ECG SIGNAL FEATURE EXTRACTION AND CLASSIFICATION BASED ON R PEAKS DETECTION IN THE PHASE SPACE

Olga Malgina [1], Jana Milenkovic [1], Emil Plesnik [2], Matej Zajc [2], Jurij F. Tasic [1, 2]

[1] Institute Jozef Stefan, Jamova cesta 39, 1000 Ljubljana
[2] University of Ljubljana, Faculty of Electrical Engineering, Trzaska cesta 25, 1000 Ljubljana

ABSTRACT

The goal of this paper is to present a novel approach in the automatic diagnosis of ECG abnormalities based on detection of R peaks in the phase space. The features are extracted from detected R peaks using their geometric position on the phase curve.
This paper is dealing with classification problem of normal and abnormal ECG signals. The proposed system has been validated with the data from the MIT-BIH database, in order to detect the cardiac arrhythmia. Support Vector Machine and K-Nearest Neighbour are used as classifiers. Results for both classifiers are similar. They are showing high accuracy in the experiment of classifying one test signal.

***Index Terms*—** ECG signal, phase space, classification, feature extraction, R peaks

1. INTRODUCTION

Computers measure more precisely than the human eye. Therefore automatic electrocardiogram (ECG) analysis is a fundamental task in cardiac monitoring and detection of cardiac abnormalities.
The R peak corresponds to the highest and sharpest position in the signal. This peak is the most important characteristic point in ECG and most of research on that field has been concentrated on analyzing it. Many of proposed methods for detection of R peaks were realized in the time domain [1-3]. Method for peak detection in phase space was chosen because it may explain the structure which is hidden in the signal dynamics and open new possibilities for feature extraction and classification.
The behavior of the ECG signal in the reconstructed phase space is used to determine features which in the best way characterize and distinguish signals. The classification of the ECG signals follows the previously extracted features. Fig. 1 presents the basics steps in the procedure for recognizing and classifying normal from abnormal heart beat rate.

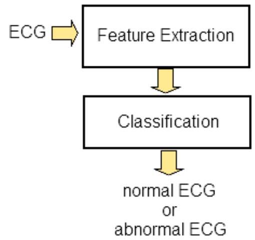

Figure 1. Basic steps in distinguishing different signals

The feasibility of the introduced solutions is demonstrated using ECG data from the public database MIT-BIH [4]. Our analysis is based on comparing two types of signals from Normal and Arrhythmia databases.
Proposed approach based on phase space reconstruction of signal can be easily implemented in portable ECG monitoring system for automatic analyzing and classification of signals.
The remainder of this paper is organized as follows. Section 2 gives a brief description of normal characteristics of the ECG signal. Section 3 provides phase space reconstruction and R peaks detection algorithm. In the last section, we present the experimental results with detailed discussion

2. NORMAL CHARACTERISTICS OF THE ECG

Every patient has his own unique ECG patterns. Nevertheless we try to use most common correspondences between characteristic points. The normal ECG is composed of the QRS complex, T and P waves. Fig. 2 depicts the basic shape of a healthy ECG heartbeat signal.

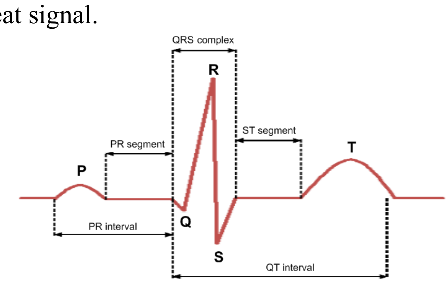

Figure 2. Normal ECG signal

The first electrical signal in the normal ECG is known as the P wave. It has positive polarity, and its duration is less than 120 milliseconds. Contraction (depolarization) of the ventricles results in usually the largest part of the ECG signal (because of the greater muscle mass in the ventricles) and it is known as the QRS complex, where:

- the Q wave is the first initial downward or “negative” deflection. It may or may not always be present,
- the R wave is always the first “positive” deflection,
- the S wave, the “negative” deflection, follows the R wave.

The T wave represents the relaxation (repolarization) of the ventricles. The rate of the normal sinus rhythm is 60 – 100 beats/min [5]. When the ECG abnormalities occur, standard features change their values, e.g. ST depression, T wave changes and so on [6].

3. ALGORITHM FOR R PEAKS DETECTION

3.1. Phase space reconstruction

The use of discrete derivative method [7 – 8] for reconstruction of the signal from time space leads to phase portrait of the signal in the phase space.

Let $s(k), k+1 \ldots N$, be the original signal in the time domain. Then for reconstruction of the ECG signal from the time domain into phase space the following formulas are going to be used:

$$x(k) = s(k), \\ y(k) = s(k+1) - s(k). \quad 1 \le k \le N-1 \tag{1}$$

The following normalization is performed:

$$x^*[k] = \frac{x[k] - \min x[k]}{\max x[k] - \min x[k]}, \forall k = 1,\ldots,N \\ y^*[k] = \frac{y[k] - \min y[k]}{\max y[k] - \min y[k]}, \forall k = 1,\ldots,N \tag{2}$$

As a result a new set of points $A = \{(x^*(n), y^*(n)), n = 1,\ldots, N\}$ is obtained. This set of points belong to the trajectory of the observed signal in the two-dimensional normalized phase space $x^* - y^*$.

3.2. R-peaks detection in the phase space

It is possible to extract R-peaks of ECG signals from the geometric structure of the curve in the phase space, also called phase portrait. Phase portraits of selected normal and abnormal ECG signals are presented in Fig. 3. It shows a significant set of loops of different sizes. R wave corresponds to the larger loop with R peak on the top of it.

Let $M = N - 1$ be the number of points on the curve, (x,y) coordinate of a point and $i \in \{1,\ldots,M\}$ index of a point. The following two conditions for R-peaks detection have to be satisfied:

$$1) \begin{cases} x(i-1) - x(i-2) > 0, \\ y(i-1) - y(i-1) > 0, \\ x(i+2) - x(i+1) < 0, \\ y(i+2) - y(i+1) > 0. \end{cases} \tag{3}$$

2) $x(i)$ and $y(i)$ have to be more then 0.6 for each $i \in \{1,\ldots,M\}$.

If previous conditions are satisfied for an $i \in \{1,\ldots,M\}$, than the R point will be determined as: $(x(i), y(i)) = R(i)$.

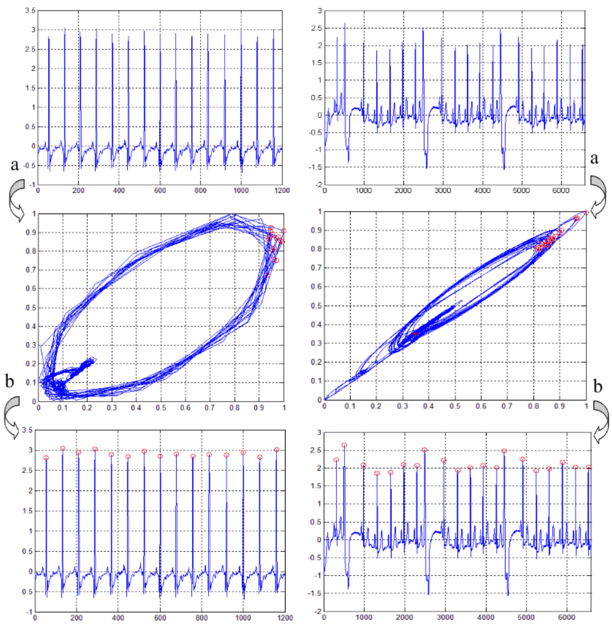

Figure 3. Signal 119 from MIT-BIH Arrhythmia database on the right and 16265 from MIT-BIH Normal Sinus Rhythm database on the left. a) reconstruction of the signal from time domain into the phase space, b) visualization in the time domain of detected R-peaks in the phase space.

Using the R peaks detection algorithm R-R duration is detected. Differences between normal and abnormal signals (arbitrary signals) can be seen in the Table 1.

Table 1. R-R intervals for normal and abnormal signals

	R-R intervals							
	1-2	**2-3**	**3-4**	**4-5**	**5-6**	**6-7**	**7-8**	**8-9**
Signal 16265 normal	96	100	96	90	89	94	109	120
Signal 119 abnormal	194	474	338	336	320	323	194	468

4. RESULTS

4.1. Feature extraction

The behavior of the ECG signal in the reconstructed phase space is used to determine all needed features for successful classification.
Features extracted from the reconstructed phase space depend on the geometric position of R-peaks on the phase curve.
A support point K calculated in [9] was determined for feature extraction on the phase curve. Example of a phase curve with marked features (length, angular) and points K and R is shown in Fig. 4.

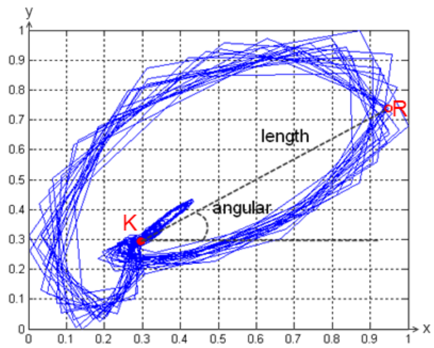

Figure 4. Phase space of a normal ECG signal with determined features

After detecting R peaks of a signal in the phase space, two features are extracted due to their good approximation of R peak characteristics. The length between the mentioned fix point K and the R peak is taken as a first feature. The angle between horizontal direction going through point K and direction that connect K and R peak is specified as the second feature.

4.2. Classification

The classification algorithm is based on R-peaks detection in the reconstructed phase space.
Six signals were selected for classification analysis. Three of them are representing normal and three of them abnormal signals.
Measurements of two selected features for three normal and three abnormal signals are presented in Table 2 and Table 3.
The time period used in our experiment was 4000 samples for both types of signals. In that time window the number of detected R peaks was 40, 49, 49 for normal and 13, 13, 12 for abnormal signals. Difference in the number of detected R peaks between both types of signals occurs due to different sampling frequencies of normal (128 Hz) and abnormal (360 Hz) signals.
Data from both tables are used as the training set and plotted in two dimensional feature space (angular versus

Table 2. Extracted features of normal signals

		Signal 16539 normal	Signal 16265 normal	Signal 16420 normal	Signal 16786 normal
Number of detected R peaks		40	49	49	37
Length	Average	0,7831	1,0395	0,6048	1,0507
	Standard deviation	0,0622	0,0781	0,0741	0,0829
Angular	Average	33,96	39,82	32,05	37,27
	Standard deviation	5,78	3,26	6,47	5,11

Table 3. Extracted features of abnormal signals

		Signal 101 abnormal	Signal 103 abnormal	Signal 113 abnormal	Signal 119 abnormal
Number of detected R peaks		13	13	11	12
Length	Average	0,9656	1,1318	1,0140	0,7158
	Standard deviation	0,0927	0,0410	0,1786	0,0825
Angular	Average	44,00	43,85	44,21	44,00
	Standard deviation	1,15	0,69	0,65	0,60

length) in Figure 5. Normal signals are represented with blue signs while abnormal with red signs.
To evaluate the performance of the presented algorithm for distinguishing normal from abnormal signals, two classifiers were used: K-Nearest Neighbour and Support vector Machine [10 – 12]. For the testing purpose three abnormal signals (signals 105, 122 and 202) were chosen and plotted with green signs. Result for the K-Nearest Neighbour method depends from the selected parameter K. For the test signal 105 using K=1 and K=3 efficiency is 86.67 percent. From the other side, for K=5 complete accuracy is achieved. The slightly higher efficiency of K-NN classifier is achieved for the test signal 122 using K=1 and K=3, where accuracy is 94.11 percent while for K=5 accuracy is maintained to 100 percent. The smallest accuracy of K-NN is obtained using test signal 202. For the previously mentioned signal accuracy of the classification is 22.22 percent for K=1, 55.55 percent for K=3 and 66.66 percent for K=5. From the other side, completely accuracy is achieved using SVM for all three aforementioned signals.

5. CONCLUSION

Phase space reconstruction proved as a reliable method for detecting specific points in ECG signals. Two features were obtained from the phase space by using the position of detected R peaks in ECG signals in association with a support point K. The extracted features were used in classification with two methods: K-Nearest Neighbour and Support Vector Machine. Results for both classifiers are highly accurate with selected test signal.

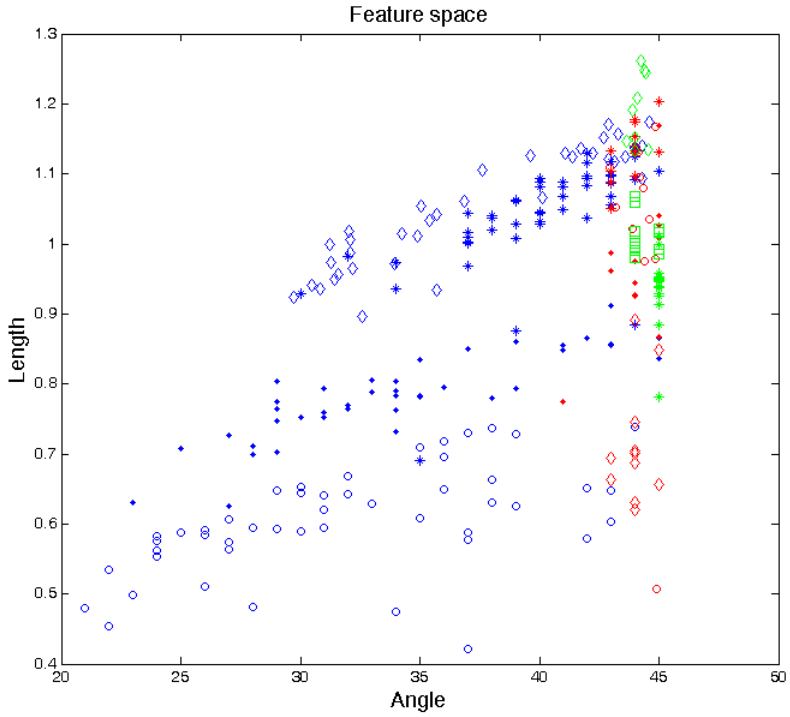

Figure 5. Training and test data represented in the feature space

This indicates that the proposed method for feature extraction can be used for distinguishing between normal and abnormal ECG signals. However, the method is not appropriate when abnormality amplitudes expand the specified threshold for R peak detection. The same problem occurs also in the time domain. The robustness of the method is going to increase with introduction of other specific points of the ECG signal.

Several other characteristic points of the ECG signal (such as P, Q, S and T point) can also be used as classification features in order of getting better results, which is a great motive for further research.

6. REFERENCES

[1] H. S. Shina, C. Leea and M. Lee, "Adaptive threshold method for the peak detection of photoplethysmographic waveform," *Computers in Biology and Medicine,* vol. 39, no. 12, pp. 1145 – 1152, Dec. 2009.

[2] G. de Lannoy, A. de Decker and M. Verleysen, "A Supervised Wavelet Transform Algorithm for R Spike Detection in Noisy ECGs," in *Proc. BIOSTEC 2008,* CCIS 25, pp. 256–264, 2008.

[3] Jing-tian Tang, Xiao-li Yang, Jun-chao Xu, Yan Tang, Qing Zou, Xiao-kai Zhang, "The Algorithm of R Peak Detection in ECG Based on Empirical Mode Decomposition," icnc, vol. 5, pp.624-627, 2008 Fourth International Conference on Natural Computation, 2008.

[4] MIT-BIH Database, http://physionet.ph.biu.ac.il/physiobank/database/mitdb, April 2010.

[5] D. H. Spodick, "Normal sinus heart rate: Sinus tachycardia and sinus bradycardia redefined," *Am Heart J,* vol. 124, no. 4, 1119–21, Oct. 1992.

[6] L. Sörnmo, P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications,* Elsevier Academic Press, Oxford, UK, 2005.

[7] L.S. Fainzilberg, "ECG Averaging based on Hausdorff Metric," *International Journal of Biomagnetism*, vol. 5, no. 1, pp. 236 – 237, 2003.

[8] L.S. Fainzilberg, "Heart functional state diagnostic using pattern recognition of phase space ECG-images," in *Proc. of the 6th European Congress on Intelligent Techniques and Soft Computing (EUFIT '98),* Germany, 1998, vol. 3, pp. 1878-1882.

[9] L.S. Fainzilberg, *Information technology for signal processing of complex shape. Theory and Practice*, Naukova Dumka, Kiev, 2008.

[10] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, Second Edition, Wiley, 2000.

[11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Lesrning Methods*, Cambridge University Press, 2000.

[12] C. j.c. Burges, "A tutorial on Support Vector Machine for pattern recognition," Data Mining and Knowledge Discovery vol. 2, no. 2, pp. 121-167, 1998.