlation dimension. In *Circuits and Systems, 2002. MWSCAS-2002. The 2002 45th Midwest Symposium on*, volume 1, pages I–447. IEEE, 2002.

[63] D. C. Michaels, D. R. Chialvo, E. P. Matyas, and J. Jalife. Chaotic activity in a mathematical model of the vagally driven sinoatrial node. *Circulation research*, 65(5): 1350–1360, 1989. doi: 10.1161/01.RES.65.5.1350.

[64] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi. Wavelet toolbox. *Matlab User's Guide*, 1997.

[65] G. B. Moody and R. G. Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001. doi: 10.1109/51.932724.

[66] M. M. Mukaka. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3):69–71, 2012.

[67] B. Olshansky, H. N. Sabbah, P. J. Hauptman, and W. S. Colucci. Parasympathetic nervous system and heart failure pathophysiology and potential implications for therapy. *Circulation*, 118(8):863–871, 2008. doi: 10.1161/CIRCULATIONAHA.107.760405.

[68] K. Orth-Gomer, C. Hogstedt, L. Bodin, and B. Söderholm. Frequency of extrasystoles in healthy male employees. *British heart journal*, 55(3):259–264, 1986.

[69] Y.-H. Park. Differences of Heart Rate Variability during Sevoflurane Anesthesia in Children by Age. *Open Journal of Anesthesiology*, 02(03):74–78, 2012. doi: 10.4236/ojanes.2012.23018.

[70] C. K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H. E. Stanley, et al. Long-range correlations in nucleotide sequences. *Nature*, 356:168–170, 1992. doi: 10.1038/356168a0.

[71] C. K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos (Woodbury, N.Y.)*, 5(1):82–87, 1995.

[72] T. Penzel, J. W. Kantelhardt, L. Grote, J.-H. Peter, and A. Bunde. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE transactions on bio-medical engineering*, 50(10):1143–1151, Oct. 2003. doi: 10.1109/TBME.2003.817636.

[73] M. Pierzchalski. Assessing cardiovascular reaction to antiarrhythmic drugs using Higuchi's fractal dimension. In *Assessing cardiovascular reaction to antiarrhythmic drugs using Higuchi's fractal dimension*, pages 200–202, 2011.

[74] M. Pierzchalski, R. A. Stepien, and P. Stepien. New Nonlinear Methods of Heart Rate Variability Analysis in Diagnostics of Atrial Fibrillation. *International Journal of Biology and Biomedical Engineering*, 5(4):201–208, 2011.

[75] B. Qian and K. Rasheed. Hurst exponent and financial market predictability. In *Proceedings of The 2nd IASTED international conference on financial engineering and applications*, pages 203–209, 2004.

[76] D. Ramaekers, H. Ector, a. E. Aubert, a. Rubens, and F. Van de Werf. Heart rate and heart rate variability in healthy volunteers: is the female autonomic nervous system cardioprotective. *Eur Heart J*, 19(SEPTEMBER):1334–1341, 1998.

[77] S. R. Redwood, O. Odemuyiwa, K. Hnatkova, a. Staunton, I. Poloniecki, a. J. Camm, and M. Malik. Selection of dichotomy limits for multifactorial prediction of arrhythmic events and mortality in survivors of acute myocardial infarction. *European heart journal*, 18(8):1278–1287, 1997.

[78] a. Ripoli, L. Zyw, C. Passino, and M. Emdin. Linear and/or nonlinear behavior of heart rate? A comparative study using Lyapunov exponents, fractal dimension, conditional entropy and spectral analysis. *Computers in Cardiology, 2003*, (3):701–704, 2003. doi: 10.1109/CIC.2003.1291252.

[79] M. Rosenstein, J. Collins, and C. D. Luca. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1-2): 117–134, 1993.

[80] K. Shafqat, S. K. Pal, S. Kumari, and P. A. Kyriacou. HRV analysis in local anesthesia using Continuous Wavelet Transform (CWT). *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4808–4811, 2011.

[81] C. R. Shalizi and E. Micheli-Tzanakou. Methods and Techniques of Complex Systems Science: An Overview. *Complex Systems Science in Biomedicine*, pages 33–114, 2006. doi: 10.1007/978-0-387-33532-2.

[82] M. Signorini, M. Ferrario, M. Marchetti, and A. Marseglia. Nonlinear analysis of heart rate variability signal for the characterization of cardiac heart failure patients. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 3431–3434. IEEE, 2006.

[83] M. G. Signorini and S. Cerutti. Lyapunov Exponents Calculated from Rate Heart Variability Time Series. *Blood Pressure*, pages 119–120, 1994.

[84] P. K. Stein, R. E. Kleiger, and J. N. Rottman. Differing Effects of Age on Heart Rate Variability in Men and Women. *The American Journal of Cardiology*, 80(3):302–305, 1997. doi: 10.1016/S0002-9149(97)00350-0.

[85] P. K. Stein, P. P. Domitrovich, R. E. Kleiger, K. B. Schechtman, and J. N. Rottman. Clinical and demographic determinants of heart rate variability in patients post myocardial infarction: insights from the cardiac arrhythmia suppression trial (CAST). *Clinical cardiology*, 23(3):187–194, 2000.

[86] Task Force of the ESC and NASPE. Guidelines Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17:354–381, 1996.

[87] C. A. S. Trial, I. Investigators, et al. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med*, 327(4):227–33, 1992.

[88] G. Valenza, P. Allegrini, A. Lanatà, and E. P. Scilingo. Dominant Lyapunov exponent and approximate entropy in heart rate variability during emotional visual elicitation. *Frontiers in neuroengineering*, 5(3):1–7, Jan. 2012. doi: 10.3389/fneng.2012.00003.

[89] S. Vandeput, B. Verheyden, A. Aubert, and S. V. Huffel. Nonlinear heart rate variability in a healthy population: Influence of age. *2008 Computers in Cardiology*, 35:53–56, 2008. doi: 10.1109/CIC.2008.4748975.

[90] N. Vandewalle and M. Ausloos. Multi-affine analysis of typical currency exchange rates. *European Physical Journal B*, 4(2):257–261, 1998. doi: 10.1007/s100510050376.

[91] A. Voss, S. Schulz, R. Schroeder, M. Baumert, and P. Caminal. Methods derived from nonlinear dynamics for analysing heart rate variability. *Philosophical transactions.*

Series A, Mathematical, physical, and engineering sciences, 367(1887):277–296, 2009. doi: 10.1098/rsta.2008.0232.

[92] A. Wolf, J. Swift, H. Swinney, and J. Vastano. Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, 16(3):285–317, 1985.

[93] L. Xia and J. Jing. An Ensemble Density-based Clustering Method. *Proceedings on Intelligent Systems and Knowledge Engineering (ISKE2007)*, 2007. doi: 10.2991/iske. 2007.45.

[94] T. Yang and M. N. Levy. Effects of intense antecedent sympathetic stimulation on sympathetic neurotransmission in the heart. *Circulation research*, 72(1):137–144, 1993. doi: 10.1161/01.RES.72.1.137.

[95] R. Yeh, G. Chen, J. Shieh, and C. Kuo. Parameter investigation of detrended fluctuation analysis for short-term human heart rate variability. *Journal of Medical and Biological Engineering*, 30(5):277–282, 2010. doi: 10.5405/jmbe.30.5.02.

[96] R.-G. Yeh, J.-S. Shieh, G.-Y. Chen, and C.-D. Kuo. Detrended fluctuation analysis of short-term heart rate variability in late pregnant women. *Autonomic neuroscience : basic & clinical*, 150(1-2):122–126, 2009. doi: 10.1016/j.autneu.2009.05.241.

[97] X. Zeng, R. Eykholt, and R. Pielke. Estimating the Lyapunov-exponent spectrum from short time series of low precision. *Physical review letters*, 66(25):3229–3232, June 1991.