HEART WORK ANALYSIS BY MEANS OF RECURRENCE-BASED METHODS

Joanna IWANIEC¹, Marek IWANIEC²

AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics
Mickiewicz Alley 30, 30-059 Krakow, Poland
¹ Department of Robotics and Mechatronics, e-mail: jiwaniec@agh.edu.pl
² Department of Process Control, e-mail: iwaniec@agh.edu.pl

Abstract

Currently, for the purposes of recorded ECG signals (electrocardiograms) interpretation, the classical methods involving analysis of geometrical properties of the recorded waveforms in time domain are used. Such an analysis consists in determining the values of parameters describing the heart rate and rhythm. However, these indicators can not be treated as an infallible criterion for diagnosis and, moreover, the limits of increasing the accuracy of ECG analysis by increasing the accuracy of determining its characteristic points have already been reached. Therefore, in the paper, for the purposes of analysis of registered ECG signals and acoustical recordings of heart work, it is proposed to use the recurrence plots and RQA analysis methods that consist in searching for the recurrence properties of the registered signals. Application of the recurrence-based methods is natural due to the cyclic character of the heart work while providing patterns characteristic for different cardiac dysfunctions supported by objective, quantitative measures will contribute to early, credible and reliable classification of cardiovascular dysfunction.

Keywords: ECG signal, heart work, cardiovascular dysfunctions, recurrence plots, RQA analysis

1. INTRODUCTION

Accelerated progress of civilization (industrialization, urbanization and nutrition) has been leading to an increased duration and quality of life and, on the other hand, to new possibilities for adverse effects on human health. Modern science, through improved sanitation, vaccination and antibiotics as well as improved social and economical conditions, has eliminated the threat of death from most infectious diseases. However, changes in the diet, lifestyle and increasing environmental pollution have brought about diseases of civilization (called also 'lifestyle diseases'), including coronary heart disease, obesity, hypertension, type 2 diabetes, epithelial cell cancers, autoimmune disease and osteoporosis.

According to the World Health Organization more people die annually from cardiovascular diseases than from any other reason. An estimated 17.5 million people died in EU from cardiovascular diseases in 2012, representing 31% of all global deaths. Due to the health statistics provided by the American Heart Association in cooperation with the Centres for Disease Control and Prevention and the National Institutes of Health in 2015:

- cardiovascular disease is the leading global cause of death, accounting for 17.3 million deaths per year, a number that is expected to grow to more than 23.6 million by 2030,
in 2008, cardiovascular deaths represented 30% of all global deaths, with 80% of those deaths taking place in low- and middle-income countries,

- about 2,150 Americans die each day from these diseases, one every 40 seconds,
- cardiovascular diseases claim more lives than all forms of cancer combined,
- about 85.6 million Americans are living with some form of cardiovascular disease,
- direct and indirect costs of cardiovascular diseases and stroke exceed $320.1 billion, including health expenditures and lost productivity.

The problem of effective and credible cardiological diagnostics in Poland is also of key importance. According to the report of the National Institute of Public Health [19]: 'For years, the most significant threat to life are the cardiovascular diseases that, in 2010, were responsible for 46% of all deaths in Poland. Cardiovascular diseases are much more common cause of premature death in Poland than average in the EU'.

ECG (electrocardiogram) is a basic diagnostic tool so widespread that it is difficult to determine the exact number of these tests performed each day. Given that a million of Poles suffer from the diagnosed coronary artery disease and each year 100 000 myocardial infarctions are registered, it can be concluded that the number of ECGs performed during the year in Poland is expressed in millions.

While it is a relatively simple test to perform, the interpretation of the ECG tracing requires significant amount of training. Despite the widespread requirement for the efficient and reliable methods of ECG signals analysis and interpretation, in practice, almost exclusively, classical time-domain methods involving analysis of geometrical properties of the recorded waveforms are used. Many scientific publications relate to algorithms for analyzing ECG signals automatically, by searching for the specific points and segments. However, the limits of increasing the accuracy of classical ECG analysis by increasing the accuracy of finding its characteristic points have already been reached. Therefore, in the paper, it is proposed to use the ECG representation in the phase space and search for the recurrence properties of the registered signals. Change of domain in which the signal is analyzed offers new, so far unreachable, opportunities for early diagnosis. The proposed approach simplifies assessment of the signal information content by providing patterns characteristic for different cardiovascular dysfunctions supported by objective, quantitative measures provided by the RQA analysis.

2. RECURRING-BASED METHODS

Recurrence, in the sense of returning to the former states, is the fundamental property of many dynamical systems. The formal concept of recurrences was introduced by Henri Poincaré [21] in 1890. In 1987, based on the Poincaré’s achievements, Eckmann, Oliffson and Ruelle formulated the Recurrence Plots method (RP) [2]. Originally, the method was used for the purposes of visualisation of system trajectories. In the following years, development of the Recurrence Quantification Analysis (RQA) [15, 25], supporting interpretation of information contained in recurrence plots, has consolidated the RP method as a tool for nonlinear data analysis. Nowadays recurrence plots and their quantification – RQA are used for the purposes of detection of qualitative changes in the behaviour of dynamical systems [20]. Discusses methods find applications in numerous fields of research, such as technique [3], astrophysics [10], cardiology, neuroscience [24] and economy [5].

2.1. Recurrence plots (RPs) method

The main idea of the RP method [2-4, 6-8, 12, 14, 15] consists in revealing all the times when the phase space trajectory of the considered dynamical system visits roughly the same area in the phase space. Such recurrence of a state \( x \) at time \( i \) at a different time \( j \) is marked within a two-dimensional squared matrix \([R] \), called recurrence matrix, with ones and zeros:

\[
[R_{ij}] = \Theta(e - \|x_i - x_j\|) \quad i, j = 1, \ldots, N \quad (1)
\]

where \( N \) is the number of considered states \( x_\ell \), \( e \) denotes a threshold distance, \( \| \| \) a norm and \( \Theta() \) the Heaviside function.

Graphical representation of such a matrix is known in the literature under the term recurrence plot. In other words, the recurrence plot represents graphically the matrix \([R_{ij}] \) given as:

\[
[R_{ij}] = \begin{cases} 1 : & \{x_i \} \approx \{x_j \} \\ 0 : & \{x_i \} \neq \{x_j \} \end{cases} \quad i, j = 1, \ldots, N \quad (2)
\]

where \( \{x_i \} \approx \{x_j \} \) are the points belonging to the neighbourhood of radius \( e \) (defined according to the applied norm).

All the possible states of the system are represented by the points of the phase space. If the state of the considered system at time \( t \) can be determined by \( d \) components (e.g. position and velocity), these parameters form a state vector:

\[
[\mathbf{x}(t)] = [x_1(t) \ x_2(t) \ldots \ x_d(t)]^T
\]

in the \( d \)-dimensional phase space of the system. Vectors \([\mathbf{x}(t)] \) define trajectory in the phase space.
\text{In the experimental research, usually, not all the components necessary to construct the state vector can be measured. In such a case it is necessary to reconstruct the phase space on the basis of a time-discrete measurement of one variable:}
\[ u_t = u(i\Delta t), \quad i = 1, \ldots, N \]
where \( \Delta t \) denotes a sampling rate.

The most popular method of reconstruction is the time delay method:
\[ \{ x_i \} = \sum_{j=1}^{m} u_j + (j-1) \{ e_j \} \]
where \( m \) is the embedding dimension and \( \tau \) is the time delay.

In order to estimate recurrence plot (for the considered dynamical system) it is necessary to specify values of three parameters: threshold \( \epsilon \), time delay \( \tau \) and embedding dimension \( m \). Values of these parameters should be selected carefully, since they have a significant influence on the informative content of the estimated recurrence plot. For instance, if the value of \( \epsilon \) is too small, there may be almost no recurrence points and the information concerning recurrence of the considered system states can be lost. On the contrary, if \( \epsilon \) is chosen too large, almost every point is a neighbour of every other point, which leads to appearance of spurious structures visible on the obtained recurrence plot. In the literature various criterions for selection of threshold \( \epsilon \) have been proposed. In [16] it has been suggested that the threshold \( \epsilon \) value should equal a few per cent of the maximum phase space diameter while in [25] it has been proposed that it should not exceed 10% of the mean or the maximum phase space diameter. In case of stationary data it is possible to select \( \epsilon \) on the basis of the recurrence point density of the RP by seeking a scaling region in the recurrence point density [25]. On the other hand, for non-stationary data, it was proposed to select \( \epsilon \) such that the recurrence point density is approximately 1%. Another criterion, which holds for a wide class of processes, assumes that a measurement of a process is composed of the real signal and some observational noise with standard deviation [23]. In order to obtain the results similar as for the noise-free situation, \( \epsilon \) has to be five times larger than the standard deviation of the observational noise. For periodic and quasi-periodic processes, optimal threshold can be determined on the basis of density distribution of recurrence points along the diagonals.

Other approaches use a fixed recurrence point density [24]. On of the most commonly used criterions is the fixed amount of nearest neighbours (FAN) criterion, which consists in fixing the number of neighbours for each point of the trajectory. In this case, the threshold is different for each point of the trajectory. The advantage of the method is that it preserves the recurrence point density and allows to compare RPs of different systems without the necessity of normalising the time series beforehand.

Appropriate time delay \( \tau \) can be determined with the application of the auto-correlation or mutual information function [9, 13]. The mutual information function (MI) method, on the contrary to the linear autocorrelation function approach, takes into account also nonlinear correlations:
\[ MI = -\sum_{ij} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j} \]
where \( p_i \) denotes probability of finding the value of system time history in the \( i^{th} \) interval, \( p_{ij} \) denotes probability that observation in a given time instant \( t \) belongs to the \( i^{th} \) interval and observation in \( (t + \tau) \) time instant to the \( j^{th} \) interval. In practical applications, as the ‘reasonable’ value of \( \tau \), the value corresponding to the first minimum of the mutual information function is assumed.

For the purposes of the smallest sufficient embedding dimension \( m \) estimation, the false nearest neighbours method [1] is frequently used. The method determines neighbours of each point of the considered system phase space trajectory for a given dimension of a phase space \( m_n \) and checks whether these points are still the closest neighbours in the \( m_{n+1} \) dimensional phase space. For the properly selected dimension of the phase space, the number of false nearest neighbours is close to zero.

In the original definition of the method [2] the \( L_2 \) norm is used and the neighbourhood radius is selected in such a way that it contains a fixed amount of states. For such a neighbourhood, the radius \( \epsilon_i \) changes for each \( x_i \) \( (i = 1, \ldots, N) \) and \( R_{ij} \neq R_{ji} \), which leads to asymmetrical recurrence diagrams. This criterion can be adjusted in such a way that the recurrence point density has a fixed predetermined value and is called fixed amount of nearest neighbours (FAN).

2.2. Recurrence Quantification Analysis (RQA)

Recurrence Quantification Analysis (RQA) [15, 25] is a method of nonlinear data analysis which quantifies the number and duration of recurrences of a dynamical system represented by its state space trajectory. Definitions of the most popular RQA measures, such as: the Recurrence Rate (\( RR \)), Determinism (\( DET \)), Laminarity (\( LAM \)), Averaged diagonal line length (\( L \)), Trapping Time (\( TT \)), Longest diagonal line (\( L_{max} \)), Longest vertical line (\( V_{max} \)), Divergence (\( DIV \)), Entropy (\( ENTR \)) can be found in [20, 25].

The simplest measure of the RQA analysis is the recurrence rate (\( RR \)):
\[ RR(\epsilon) = \frac{1}{N} \sum_{i,j=1}^{N} R_{ij}(\epsilon) \]
defined as a ratio of all the recurrence states (points) to all the possible system states. Therefore
RR describes the probability that a state recurs in its $\varepsilon$-neighbourhood in phase space.

Definition of another measure – DET – is based on the observation that the processes with uncorrelated or weakly correlated, stochastic or chaotic behaviour cause none or very short diagonals, while deterministic processes result in longer diagonals. Therefore, the ratio of recurrence points forming diagonal lines to all recurrence points is treated as a measure for determinism (predictability) of the system:

$$DET = \frac{\sum_{l=1}^{N} P(l)}{N}$$

where $P(l)$ denotes a histogram of diagonal lines of the length $l$:

$$P(l) = \sum_{i,j=1}^{N} \delta[l - R_{i,j}(\varepsilon)]$$

(9)

A diagonal line of length $l$ means that a segment of the trajectory is rather close during $l$ time steps to another segment of the trajectory at a different time; thus these lines are related to the divergence of the trajectory segments. The average diagonal line length $L$:

$$L = \frac{\sum_{l=1}^{N} l \cdot P(l)}{\sum_{l=1}^{N} P(l)}$$

(10)

is the average time that two segments of the trajectory are close to each other.

The ratio between the recurrence points forming the vertical structures and all the recurrence points is called laminarity (LAM). A value of LAM decreases if the RP consists of more separated recurrence points than vertical structures:

$$LAM = \frac{\sum_{v=\min}^{N} v \cdot P(v)}{\sum_{v=\min}^{N} P(v)}$$

(11)

Trapping time $TT$ expresses the average length of vertical structures:

$$TT = \frac{\sum_{v=\min}^{N} v \cdot P(v)}{\sum_{v=\min}^{N} P(v)}$$

(12)

It can be interpreted as the time for which the system stays in a particular state.

3. RESULTS OF EXPERIMENTAL RESEARCH

In the course of research carried out over the last few years [6, 7, 8] it has been proven that the recurrence-based methods are particularly sensitive to changes in the dynamic behaviour of mechanical systems. In this paper, the methodology based on recurrence plots (RP) method was applied to analysis of the ECG signals of patients with various types of cardiac diseases available in the MIT-BIH Arrhythmia Database Directory [17].

The MIT-BIH Arrhythmia Database was created in 1980 as the result of cooperation between Massachusetts Institute of Technology and cardiologists from Boston’s Beth Israel Hospital (presently Beth Israel Deaconess Medical Center, Boston, MA). Recordings were obtained from 47 people: 22 women (aged 23 to 89 years) and 25 men (aged 32 to 89 years). Data was collected from both inpatients and outpatients in proportions close to 3:2. Each recording contains detailed description concerning ECG lead configuration type, patient’s sex, age, medications taken by a patient, beats annotations, ectopic beats annotations, heart rhythm types and their durations, quality of parts of the signal, additional notes and points of interest (time and description). Each signal has sampling frequency 360 Hz and lasts exactly 30 minutes and 5.556 seconds.

In the course of the research, the signals of interest were downloaded from the MIT-BIH Arrhythmia Database using the LightWAVE 0.63 tool, which is an open-source software for viewing ECGs and other physiologic waveforms with associated annotations (event markers) [22]. LightWAVE tool is provided by PhysioNet online resource centre, doesn’t require installation on the user’s computer and runs within any modern web browser. 6-second parts of the signals were analysed. RP and RQA analysis was predominantly performed for MLII leads.
of the false nearest neighbours algorithm. Appropriate time delay $\tau$ was determined by searching for the first minimum of the mutual information function. All the computations were performed using the CRP Toolbox for MATLAB [13]. The scheme of performed data processing and computations is presented in Fig. 2.

Selected results of the research are presented in the Fig. 3, Fig. 4, Table 1 - Table 3. Recurrence plots typical for a normal heart rhythm (a) and arrhythmias resulting from the presence of atrial pacemaker (b), right bundle branch block (c), left bundle branch block (d), atrial fibrillation (e) and ventricular flutter (f) are presented in the Fig. 3. Corresponding phase space trajectories are gathered in the Fig. 4. Results of the RQA analysis, performed for the considered ECG signals, are shown in Table 1, Table 2 and Table 3.

In case of recurrence plots determined for the healthy people with normal heart rhythm the number of diagonal lines corresponds to the number of QRS complexes on the ECG plot display. Their presence results from the deterministic character of heart work, which is reflected in the ECG recording (Fig. 3a). Measurement noise is visible in the form of separated points. In most cases, phase spaces of normal beats display strong directional character and resemble three perpendicular segments with a common beginning (Fig. 4a).

The ECG signal recorded from a patient with implanted cardiac pacemaker consists of three main components: the natural ECG signal, pacemaker pulses and noise. Therefore the pattern visible on the recurrence plots (Fig. 3b) differs significantly from the pattern obtained for the normal heart rhythm (Fig. 3a):

- many separated dots (clusters of points) are present along the diagonal lines (instead of a few lying close together),
- diagonal lines are thicker.

What’s more, for different patients, phase space trajectories take various shapes, which makes them one of the hardest to identify. In the course of the research, the shape of the phase space trajectory (Fig. 4b) corresponding to the recurrence plot presented in Fig. 3b was encountered most frequently.

From the medical point of view, the right bundle branch block is reflected in the electrocardiogram by:

- QRS complexes lasting more than 100 ms (in case of incomplete block) or more than 120 ms (in case of complete block) [26],
- presence of terminal $R$ wave in lead V1,
- presence of a slurred $S$ wave in leads I and V6.

Representation of the ECG signal characteristic for the right bundle branch block on the recurrence plot (Fig. 3c) is ambiguous and resembles samples
with normal beats. Corresponding phase space trajectory (Fig. 4c) possess distinctly rounded ends of perpendicular segment-formations (not sharp, like in case of other heart rhythms).

The presence of the left bundle branch block is manifested in the ECG by wide QRS complexes (beyond 0.12 s) and the change in their shape. This feature is reflected in the estimated recurrence plots (Fig. 3d) by wider distances between diagonal lines, less complicated structures with less bead formations and smoother diagonal lines (than in case of normal heart rhythm). Phase spaces (Fig. 5a) differ significantly for different patients (for V1 leads they are inverted with curly lines and open shape in the upper part).

Fig. 4. Phase space trajectories determined for: (a) normal beat, (b) atrial pacemaker, (c) right bundle branch block.

(a) 
(b) 
(c) 

Fig. 5. Phase space trajectories determined for: (a) left bundle branch block, (b) atrial fibrillation (c) ventricular flutter.

Atrial fibrillation is a chaotic atrial activity and, in comparison to e.g. atrial flutter, its behaviour is entirely unpredictable. Only slight or flat P waves are visible [26]. Obtained recurrence plots are jagged, with shorter and sometimes dashed diagonal lines (Fig. 3e). Bead or rectangular formations appearing along the diagonal lines are composed of many black and white points. Occasional white spaces are visible. Phase space trajectories are similar to the case of normal heartbeats. Sometimes singular lines, which slightly deviate from the path of the main trajectory, are present (Fig. 5b).

Ventricular flutter is a tachycardia, which affects patient’s ventricles. Frequency of appearance of beat formations suggests a very high
heart rate. Without an immediate medical help it may lead to the ventricular fibrillation and a sudden cardiac death. Obtained recurrence plots (Fig. 3f) are covered with numerous round and white areas, which proves sudden changes and differing mean value of the height of the peaks. Diagonal lines are discontinuous with no clear structures along diagonal lines. Strong fluctuations of the analysed signal result in a peculiar impression that the phase space trajectory (Fig. 5c) was computed on the basis of noisy samples.

In case of patients with various cardiac dysfunctions (Fig. 3 b + f) complexity ('degree of disorder') of phase space trajectories (Fig. 4 b + f) increases significantly. Obtained diagrams differ significantly among themselves and also from the diagram computed for the healthy patient. This regularity is reflected in the values of RQA measures (Tab. 1, Tab. 2, Tab. 3).

Results of ECG analysis obtained with the application of the RP method are very promising - for each considered cardiac disease the characteristic (reference) pattern was obtained while the values of RQA measures (DET, L, ENTR, LAM, TT) in most cases are significantly higher than for the healthy patient (denoted by †).

4. CONCLUSIONS AND FINAL REMARKS

For many years, by far the greatest threat to the health and life all over the World are diseases of the circulatory system. The basic and most efficient tool used in cardiovascular diagnostic is ECG, which is so widespread that it is difficult to determine the exact number of these tests performed each day. Despite the requirement for the efficient and reliable methods of ECG signals analysis and interpretation, in practice, almost exclusively, classical methods involving analysis of geometrical properties of the recorded waveforms in the time domain are used. The scientific works carried out recently have been focused on increasing the accuracy of ECG analysis by increasing the accuracy of finding its characteristic points. However, the accuracy limits of such a geometrical analysis have already been reached. Therefore the quest of formulating novel methods supporting classical analysis of ECG signals is still open.

Therefore, in the paper, it is proposed to use the ECG representation in the phase space and search for the recurrence properties of the registered signals. Change of domain in which the signal is analyzed offers new opportunities for early diagnosis. The proposed approach simplifies assessment of the signal information content by providing patterns characteristic for different cardiovascular dysfunctions supported by objective, quantitative measures provided by the RQA analysis.

In this paper, the methodology based on recurrence plots (RP) method was applied to analysis of the ECG signals of patients with various types of cardiac diseases available in the MIT-BIH Arrhythmia Database Directory. Results of ECG analysis obtained with the application of the RP method are very promising - for each considered cardiac disease the characteristic (reference) pattern was obtained while the values of RQA measures are significantly higher than for the healthy patient.

Conducted studies have to be continued and complemented by statistical analysis for different age groups and various degrees of severity of dysfunctions. However, at the present stage of the research, it can be concluded that the RP method supported by the RQA analysis can be successfully used for ECG rhythm analysis, searching of fragments of ECG signals characteristic for abnormal states, detecting states prior to ventricular
fibrillation and flutter. They can also find applications in monitoring patients after surgery (not only cardiological), and in interpreting the results of vectocardiography.

REFERENCES

18. Mygal V.P., But A.V., Mygal G.V, Klimenko I. A. An interdisciplinary approach to study individuality in biological and physical systems functioning. Scientific Reports 2016; 6. DOI:10.1038/srep29512.