Rhythm Analysis by Heartbeat Classification in the Electrocardiogram
(Review article of the research achievements of the members of the Centre of Biomedical Engineering, Bulgarian Academy of Sciences)

Ivaylo Christov*, Irena Jekova, Vessela Krasteva, Ivan Dotsinsky, Todor Stoyanov

Centre of Biomedical Engineering – Bulgarian Academy of Sciences,
105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
E-mail: ivaylo.christov@clbme.bas.bg

* Corresponding author

Received: June 16, 2009 Accepted: July 1, 2009

Published: August 5, 2009

Abstract: The morphological and rhythm analysis of the electrocardiogram (ECG) is based on ventricular beats detection, wave parameters measurement, as amplitudes, widths, polarities, intervals and relations between them, and a subsequent classification supporting the diagnostic process. Number of algorithms for detection and classification of the QRS complexes have been developed by researchers in the Centre of Biomedical Engineering – Bulgarian Academy of Sciences, and are reviewed in this material. Combined criteria have been introduced dealing with the QRS areas and amplitudes, the waveshapes evaluated by steep slopes and sharp peaks, vectorcardiographic (VCG) loop descriptors, RR intervals irregularities. Algorithms have been designed for application on a single ECG lead, a synthesized lead derived by multichannel synchronous recordings, or simultaneous multilead analysis. Some approaches are based on templates matching, cross-correlation or rely on a continuous updating of adaptive thresholds. Various beat classification methods have been designed involving discriminant analysis, the K-th nearest neighbors, fuzzy sets, genetic algorithms, neural networks, etc. The efficiency of the developed methods has been assessed using internationally recognized arrhythmia ECG databases with annotated beats and rhythm disturbances. In general, high values for specificity and sensitivity competitive to those reported in the literature have been achieved.

Keywords: Electrocardiography, QRS detection, Rhythm analysis, Automatic beat classification, Morphological parameters, Time-frequency analysis, Template matching, Karhunen-Loève transform, Independent component analysis, K-th nearest neighbors.

Introduction
The ventricular contractions and the depolarization phenomenon are identified in the electrocardiogram (ECG) by characteristic high-amplitude waves, named QRS complexes. Distances between them (RR intervals) define the rhythm, which is strongly influenced by the emotions and the physical activity and less in line by the respiratory act. In heart conduction disorders, ventricular excitation may not originate as it is normal from the sinus node, but from other ectopic centers in the myocardium. Thus premature contractions are generated, called also extrasystoles or ectopic beats. Typically, they are recognized by the irregular coupling RR intervals. The premature atrial contractions (PACs) produce normally shaped QRS complexes, while the premature ventricular contractions (PVCs) are generating a variety of QRS waveforms, quite differing from the normal ones. The premature beat itself does not cause symptoms but the occurrence of multiple single premature beats is considered clinically important, since it is a sign for disturbance in the depolarization process preceding in many cases the appearance of malignant cardiac arrhythmias.
The automatic detection and classification of the ventricular contractions as normal or premature is a subject of long-term studies. This is the basis of the rhythm analysis, which is usually applied to continuous 24-hour ECG recordings (Holter systems) and monitors in surgical and intensive care rooms to identify rhythm disorders.

Recognition of premature contractions and their classification as PACs or PVCs is requires a reliable detection of the QRS complexes [3, 4, 9, 10]. During the QRS detection process, some extrasystoles may be initially identified as normal QRS complexes and consecutively separated by means of additional criteria.

Two main strategies are set apart in the development of automatic beat classification methods in consideration to their application: (i) fast real-time or pseudo-real time implementations and (ii) complex offline classifiers aiming at higher accuracy and reliability of the clustering results in more classification groups.

**Automatic classification in real or pseudo-real time**

Algorithm for PAC and PVC recognition has been developed by Christov [5], working in pseudo-real time. Two different areas are calculated for each beat within a window defined around the largest positive and the largest negative peaks, and the smaller area is selected. Adaptive thresholding is applied, depending on the difference in areas and the difference between the current RR interval and the previous regular RR interval. PVC is detected when the difference between the area of the analyzed beat and the mean area of the five preceding normal QRS complexes, stored in a reference buffer, exceeds a threshold value. PAC is discovered by comparison of the two coupling RR intervals around the current complex with the mean RR-interval calculated as an average of the five preceding regular RR intervals. If a normal ventricular contraction is recognized, the QRS complex is used to update the reference buffer. High values for sensitivity, specificity and positive prediction accuracy, respectively 95.54%, 99.66% and 96.55%, have been reported for this method.

Dotsinsky and Christov [11] have clustered ventricular contractions by delineation of QRS onsets and offsets, supported by an algorithm for rejection of high amplitude T-waves, and some untypical P-waves. The algorithm examines the slopes of the ventricular contractions. The QRS onsets are defined by two alternative criteria that follow two consecutive differences of samples, standing 20 ms apart from each other (one period of the 50 Hz powerline interference). The first criterion requires the magnitude of the two equal-signed differences to exceed a threshold of $L = 280 \, \mu V$. In order to cancel false threshold activation by sharp peaks, it is additionally required that the absolute value of the difference between the candidate central sample and the last detected offset is $< L$. The second criterion provides recognition of extrasystoles of low slew rates but of high amplitudes if the absolute value of the difference between the candidate and the last detected QRS offset is $> 2L$ and if the candidate is $< 240ms$ apart from the previous QRS offset.

The QRS offset is marked if the absolute difference between the candidate and the last onset $< L$ and all samples within the next 20 ms do not differ considerably in amplitude (the absolute values of the successive intersample differences are $< L/4$). Fig. 1 illustrates two ECG signals with normal QRS and extrasystoles for which the onsets and offsets are measured (circle ‘o’ marks) and peaks of the QRS complexes are recognized (asterisk ‘*’ marks).
Fig. 1 Delineation of onsets, offsets (marked with circles), and peaks (marked with asterisks) of the QRS complexes. On the left – 1st, 2nd and 4th complexes are normal and the 3rd is extrasystole. On the right – 1st and 3rd are extrasystoles and the 2nd is normal QRS.

Fig. 2 Examples of different cardiac arrhythmias with correct (a-d) and erroneous (e-h) detection of A-beats according to the method in [19]. Detection labels are marked at the top of each subplot, while the annotation labels are shown between the two ECG leads.

The resolution on the x-axis is 50mm/s and on the y-axis is 20mm/mV.
(a) Arrhythmia with premature contractions excited from ectopic centers both in the atria and in the ventricles; (b) Atrial fibrillation; (c) - Atrial Tachycardia; (d) - Blocked rhythm: A-beats between 2 pathologic pauses; (e) - Arrhythmia with polymorphic ventricular contractions; (f) - arrhythmia with atrial PBs; (g) - Atrial fibrillation at rapid heart rate; (h) - Arrhythmia with interpolated V-beat.
The method by Dotsinsky and Stoyanov [9] applies searching for a PVC starting 120 ms after the previously detected QRS complex. The basic criterion is the presence of a biphasic wave, which exceeds predefined thresholds for durations and amplitudes of the two phases. If no candidate for extrasystole is found, the threshold requirements are reduced and a second verification is applied. Meeting the reduced criteria clusters the biphasic wave as a ‘suspicious’ candidate for an extrasystole. A second branch of the algorithm verifies whether the candidate is located between two normal QRS complexes, < 1.3 s apart from each other. A third branch compares the parameters of the currently detected QRS complex to a QRS template selected initially as normal QRS. The parameters are: maximal positive and negative amplitudes, peak-to-peak amplitude, and the number of samples exceeding the defined above threshold. The candidate is clustered as extrasystole in case of significant difference between the parameters.

Krasteva et al [19] have suggested a method for detection of single PACs, or PACs propagating during supraventricular tachycardias, postoperative or paroxysmal atrial fibrillations. The PACs detection algorithm applies heartbeat classification by estimation of the QRS waveform morphology and RR-intervals irregularity in two-channel ECGs. The estimator of the RR-intervals is adopted from [5] taking the difference between the two surrounding RR intervals, normalized to the mean value of the five preceding RR intervals. Three other parameters estimate the QRS waveform morphology by comparing the current heartbeat to a reference QRS pattern calculated after averaging of the five preceding beats. The suggested parameters evaluate: (i) the difference in QRS widths; (ii) the difference in QRS areas; (iii) the difference between the amplitudes of the maximal QRS-loop vector in the vectorcardiographic (VCG) plane. The developed method applies a sequence of tests to search for different types of atrial arrhythmias, including test for absolute arrhythmia, for supraventricular tachycardia, for pathologic pause, for single premature heartbeat, including additional verification for PAC. The proposed decision-tree classifier clusters the heartbeats as normal, PACs and PVCs, marked respectively by ‘N’, ‘A’ and ‘V’ in Fig. 2. The testing of the algorithm with the publicly available MIT-BIH arrhythmia database presented a relatively high accuracy with sensitivity of 92.2% and specificity of 96%.

Krasteva and Jekova [20] have proposed a method for discrimination of PVCs from normal beats, PACs and paced beats (PBs) by analysis of single channel ECG with resource efficient algorithms. The method requires minimal expert annotation of a few normal ventricular complexes, accepted as static QRS patterns. During the ECG monitoring, self-learning of the method is achieved by permanent update of dynamical QRS patterns to capture slight

![Fig. 3 Heartbeat waveform descriptors evaluated in [20](based on area difference (on the left), frequency spectrum difference (on the right-top), and maximal cross-correlation (right-bottom).](image)
variations in the heartbeat waveforms of the patient’s sustained rhythm. The method is based on matching of the evaluated heartbeat with the QRS templates by a complex set of ECG descriptors, including maximal cross-correlation, area difference, and frequency spectrum difference. The example in Fig. 3 illustrates the potential of the three descriptors to distinguish a normal QRS from PVC when they both are compared to the QRS template. Temporal features are also evaluated by analyzing the R-R intervals. Fuzzy classification rule is integrated. Analysis of all recordings in MIT-BIH and the MIT-SVDB databases show high values for sensitivity (98.4%) and specificity (98.9%) of the developed method.

Iliev et al [15] have suggested a very fast software technique for real-time detection of pathological cardiac events in ECG designed especially for event/alarm recorders carried by high-risk cardiac patients. The method implements simple QRS detection by amplitude and slopes thresholding (Fig. 4 – left plots), and QRS waveform evaluation by 64x32 histogram matrix, which accumulates dynamically the amplitude-temporal distribution of the successive heartbeat waveforms. Fig. 4 (right plot) depicts such a matrix, where normal and PAC beats are superimposed within the dark area of repeating waveforms, while PVC beats fall outside this area. The evaluation of the heartbeats by this histogram matrix provides high rating for normal and PAC beats but low rating for PVC beats. Simple decision-tree classifier is implemented. The performance of the method is tested with AHA, MIT-BIH and the European ST-T Databases. The specificity and the sensitivity are reported to be about 99.5% and 95.7% for all databases and about 99.81% and 98.87% for the noise free dataset.

Jekova and Krasteva [16] have investigated the projection of the cardiac electrical vector on the VCG plane formed by two non-orthogonal Holter chest leads, searching for significant differences between the QRS loop dispositions of normal and PVC beats. Fast calculations are aimed during the analysis of the spatial correlation of two QRS loops. They are achieved by approximation of the QRS loop in the VCG matrix space of 100x100 elements, each element taking a value of 0 or 1, assigned by a fast algorithm for verification of whether the element is external or internal to the QRS loop (Fig. 5). Thus the complexity of the calculations is reduced to operations with binary matrixes for assessment of the spatial displacement of the tested QRS loop area to a reference QRS loop area. In addition to the VCG analysis, which considered alone does not contain information about the temporal ECG characteristics, a parameter for assessment of the interbeat RR interval differences independent from the momentous heartrate is introduced. The reliability of the proposed parameter set for clustering
of normal and PVC beats was estimated by two classification methods – a stepwise discriminant analysis and a decision-tree-like classification algorithm, both tested with the MIT-BIH arrhythmia database. The accuracy achieved for the stepwise discriminant analysis is sensitivity of 91% and specificity of 95.6%. Comparable results are also achieved with the simpler decision-tree-like technique, including sensitivity of 93.3% and specificity of 94.6%.

Fig. 5 Illustration of 2-lead ECG and the QRS loop approximation in the VCG matrix plane. (top plot): Extraction of N and PVC beats in a fixed size window around the fiducial point. (bottom plots): The QRS loops and their approximated areas for the beats with indexes from T-5 to T-1. The black elements are assigned with a value of 1, the white elements are assigned with a value of 0, used in the calculations of the QRS loop areas. The last subplot illustrates the calculated reference VCG Matrix, which summarizes the QRS loop spatial distributions of the five consecutive QRS complexes.

A review article of real or pseudo real-time methods and algorithms for detection of extrasystoles has been presented by Dotsinsky et al [4].
Advanced algorithms for automatic heartbeat classification
The goal of the advanced algorithms for automatic heartbeat classification is higher accuracy and reliability of the clustering results in more classification groups. They are implemented in systems for which working in real or pseudo-real time is not imperative.

Three major trends in the advanced algorithms for automatic heartbeat classification were pursued during the last decade: (i) evaluation of reliable heartbeat descriptors from the ECG or VCG; (ii) application of different classification methods; (iii) comparative studies of descriptors, classification methods and learning datasets of different sizes.

Descriptors of the ventricular contractions
Christov and Bortolan [6] have defined and measured 26 morphological parameters, including: width of the QRS complex, its positive and negative peak amplitudes, positive and negative areas, slew-rate of different segments, magnitude and angle of the main VCG vector, etc (Fig. 6) The parameters are measured for all QRS complexes annotated as ‘normals’ (N) and PVCs from the 48 ECG recordings of the MIT-BIH arrhythmia database. Neural networks (NN) were used for the analysis of the large quantities of descriptors. Separate ranking of any descriptor and homogeneous group ranking (amplitude, area, interval, slope and VCG vector) were performed. From the two ECG leads, the first three ranked parameter groups for clustering of PVCs are QRS amplitude, slope and interval, while for N clustering they are VCG vector, QRS amplitude and area.

Feasibility of the Karhunen-Loève transform (KLT) for detection of ventricular ectopic beats is studied by Gómez-Herrero et al [13]. The KLT basis functions are derived for a small-sized training set of normal QRS complexes to extract their major components. The relevant KLT features are obtained by comparison between five selected heartbeats of the predominant rhythm and all other heartbeats in the tested ECG recording. Statistical analysis of the KLT features for MIT-BIH arrhythmia database contributes to the definition of threshold criteria for discrimination between the predominant and the ventricular ectopic beats. The achieved accuracy is about 97.7% for single-lead analysis and 98.3% for joint two-lead processing. The method is attractive and suitable for implementation in an automatic analysis module because of the necessity for supervisor annotation of only five beats of the predominant rhythm in one ECG recording.
Gómez-Herrero et al [14] have proposed a method based on the Matching Pursuits algorithm for selecting time-frequency descriptors that can be used for heartbeat classification. The clustered groups are not only ‘normal’ (N) and PVC, but also left bundle branch blocks (LBBB), right bundle branch block (RBBB), and paced beats (PB). The authors have investigated the usefulness of Independent Component Analysis for extracting additional spatial features from multichannel ECG recordings. The computing resources required by the proposed system are high during the training of the feature extractors. However, once the system has been trained, the extraction of the time-frequency features only requires the projection of the new beats into the selected wavelet packets atoms.

Classification methods for clustering of the ventricular contractions
The potential of the 26 morphological parameters as defined and measured by Christov and Bortolan [6], has been tested to discriminate all N and PVC beats in MIT-BIH arrhythmia database, applying the following classification methods:
- Neural networks (NN) [6]. Very good clustering of the two heartbeat groups is achieved using all 26 descriptors. The NN classifier can be quite simplified (with some small compromise towards the accuracy), by decreasing the number of the descriptors;
- K-th nearest neighbor rule – Christov et al [7];
- Discriminant analysis – Jekova et al [17]. Considering the two available ECG leads, 7 parameters with the highest discriminant power for N and PVC has been extracted.
- Hyperbox (HB) or hyperellipsoid classifier [1]. In order to characterize and to search in the feature space for the optimal HB that better characterize the considered class, different learning processes have been developed with the use and the combination of fuzzy clustering and genetic algorithms. Example of hyperellipsoids classifier is shown in Fig. 7.

![Fig. 7 Example of classification with two hyperellipsoids (PVC: red dots; N: blue ‘+’ marks)](image)

Comparative studies of descriptors, classification methods and learning atasets of different sizes
Christov et al [8] have presented a comparative study of the heartbeat classification abilities of two techniques for extraction of heartbeat features from the ECG: (i) QRS pattern recognition method for computation of a large collection of morphological QRS descriptors
and (ii) Matching Pursuits algorithm for calculation of expansion coefficients, which represent the time-frequency correlation of the heartbeats with extracted learning basic waveforms (Fig. 8). The K-th nearest neighbor classification rule [7] has been applied for assessment of the performances of the two ECG feature sets with the MIT-BIH arrhythmia database for QRS classification in five heartbeat classes: N, PVC, LBBB, RBBB, and PB. Five learning datasets are explored: one general learning set (GLS, containing 424 heartbeats), and four local sets, respectively GLS + about 0.5, 3, 6, 12 min from the beginning of the ECG recording. The achieved accuracies by the two methods are sufficiently high and do not show significant differences. The optimal size of the learning set is found to be about 3 min for which sensitivity between 94.8% and 99.9%, and specificity between 98.6% and 99.9%, are reported. The repeating waveforms, like N, RBBB, LBBB, and PB are better classified by the Matching Pursuits time-frequency descriptors, while the wide variety of bizarre PVCs are better recognized by the morphological descriptors.

Fig. 8 Five classes of heartbeats (on the right-bottom) analyzed by the Matching Pursuits algorithm in [8]. Time-frequency support (a-e on the top) and time-domain representation (a-e on the bottom) of the ten top-ranked time-frequency atoms from the Symmlet 8 wavelet packet, which were selected by the Matching Pursuits algorithm to best correlate with the signal structures of each heartbeat class in GLS.

The learning capacity and the classification ability for N and PVC clustering by four classification methods have been compared by Bortolan et al [2]. The classifiers are: neural networks (NN), K-th nearest neighbour rule (Knn), discriminant analysis (DA) and fuzzy logic (FL). The descriptors are the 26 morphological parameters [6]. One global and two local learning sets are tested. Better accuracy is reported for the local classifiers because of their good adaptation to the patient rhythm, while the capacity of the global classifier to process
new records without additional learning is expectedly balanced by lower accuracies. NN assure the best results (high and balanced indices for specificity and sensitivity) using one of the local learning set, while the Knn provides the best results with the other local learning set. Using the global learning set, DA and the FL methods perform better than the NN and Knn.

As an addition to the work of Bortolan et al [2], Jekova et al [18] presented comparative study of the learning capacity of the four classification methods – Knn, NN, DA, and FL using the 26 morphologic parameters [6]. One global, one basic and two local learning sets have been used. A small-sized learning set, containing the five types of QRS complexes collected from all patients in the MIT-BIH database, was used either with or without applying the leave one out rule, thus representing the global and the basic learning set, respectively. The local learning sets consist of heartbeats only from the tested patient, which are taken either consecutively or randomly. Using the local learning sets the assessed methods achieve high accuracies, while the small size of the basic learning set is balanced by reduced classification ability. Expectedly, the worst results have been obtained with the global learning set.

Acknowledgements

This article is done as a review of the research achievements of the members of the Centre of Biomedical Engineering – Bulgarian Academy of Sciences during the years. We acknowledge all colleagues for the partnership in some of the studies cited in the reference list for the successful collaboration and assistance:

- Senior Researcher Giovanni Bortolan – ISIB-CNR, Padova, Italy;
- Prof. Karen Egiazarian, head of department Transforms and Spectral Techniques Group, Tampere International Centre for Signal Processing, Finland;
- Senior Researcher Atanas Gotchev – Tampere International Centre for Signal Processing, Finland;
- Researcher, PhD student German Gómez-Herrero – Tampere International Centre for Signal Processing, Finland;
- Prof. Witold Pedrycz – chairman of Canada Research Chair, Department of Electrical and Computer Engineering;
- Assoc. Prof Ivo Iliiev – Technical University, Sofia;
- Senior Assistant Serafim Tabakov – Technical University, Sofia.

References

Prof. Ivaylo Christov, D.Sci., Ph.D.
E-mail: ivaylo.christov@clbme.bas.bg

Ivaylo Christov has graduated at the Technical University – Sofia, Faculty of Electronic Engineering and Technology, specialization Medical and Nuclear Instrumentation in 1978. He has been working at the Central laboratory of Biomedical Instrumentation – Medical Academy, Sofia, during 1979-1994. Since 1994, he is part of the team of the Centre of Biomedical Engineering – Bulgarian Academy of Science. His current positions are: deputy director and head of department Analysis of Processing of Biomedical signals and Data. He defended his Ph.D. thesis in 1989, and a D.Sci. thesis in 2006. He is Assoc. Prof. since 1996, and Prof since 2007.

Assoc. Prof. Irena Jekova, Ph.D.
E-mail: irena@clbme.bas.bg

Irena Jekova graduated Technical University – Sofia, Faculty of Electronic Engineering and Technology, specialization Electronic Medical Equipment in 1998. She has been working in the Centre of Biomedical Engineering, Bulgarian Academy of Sciences since 1999. She defended her Ph.D. thesis in the field of ventricular fibrillation detection in 2001. In 2007 she became Associate Professor. I. Jekova was awarded as a young scientist by the Bulgarian Academy of Sciences (2002) and by the Union of Scientists in Bulgaria (2007) for her outstanding scientific achievements in the development of methods and algorithms for ECG signal processing. Her scientific interests are in the field of biomedical signals and data processing, ECG analysis and arrhythmia classification.

Assoc. Prof. Vessela Krasteva, Ph.D.
E-mail: vessika@clbme.bas.bg

Vessela Krasteva graduated Technical University – Sofia, Faculty of Electronic Engineering and Technology, specialization Electronic Medical Equipment in 1998. Since 1999, she is working in the Centre of Biomedical Engineering, Bulgarian Academy of Sciences. In 2001 she held Ph.D. degree in the field of defibrillation. In 2007 she became Associate Professor. V. Krasteva has been honored with a number of awards for Outstanding Scientific Contributions for Young Scientists by the Ministry of Education and Science (2006), Bulgarian Academy of Sciences (2006), Union of Scientists in Bulgaria (2007). Her general interests are focused on development of models, methods and algorithms in the field of biomedical signal processing and electrical therapy, especially with applications in ECG instrumentation and defibrillation.
Prof. Ivan Dotsinsky, D.Sci., Ph.D.
E-mail: iadoc@bas.bg

Ivan Dotsinsky obtained his M.Sc. degree from the Faculty of Electrical Engineering, Technical University of Sofia. His PhD thesis was on the statistical assessment of the reliability of electrical and electronic circuitry. In 1987 he obtained the Dr. Eng. Sci. on instrumentation of electrocardiology. Since 1989, he has been Professor in Biomedical Engineering. Since 1994, he is a Professor with the Centre of Biomedical Engineering, Bulgarian Academy of Sciences, and part-time Professor in the Technical University of Sofia. His interests are mainly in the field of acquisition, preprocessing, analysis and recording of Biomedical Signals.

Todor Stoyanov, Ph.D.
E-mail: todor@clbme.bas.bg

Todor Stoyanov graduated as M.Sc. from the Faculty of Electronics, Technical University of Sofia, in 1999. Since 1999 he is with the Centre of Biomedical Engineering, Bulgarian Academy of Sciences. He is Research Associate since 2002. He obtained Ph.D. degree in 2005 on computer aided processing and analysis of electrocardiograms. His interests are in developing embedded systems for Biomedical Signal Analysis.