Methods of visualization and analysis of cardiac depolarization in the three dimensional space

Master Thesis

Submitted to the

Department of Computer Science and Languages at

Anhalt University of Applied Sciences

in fulfillment of the requirements for the degree of

Master of Science

Student: S. Artemov

(Matr. Nr.: 4063177)

supervisor: Dr. B. Krause

June 2016

Annotation

The master thesis presents methods for intellectual analysis and visualization 3D EKG in order to increase the efficiency of ECG analysis by extracting additional data. Visualization is presented as part of the signal analysis tasks considered imaging techniques and their mathematical description.

Have been developed algorithms for calculating and visualizing the signal attributes are described using mathematical methods and tools for mining signal. The model of patterns searching for comparison purposes of accuracy of methods was constructed, problems of a clustering and classification of data are solved, the program of visualization of data is also developed.

This approach gives the largest accuracy in a task of the intellectual analysis that is confirmed in this work.

Considered visualization and analysis techniques are also applicable to the multi-dimensional signals of a different kind.

Contents

Introduction

With the proliferation of software and hardware systems of recording and identification of biological signals, as well as mathematical and statistical data processing tools, there were a means allowing making diagnostics of any organ.

Based on these data it is possible to predict violations in work of body for long before their emergence, as well as identify the thinnest anomalies in their work. Solving these problems is the task of data mining. Data mining means extracting information from input data sets that show the some relationships with certain accuracy. Increase of accuracy is the main problem of data mining, the solution is attributes searching which more specifically will describe a signal behavior.

The null hypothesis of this work consists that the analysis of 3D EKG will provide additional data to improve the method of diagnosing abnormalities that are hidden in the correlation between the signals on axes. For searching of additional data and visual display of data 3D EKG visualization tools are used.

1. Description of the research object

The cardiography - is a common and simple method for studying heart based on the registration of low-voltage pulses produced by nervous system and heart muscle. There are different methods and ways of recording the signal differentiated by the place of collecting, time and other physical human features.

ECG represent multidimensional timelines potentials in certain points on human body.

Representation of heart activity is formed as time series of a potentials difference with reference to the pickup points on the human body. If to present a body as three-dimensional object, then it is possible to find a cooperative vector of change of potentials in space and time – 3D EKG.

Collection information about a cycle of heart work is only the first part of an electrocardiography the main and laborious part is directly research ECG. For this purpose developed special software and information store.

2. Purposes and tasks

The project purpose is increase of effectiveness of the ECG analysis by application research technique of 3D EKG that providing high accuracy in data mining tasks.

The task of three-dimensional analysis of ECG in work is complex it is divided into two subtasks directly on the analysis and visualization. These subtasks should be solved in parallel because the visualization without the analysis has no practical use and without visualization it is impossible to carry out analysis subtasks such as patterns searching, predicting ECG behavior etc.

1) The task the three-dimensional visualization consists in searching and adaptation practically applicable methods of signals representation and additional data in form of graphic information obtained as a result of data mining. To implement visualization tasks necessary to develop software that allows produce mathematical processing and display large amounts of data in three-dimensional space.

2) The data mining task includes finding repeated data sets - in order to find patterns and predict, unusual for a normal ECG abnormalities in the signal. The task contains the comparative analysis of models and divided into stages. The first step consists in constructing search pattern models based on the use the available attributes in the first case and the additional attributes in the second case, taking into account the analytical data. The second stage consists in comparison of two various models taking into account change of parameters and attributes of models.

3. Relevance and practical importance

The main type of ECG analysis for a long time was decoding ECG parameters (frequency, HEA, width of the complexes). With advent of computer means signal processing appears the task of data mining of time series. A new approach to analysis of ECG is the study heart function in three-dimensional space with the exception of time dependence. Analysis 3D ECG is more difficult because it requires a new visualization tools and data processing for purpose of solving the problem of data mining.

Based on the obtained graphs clearly visible some additional properties of ECG such as small curves, stretching and compression loop which can be processed and presented in numerical form to improve analysis accuracy.

The practical importance is to improve the efficiency and accuracy of the evaluation of ECG signal parameters. And also in obviousness of obtained diagnostic information in form of their combination with 3D ECG signal. Using visualization TEKG allows for personnel to estimate a correlation between signals, to consider HEA vector not in a projection but in 3D, to estimate deviations in form and signal distribution that provides a greater understanding of the processes taking place.

4. The modern methods of analysis of electrocardiographic signals

The task of cardiography consists in graphical representation a cardiac muscle beat during a certain time period depending on its purposes and the subsequent detection of ECG analysis options to determine deviations [1].

4.1 Recording ECG signal and physical signal interpretation

ECG recording is carried out using specialized devices that differentiated by the number of leads, resolution, availability filters (50 Hz, HF and LF), portability and a type of store devices. The graphical representation of the beat is a set of diagrams.

Graphs include full information about heart working. After the recording ECG signal carry out decoding of identification parameters which description will be presented in next part.

The projection of diagrams of leads is three-dimensional vector. The projection of it vector on the first standard axis is called the heart electrical axis (HEA) [2].

In medical practice there are several types of leads but more often use 12 major leads such as:

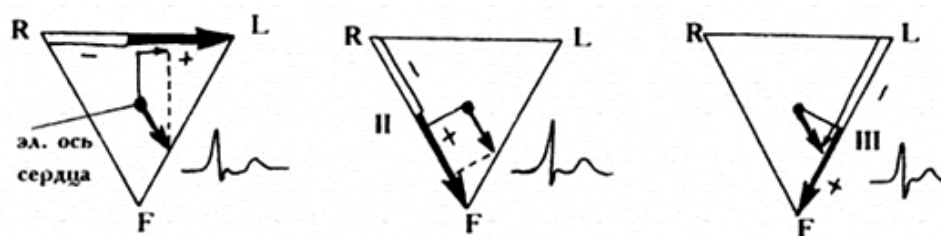1) The three standard leads (Fig. 4.1)

Figure 4.1 - Projections of the heart electrical vector

a) left hand (+) - right hand (-);

d) left leg (+) - right hand (-);

c) left leg (+) - left hand (-).

2) Three amplified single-pole leads (Fig. 4.2)
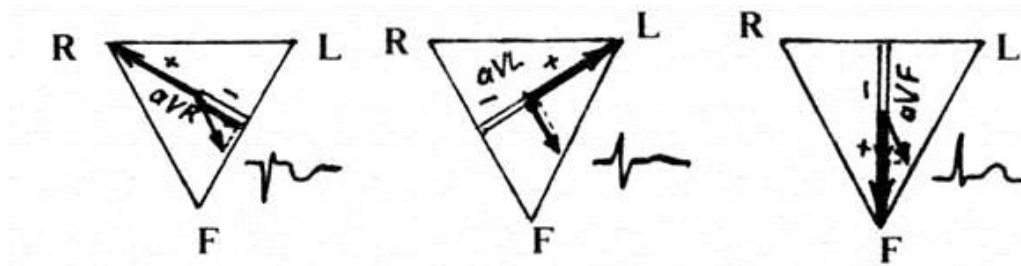
Figure 4.2 - Projections of the heart electrical vector

a) aVR - amplified lead from right hand;

b) aVL - amplified lead from left hand;

c) aVF - amplified lead from left foot.

3) Six chest leads (Fig. 4.3)

V1, V2, V3, V4, V5, V6;

Also are applied three additional leads for the purpose of sighting the diagnosis of myocardial hearths: V7, V8, V9;

Figure 4.3 shows a projection of the HEA obtained from the chest leads.

Figure 4.3 - Position chest leads and projection HEA

Furthermore there are additional leads types which are recorded separately and necessary to identify a particular behavior heart function and areas with violation or abnormal beat.

To such leads include:

1) Three bipolar leads by W. Nebh (Fig. 4.4)

Used for additional diagnosis of myocardial changes anterolateral, back and of upper parts the front wall of the ventricle.
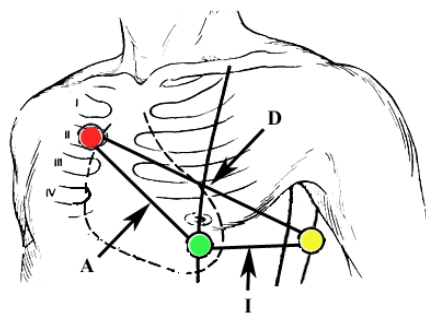
Figure 4.4 – Position leads by W. Nebh

2) Orthogonal Frank leads

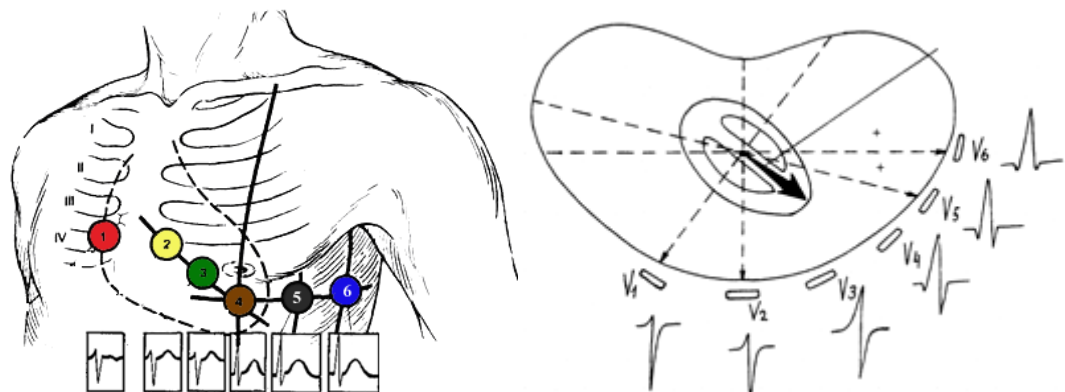The orthogonal signals is three chest leads. In this case the most simple and convenient X, Y, Z leads. The axes are perpendicular to each other and perpendicular to the person plane. Figure 4.5 shows an image of relative location of the ECG leads in space.
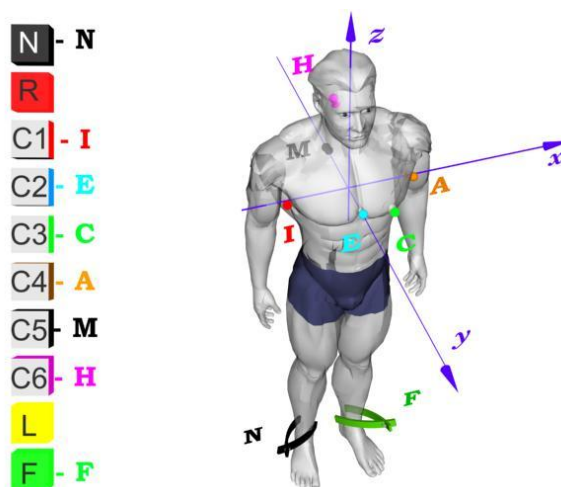
Figure 4.5 - The relative position leads in space

Thus if we imagine the human body as a three-dimensional object that is quite simple make signals to spatial coordinates which is used as initial data herein.

4.2 Description of source data

As input data used projections of potentials to XYZ axes. Information is collected by the device with the sampling frequency 1 kHz. The projections on the axes are calculated with respect to 12 major leads or orthogonal Frank leads. Three variables define a set of XYZ points in the three-dimensional space.

After converting signals to study the characteristics of the heart carried out a preliminary classification of the beats and the received set of files with the following variables:

- Source data array XYZ with dimension:

$$N_{beats} \text{ x } N_{samples} \text{ x } N_{axes} \qquad (1)$$

where N_{beats} - number of beats; $N_{samples} = 1500$ – count samples in beat; N_{axes} - 1-x, 2-y, 3-z - axis number.

- Description of beats according to the classification (normal heartbeat, tachycardia, interference, etc.);

- Additional metadata (patient information).

All information presented in MatLab file. Software MatLab has a number of benefits required for quality data processing and visualization. One of these features is three-dimensional data visualization MatLab allows flexible and quite simple to implement this functionality.

4.3 Methods of analysis and signal parameters

ECG analysis is based on the evaluation the graphs leads.

Normal ECG graph (Fig. 4.6) contains usual 5 peaks: P, Q, R, S, T, and sometimes peak U. QRS complex peaks corresponds to ventricular depolarization. The remaining ST segment and T waves characterize the reverse process - repolarization.
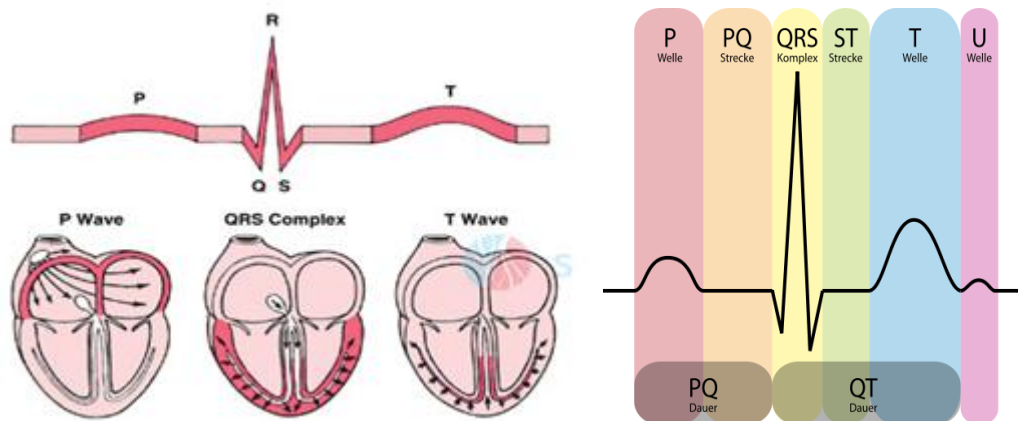

Figure 4.6 - Normal form ECG and splitting by intervals

Based on this structure is constructed following analysis that identifies additional parameters.

ECG analysis is performed in sequential order:

1) Evaluated rate and measured the frequency of contractions (at normal ECG - sinus rhythm, heart rate - 60 ... 80 beats per minute);

2) Calculated intervals (QT, normal - 390 ... 450 ms) which characterize the duration of compression phase by special formula. It reflected by pulse intervals and calculated using specialized software;

3) The position of HEA calculated by isoline at the height peaks;

4) Is studied QRS complex:

- The width of complex;

- The amplitude of complex;

- The amplitude of peaks.

5) Is studied segment ST reflecting recovery period of heart muscle initial state after depolarization and T wave characterizing the repolarization process.

This operation is very laborious and requires a doctor's attention, but there are software systems that help doctor to identify the main ECG parameters and increase the data values accuracy.

4.4 Software for visualization and analysis of ECG signal

With the proliferation of computer diagnostic equipment with electronic systems of data collection and storage processing ECG analysis on paper become impractical. Special software is used to analyze the ECG that allows doctors to transfer some functions in automatic mode and thus increase diagnosis accuracy.

To analyze the means used for visualization and analysis of ECG need to search and analyze the functionality of existing software.

The obtained information about functionality must be considered in implementation of tasks of analysis and visualization 3D EKG in this work.

1) ECGSIM program

This program (Fig. 4.7) is a tool heart simulating and find the relationship between electrical potentials of leads as well as for search and localization areas of heart that may be different kinds of deviations.

Figure 4.7 - ECGSIM program interface
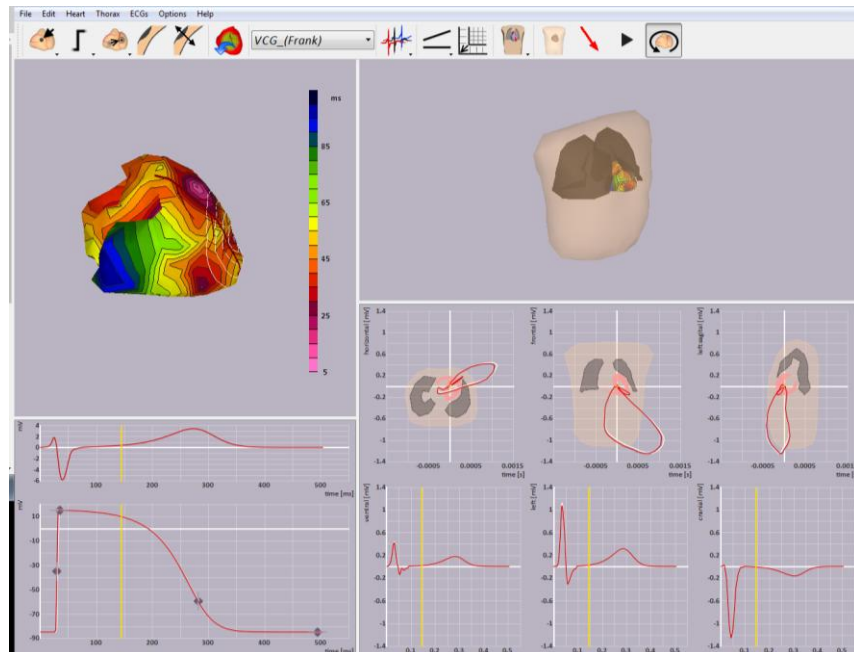
The program works with three-dimensional images of heart muscle and body, builds and animates the vector summary potential, calculates the surface map the spread of electrical wave velocity.

ECGSIM allows working with the above types of leads: 12 major leads are shown in Figure 4.8, with orthogonal Frank leads (Fig. 4.5) and others.
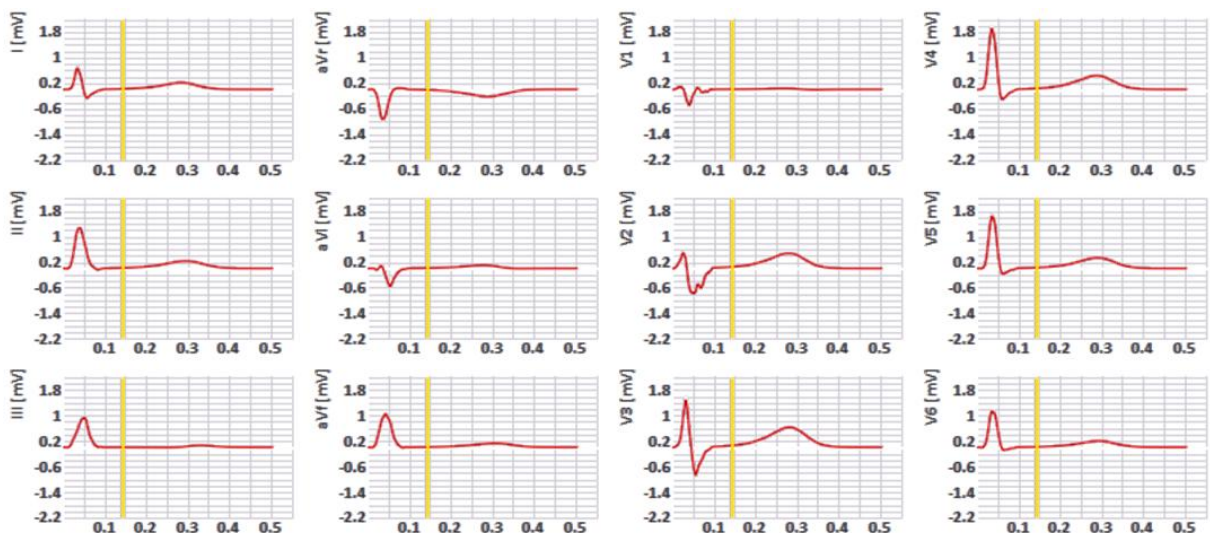
Figure 4.8 - Visualizing graphs of 12 leads

Visualizing the summary potential vector is shown in Figure 4.9, program allows playing animation of the vector, the image clearly shows the offset and amplitude of vector.
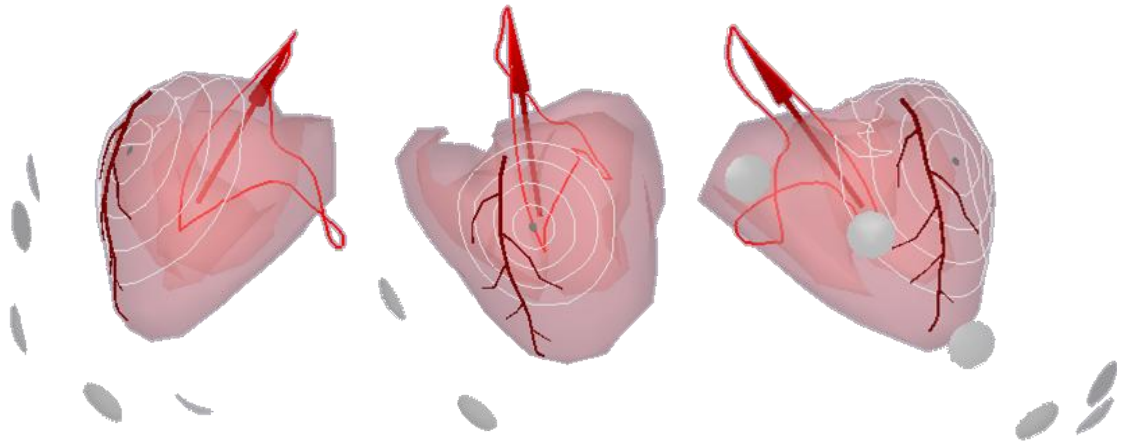
Figure 4.9 - Visualizing the heart's electrical axis

In addition ECGSIM allow build a map of the surface wave propagation velocity (Figure 4.10) it can be used as a means of search and localization areas with abnormal behavior.
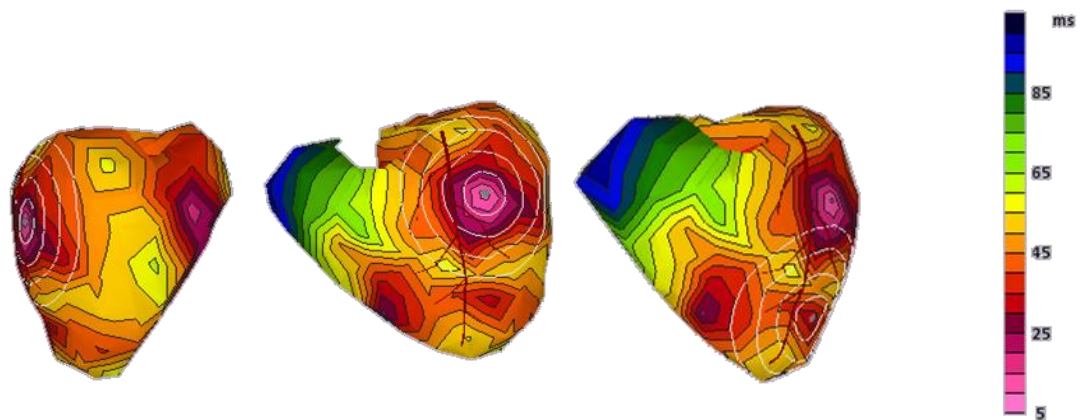
Figure 4.10 - Visualizing time passing electrical impulse to the heart areas

2) Hardware-software complex cardiometry-MT

This complex is a doctor workplace and consists of a cardiograph and direct analysis and visualization program (Fig. 4.11) which recorded all information about the patient.
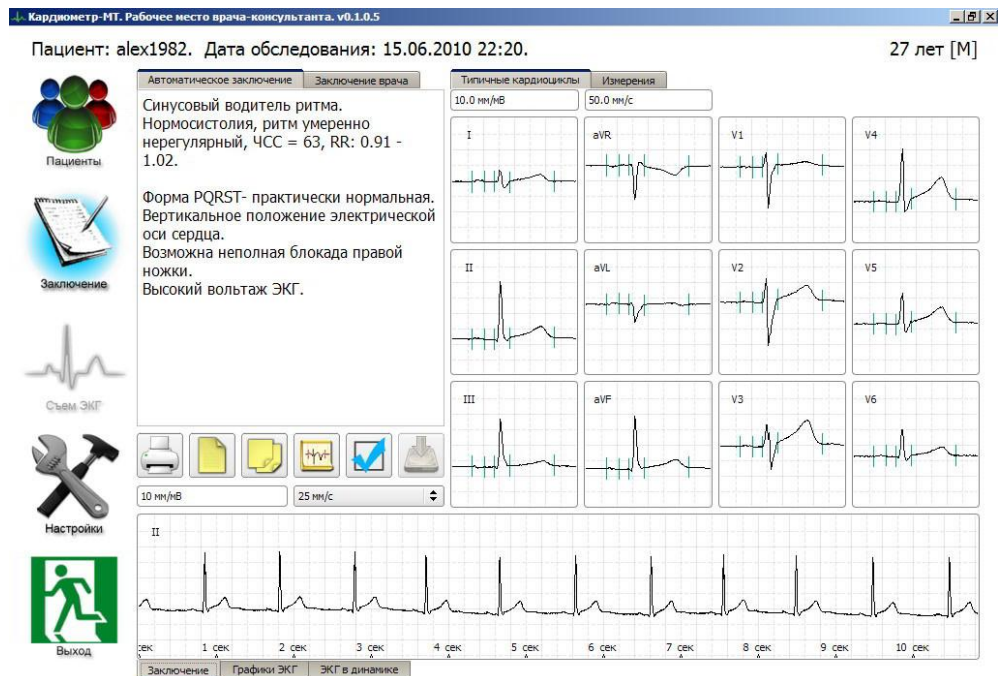
Figure 4.11 - Interface of program cardiometry-MT

The program allows to:

- to evaluate dynamics of changes ECG in time;

- to measure value of the amplitude-time ECG parameters;

- automatically mark of beats cardiac complexes described above;

- to display leads graphs;

- to evaluated using embedded items amplitude and delays.

3) EKG Viewer program

EKG Viewer is a visualizer the ECG signal with additional features:

- measurement of the length and amplitude;

- graphs rescaling;

- analysis and measurement the main ECG parameters.
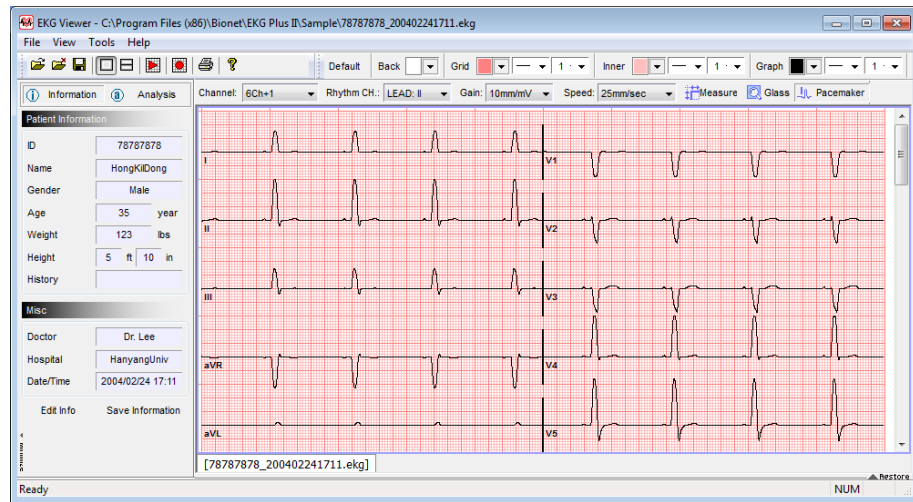
Figure 4.12 - ECG Viewer software interface

4) The hardware-software complex myocardium 12

The complex also provides a set of equipment with supplied software. This program (Fig. 4.13) is a working doctor place with data base of ECG data. To work with ECG doctor is able to:

- also display all leads;

- calculate the main ECG parameters;

- make amplitude and length measurement.

Figure 4.13 - Myocardium 12 program Interface

The survey showed that the main features of existing ECG analysis are:

1) finding all showing above ECG parameters;

2) partition ECG heartbeat;

3) classification of beats waveform;

4) simulation of the heart as graphing relative touch point to the heart muscle;

5) finding shifted and compressed signal;

6) normalization and alignment;

7) visualization leads by graphs;

8) visualization special search time delay functions motion vector heart potentials to identify abnormal behavior.

These functions will be considered for the work tasks.

5. Visualizing three-dimensional ECG signal

Visualizing beats in three-dimensional space does not much popularity and practical use because this type ECG is difficult to analyze. With computer technology application to this knowledge field manifested more interest. ECG representation in this way has certain features, since a time value is lost and instead comes into effect signal propagation speed and also produces visible signal correlation with respect to each other considered in a small degree at analysis in the time space.

5.1 Three-dimensional QRS complex representation

The graph 3D EKG (Fig. 5.1) is oriented at an angle loop with different kinds curves, spiral shape and a small loops in myocardial depolarization space. On graph can be seen 10 beats 3D EKG which have similar shape with a small but significant differences.
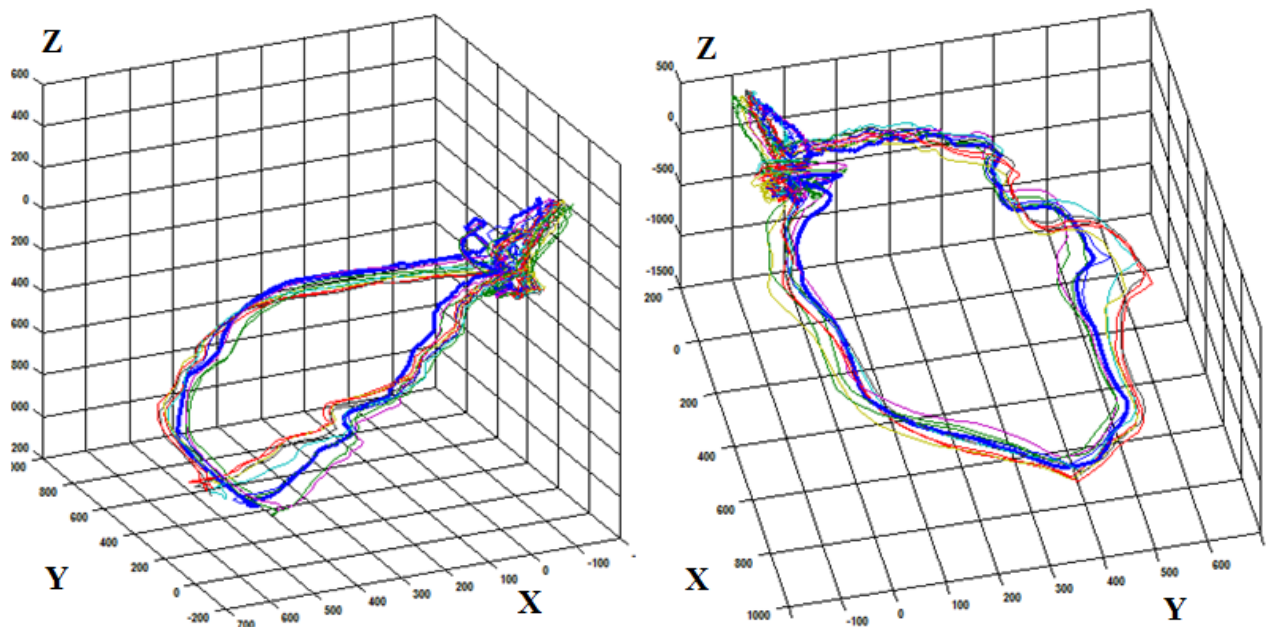

Figure 5.1 - Visualizing beats complex of 3D ECG

However more interesting visualization of the QRS complex because its shape and orientation is indicator of most significant deviations in heart. These include deviation HEA, the conduction abnormality, the heart hypertrophy, the

blockade of atrioventricular bundle, beats and etc. Figure 5.2 shows the QRS complex beats.
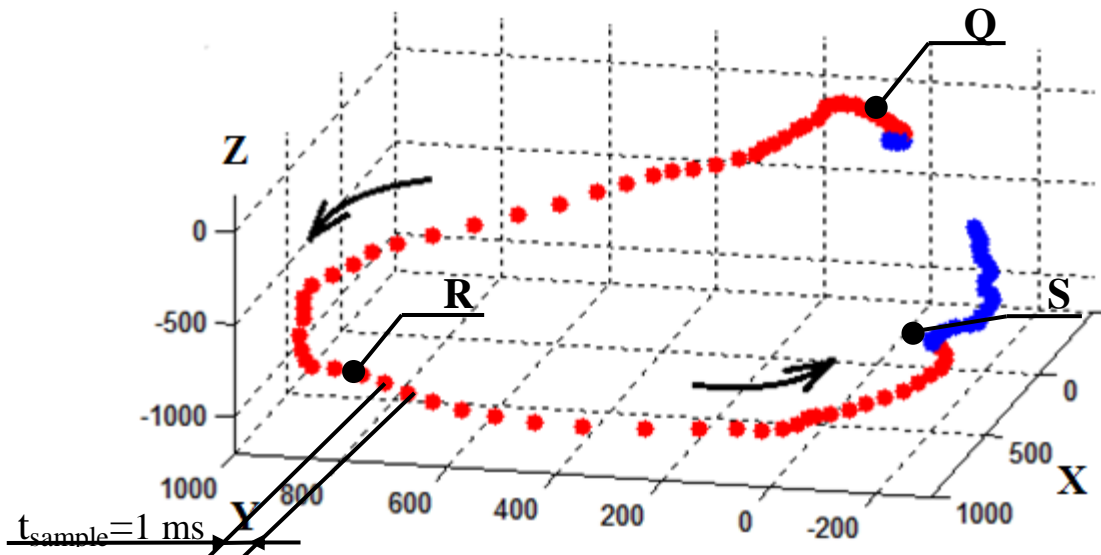
Figure 5.2 - The graph of the QRS complex in three-dimensional space

The graph has expressed the QRS peaks and also can be see the difference distances between sampling points with a period of 1 ms and direction of graph propagation shown by arrows.

5.2 Features of ECG signal research in three-dimensional representation

For comparison shown two heartbeats (Figure 5.3) a single patient in different intervals on axis X. As can be seen graphs have similar amplitude and width of QRS complex but the number beating 2 is before further tachycardia. Changes in the time graph are not visible.
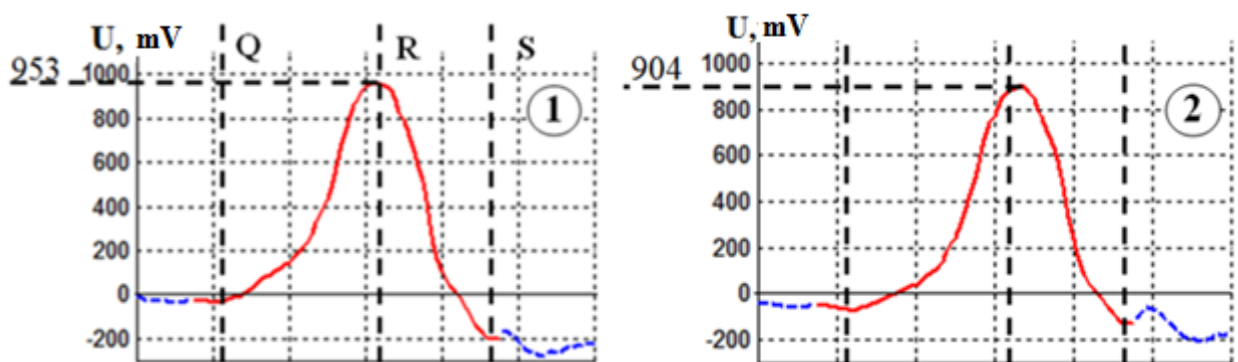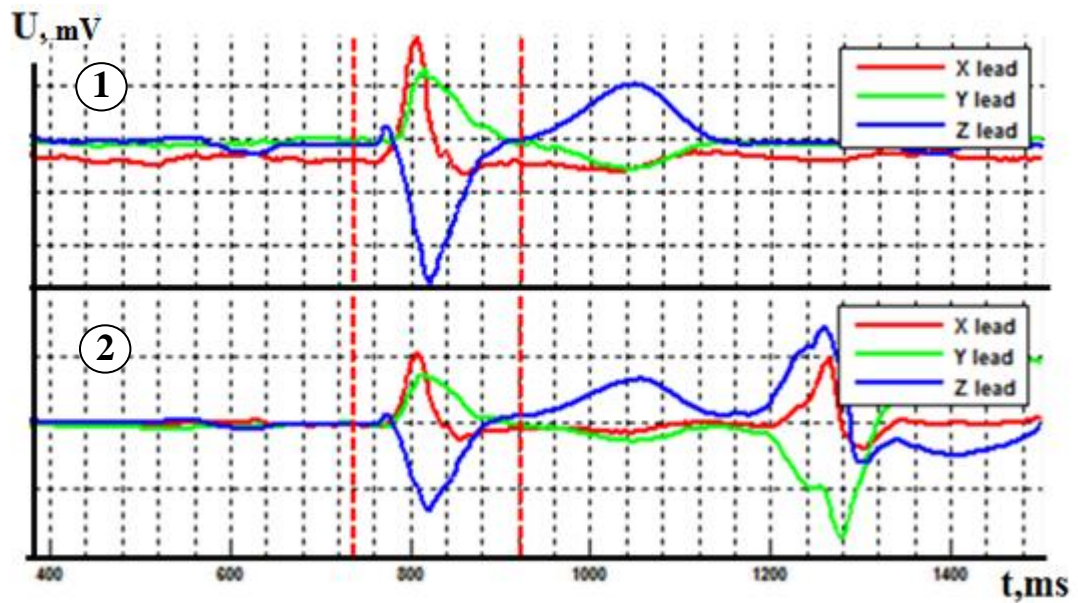
Figure 5.3 - The graphs two beats of QRS complexes

Figure 5.4 show the same beats in three-dimensional space. As can be seen from graphs beat number 2 has a similar shape with beats number 1 but with peaks shift of QRS complex and with rotation angle in space.
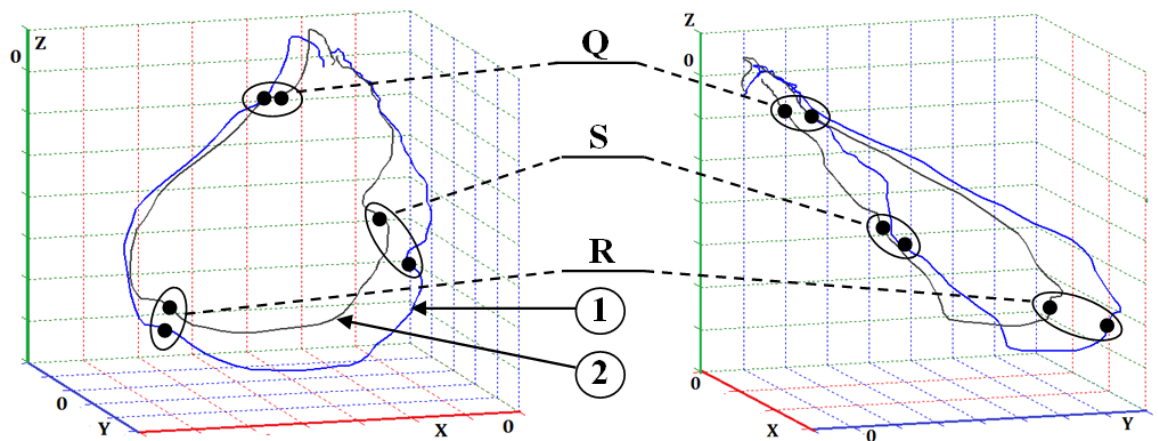
Figure 5.4 - The graphs QRS complexes of two beats in three-dimensional space from different perspectives

The relative angle of rotation α is shown in Figure 5.5. Calculation of average planes will be presented in the next part.
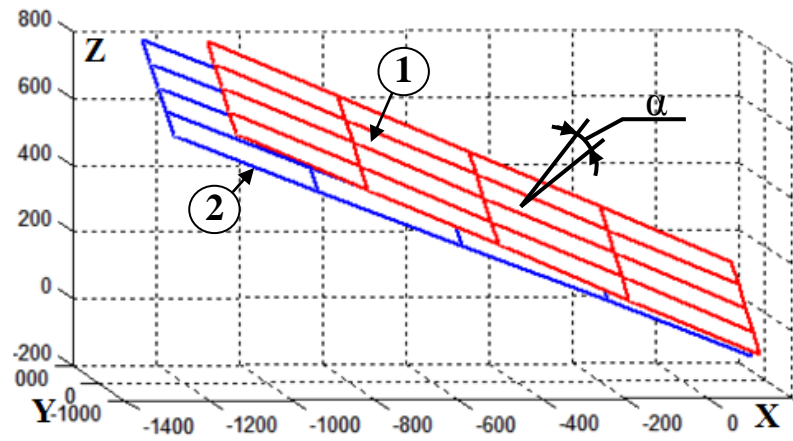
Figure 5.5 - The average ratio of two beats planes.

Thus, the presentation of data in three-dimensional space makes it possible to evaluate invisible changes and deflections.

5.3 Visualization methods of ECG signal three-dimensional representation

5.3.1 Finding and visualization of middle dynamic vector of heart electric axis

Earlier in part 4.1 explain the concept of the electric vector of heart axis. Visualizing vector on leads projections are shown on Figures 4.1 and 4.2.
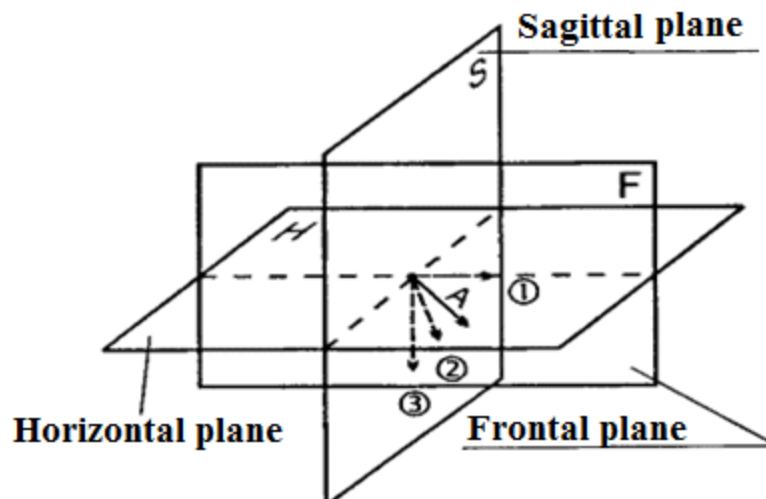
Figure 5.6 - The projection of electrical vector at different planes

However the precise vector provision in space at interpretation state of ECG is not considered and based on its projection to leads in a certain plane. Figure 5.6

shows the plane with respect to that calculated position vector axis. The most commonly used is the frontal plane.

Figure 5.7 shows the graphs showing the find direction of the electric vector of the heart axis in the frontal plane. Where the graph axes is standard and enhanced leads.

a.   b.
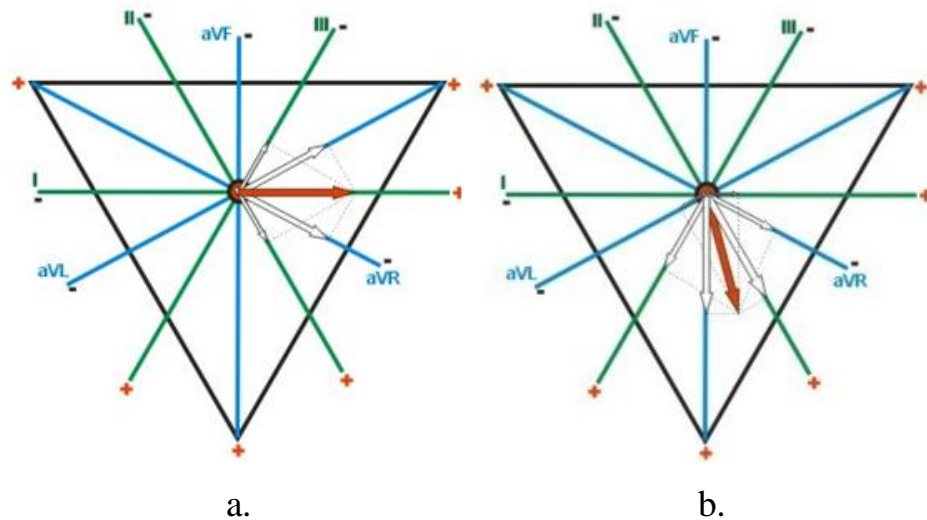
Figure 5.7 - Projections of heart electric vector: a - deviation from the norm; b- average direction

However with projections some data is lost which characterizes the behavior of beats in the other projections, but this information should be considered at diagnostic state.

Another important parameter is the dynamics of vector position HEA. This dynamics reflects the position vector at each time point and at each sampling point in graph. However this is not considered in the normal ECG analysis because of complexity of understanding and displaying of this parameter.

5.3.1.1 Description of the mathematical method for calculating the electrical heart axis

The basis of algorithm to solve the problem of determining HEA is method of principal component analysis (PCA).

This method allows to transform a set of variables dependent variables (the correlating data) in this case located in three-dimensional space, to space with

smaller number of measurements, in this case the plane with explanatory variables (not correlating with each other). PCA simplifies data presentation for consideration and processing.

The presented algorithm PCA solves the problem of data approximation by linear manifold which is one of the four main applications of method.

The mathematical description of these approximations of linear manifolds given below.

Given the set of finite vectors (points):

$$x_1, x_1, \ldots, x_m \in \mathbb{R}^n \quad (2)$$

where $\mathbb{R}^n$ - space linear manifolds

The task is to find this:

$$L_k \in \mathbb{R}^n \quad (3)$$

where $k = 0,1, \ldots, n-1$

that the sum of squared deviations by x_i from L_k was minimal (Fig. 5.6):

$$\sum_{i=1}^{m} dist^2(x_i, L_k) \rightarrow min \quad (4)$$

where $dist(x_i, L_k)$ - the distance from point to linear manifold.

Any k - dimensional linear manifold $\mathbb{R}^n$ can be defined as:

$$L_k = \{\alpha_0 + \beta_1\alpha_1 + \cdots + \beta_k\alpha_k | \beta_i \in \mathbb{R}\} \quad (5)$$

From all this, the principal component vectors can be found as:

$$\alpha_k = arg\min_{\|\alpha_k\|=1}(\sum_{i=1}^{m}\|x_i - \alpha_k(\alpha_k, x_i)\|^2) \quad (6)$$
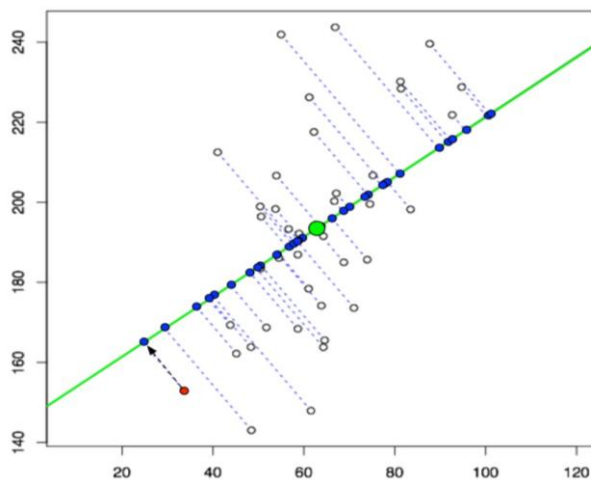
Figure 5.8 - Example of finding the optimal approximation in two-dimensional space

5.3.1.2 Description of the algorithm implementation of the method of principal component analysis in MatLab

The algorithm for calculating the position HEA is implemented as a function of MatLab:

function [normal,Err] = func_normal_vector(XYZ, win_size, start_point, end_point)

Input variables:

1) XYZ - a variable that represents a set of points in three-dimensional space (Fig. 2) in the format (number of beats, the sample number; axis);

2) win_size - the size of the window construction of the plane;

3) start_point, end_point - start and end point

Output variables:

1) normal - an array of normal vectors to the planes;

2) Err - total error, the amount of deviation from the plane of each point.

This function works in cycle with the following procedures:

1) is defined by the start and end point:

*p1 = start_point+(step*i-step);*

p2 = p1 + win_size;

2) from the array XYZ extracted point which amount is determined by win_size:

XData =(XYZ(1,p1:p2,3))';

YData =(XYZ(1,p1:p2,1))';

ZData =(XYZ(1,p1:p2,2))';

3) for implementing PCA search function uses a standard feature:

[coeff] = princomp(X);

Which input variable is an array of points. At the output vector of linear coefficients of plane equation (4) - coeff.

4) Here normal vectors are the coefficients of equation:

normal(:,i) = coeff(:,3);

For their building on the graph using that expression:

N = [normal(1,i)*350+XData(1,1); normal(2,i)*350+YData(1,1);

normal(3,i)*350+ZData(1,1)];

5) For error finding (deviation) using the sum of distances between the planes and the points for each iteration:

error = abs((X - repmat(meanX,n,1))*normal(:,i));

Err(:,i) = sum(error);

6) After calculation the normal vector then shift window by size of step and so on.
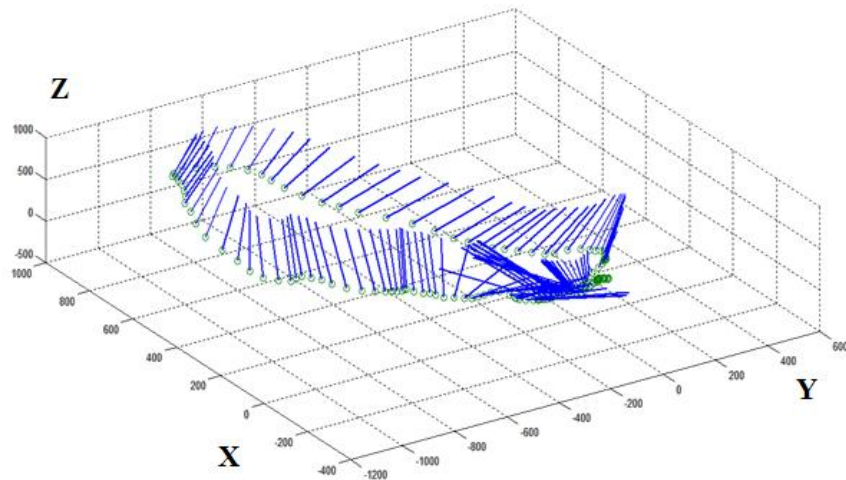

Figure 5.9 - Distribution of normal vectors in space

As a algorithm result is a set of vectors (Fig. 5.9) and the corresponding total points of deviation by plane - approximation error (Fig. 5.10).
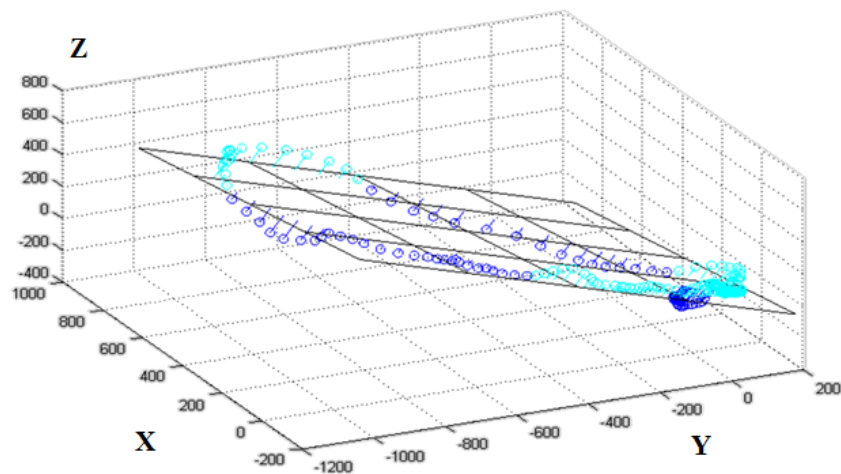

Figure 5.10 - The deviation of the points from the fitted plane

5.3.2 Finding approximating surfaces

As in the previous task the search approximating surface in this project is used for clear visualization of the heart vector changes.

The idea is to find a curved surface which shows the averages and bending and twisting of 3D EKG.

The task is to find non-linear regression model and subsequent construction of the surface based on non-linear equations and calculated coefficients.

5.3.2.1 Description of the mathematical method of surface search

The simplest regression search method - calculation of linear regression.

Linear Regression - this is the approach to modeling the relationship between scalar values of the positions of points in space.

Nonlinearity in this case considered at data preparation state where one of variables costs from two previous. Thus the program work with linear regression but at output is non-linear equation.

Since MatLab working with matrices and vectors the method will describe in vector form.

The basic concept is the regression model:

$$Y = X\beta + E \quad (7)$$

where

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ x_{21} & \cdots & x_{2k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix} \text{ - matrix of covariates (factors)}$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \text{ - values vector}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \text{ – coefficients vector (model parameters)}$$

$$E = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} - \text{error vector}$$

Then the regression equation looks like a set of solutions:

$$y_i = \beta_i x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad (8)$$

where $i = 1 \dots n$

To find the coefficients β, need to find the rate of change covariates:

$$\beta_i = \frac{\partial f}{\partial x_i} = \text{const} \quad (9)$$

Thus the problem reduces to finding a model with a minimum value of the approximation error and calculation coefficient values.

5.3.2.2 Description of algorithm implementation in MatLab

To perform these tasks write another function:

function [mdl] = func_non_leniar_plane_weidth(XYZ, beat_n, start_point, end_point)

Input variables:

1) XYZ - a set of points in three-dimensional space (Fig. 2);

2) beat_n – beat number;

3) start_point, end_point - start and end point.

The output variables:

1) mdl - structure in MatLab format which elements comprise:

3) Coefficients - the coefficients of regression equation;

4) Formula - regression formula description;

5) Coefficient_Covariance - covariance matrix.

To calculate the coefficients of the equation prepared data by:

X1 =(XYZ(beat_n,start_point:end_point,3))';

X2 =(XYZ(beat_n,start_point:end_point,1))';

X3 =(XYZ(beat_n,start_point:end_point,2))';

*X = [ones(size(X1)) X1 X2 X1.*X2];*

For this task was decided to apply the weighting factors for calculating the coefficients of the regression equation:

w = ((X1.^2)+(X2.^2)+(X3.^2)).^(0.5);

The weighting factors represent modules distance between points in three-dimensional space.

Using standard function of MatLab Fitting Toolbox:

mdl = fitlm(X,X3,'linear','Weights',w);

Where parameters include regression vector X, vector values X3, linear regression type and weight of the additional parameter. At the output structure described above.

Next, build a grid with a period of 100:

x1fit = min(X1):100:max(X1);

x2fit = min(X2):100:max(X2);

[X1FIT,X2FIT] = meshgrid(x1fit,x2fit);

Calculate the plane on this grid:

*YFIT = table2array(mdl.Coefficients('x1','Estimate')) + table2array(mdl.Coefficients('x2','Estimate'))*X1FIT + table2array(mdl.Coefficients('x3','Estimate'))*X2FIT +table2array(mdl.Coefficients('x4','Estimate'))*X1FIT.*X2FIT;*

The last step - to build this surface with the perpendicular lines at each grid point:

mesh(handle_in,X1FIT,X2FIT,YFIT);

surfnorm(handle_in,X1FIT,X2FIT,YFIT);

The result of this function is the visual representation of changes in vector heart - perpendicular to the plane of surface (Fig. 5.11).
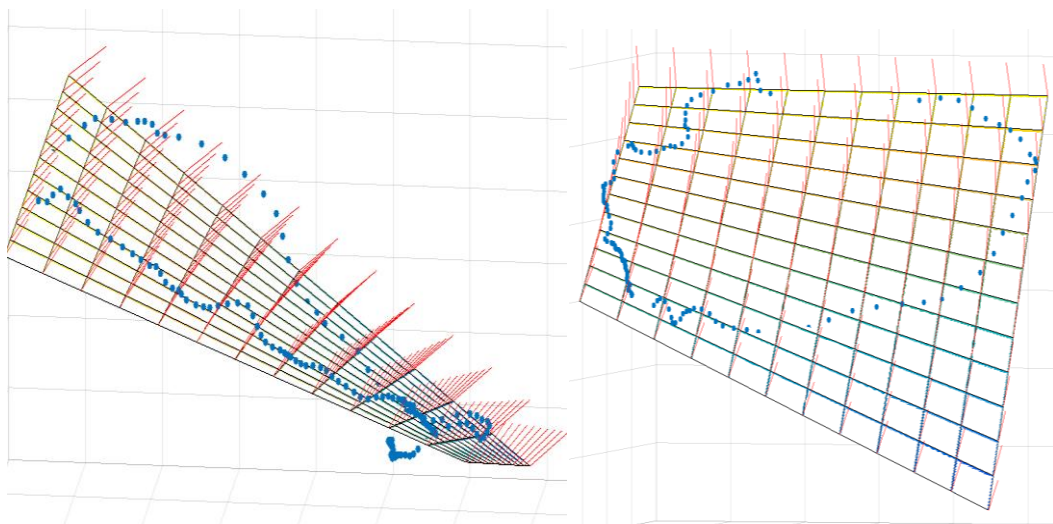

Figure 5.11 - surface approximating from different perspectives

5.3.3 Visualizing dynamic electrical axis vector via a spherical surface

Another type of HEA visualization is transition to spherical coordinates (see Fig. 5.12, a), in which vector is a segment OM constant length the coordinates of which are given by the angles φ - azimuth and θ - elevation.

a.   b.
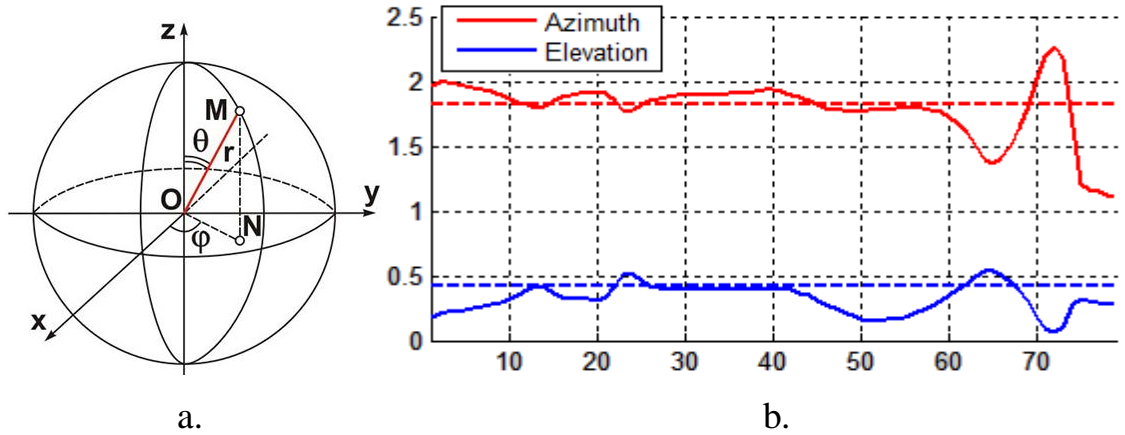
Figure 5.12 - Position HEA vector in spherical coordinates: a - spherical coordinates; b - a graph of position vector.

Figure 5.12, b shows change in the position heart vector of one beats, the dotted line shows mean angles of complex designed according to previous chapter.

Another visualization type of the vector in spherical coordinates - using surface areas for displaying motion track vector on surface. Figure 5.13 shows the track of the motion vector as seen at sphere displayed tracks and points clusters.

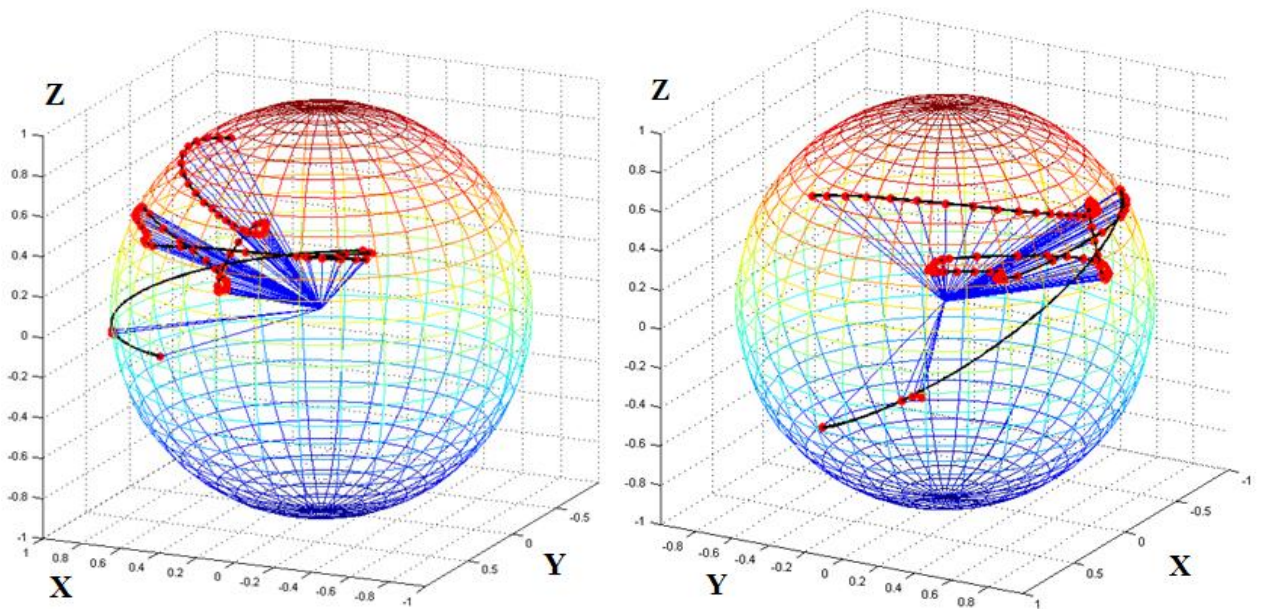Figure 5.13 - Visualizing of the electric heart vector position on spherical surface

The main track has two loops are arranged symmetrically relative to each other, the clusters formed at the ends of QRS complex, where position vector is not significant changes. Thus we obtain a way to visualization, of track displaying relative shape and position which clusters may determine deviation in the direction of propagation of electrical work of the heart waves.

5.3.4 The use algorithm of dynamic time warping to visualize the relative displacement signal

DTW is used to align the graphs of same length relative to each other namely to calculate the difference between the graphs. It is widely used for speech recognition and data mining problems. However, it can apply to other time-varying signals.

The basic idea - use function DTW for visualization of ECG deviations in three dimensions from each other using the additional lines connecting to corresponding points.

5.3.4.1 Mathematical description of DTW algorithm

The classical approach to calculating DTW considered further:

1) calculated distance matrix - this is usually the Euclidean distance corresponding points:

$$d(q_i, c_i) = |q_i - c_i| \quad \text{or} \quad d(q_i, c_i) = (q_i - c_i)^2 \quad (10)$$

where q_i, c_i - graphics shown at Fig. 5.14, a.

Figure 5.14 shows a graph of distances distribution between points of two signals represented as a distance matrix d.

2) calculated matrix D, which represents the transformation of the coordinate axes:

$$D_{i,j} = \min(D_{i-1,j}; D_{i,j-1}; D_{i-1,j-1}) + d_{i,j} \quad (11)$$

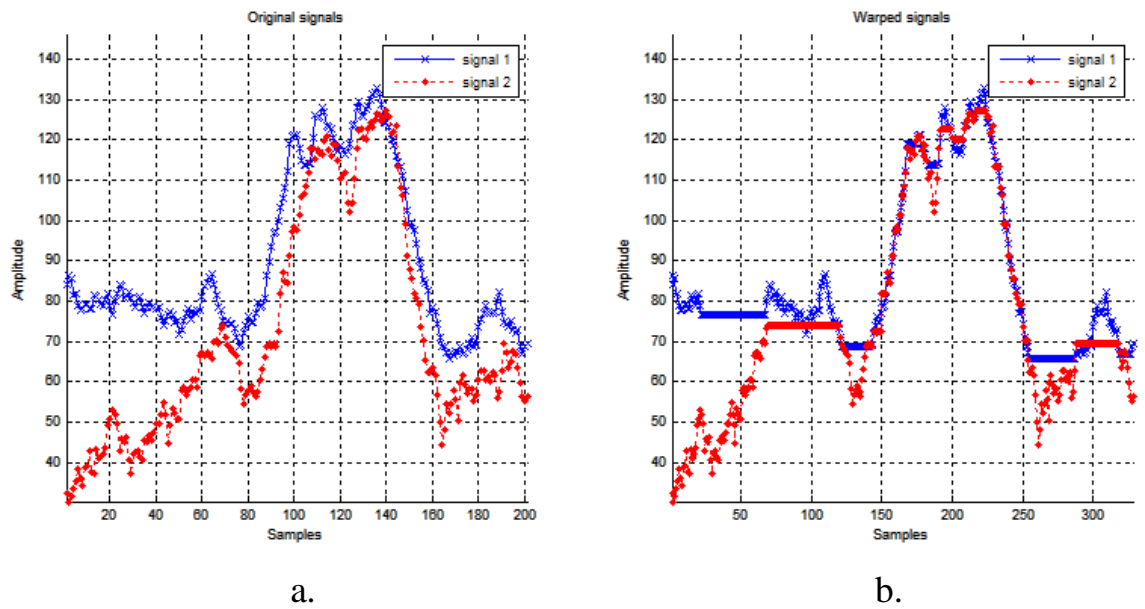a.   b.

Figure 5.14 - The graphs before and after transformation a - source graphics; b - graphs after DTW.
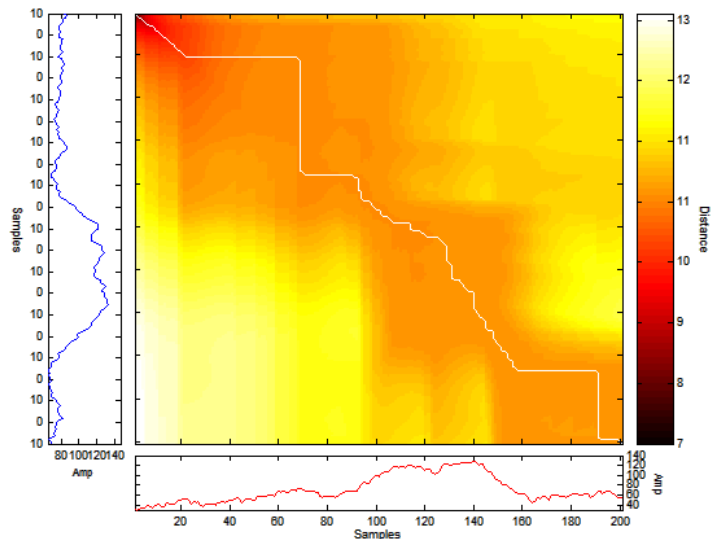

Figure 5.15 - Calculation of the optimal path transformation

Matrix D represents minimum length of path in the graph.

As can be seen from the graph, the most optimal path will be a line that runs along closest to the red shade of the way.

Figure 5.15 shows the matrix with signals placed in upper row and first column to data signals calculated distance between points d. In this case, there are three directions of movement of the optimal path: right, down, right and down. The direction is based on the minimum accumulated distance.

For example, the optimal path is shown lighter color in Figure 5.16. At point are summed value of 1.1 at the point 0 and all values of surrounding matrix in three directions, in this situation there are two optimum direction the right and down. Then for these directions calculate sum of values of others and their selected minimum value. Calculation make according to above formula.

	1	2	4	2	5	6	3	3	0	-1	0
1	0	1	3	1	4	5	2	2	1	2	1
3	2	1	1	1	2	3	0	0	3	4	3
5	4	3	1	3	0	1	2	2	5	6	5
2	1	0	2	0	3	4	1	1	2	3	2
3	2	1	1	1	2	3	0	0	3	4	3
6	5	4	2	4	1	0	3	3	6	7	6
7	6	5	3	5	2	1	4	4	7	8	7
5	4	3	1	3	0	1	2	2	5	6	5
3	2	1	1	1	2	3	0	0	3	4	3
-2	3	4	6	4	7	8	5	5	2	1	2

Figure 5.16 – Calculating example of optimal path in DTW algorithm

For a more visual picture of search path in Figure 5.17 displays the optimal path search graph in 3D form where the values of distances matrix d, located on Z axis and marked hot-map color.
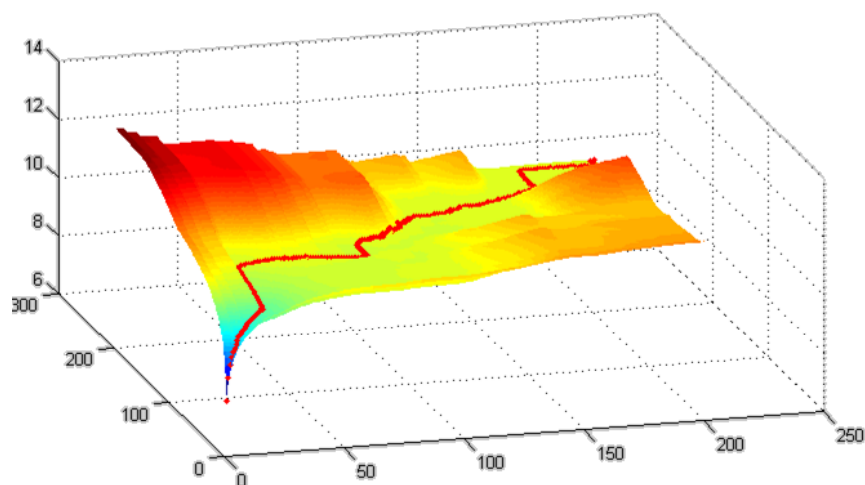

Figure 5.17 - Visualization of optimal path transformation in three-dimensional form

3) The next step is the calculation matrix W, which is array of points compounds. At this stage the problem of defining optimal path. In Figure 5.14, b shows graphs after transformation.

5.3.4.2 Description of algorithms in MatLab

DTW algorithm is implemented as a function of:

function [Dist,D,k,w,rw,tw]=dtw(r,t,pflag,version)

Input parameters:

1) r - reference signal;

2) t - transformed signal;

3) pflag - rendering flag graphics;

4) version - the calculation of DTW type of 2D or 3D.

Output variables:

1) Dist - an array of distances between signals;

2) D - accumulating matrix of distances;

3) k - normalization factor;

4) w - matrix of optimal path;

5) rw, tw - transformation vector r and t respectively signals.

According to the mathematical description the first thing to do - is to find the Euclidean distance between points, the operation is carried according to formula:

$$d = (X1 - X2)^2 + (Y1 - Y2)^2 + (Z1 - Z2)^2 \quad (12)$$

According to scripting language MatLab the operation is written as follows:

d=((repmat(X1',1,N)-repmat(X2,M,1)).^2+(repmat(Y1',1,N)-
repmat(Y2,M,1)).^2+(repmat(Z1',1,N)-repmat(Z2,M,1)).^2);

Then calculated distance matrix is accumulating in the loop in accordance with:

```
for m=2:M
   for n=2:N
      D(m,n)=d(m,n)+min(D(m-1,n),min(D(m-1,n-1),D(m,n-1)));
   end
end
 [row,M]=size(X1);
[row,N]=size(X2);
```

Next step - find the matrix w - optimal path. The operation is performed in cycle:

```
w=[M N];
 while ((n+m)~=2)
   if (n-1)==0
```

```
    m=m-1;
  elseif (m-1)==0
    n=n-1;
  else
    [values,number]=min([D(m-1,n),D(m,n-1),D(m-1,n-1)]);
    switch number
    case 1 m=m-1;
    case 2 n=n-1;
    case 3 m=m-1;  n=n-1;
    end  end
  k=k+1;  w=[m n; w];
```

Further, if exposed pflag necessary to draw graph.

A visualization task is to connect the corresponding points according to the matrix w. Figure 5.18 shows a rendering method using a DTW algorithm.
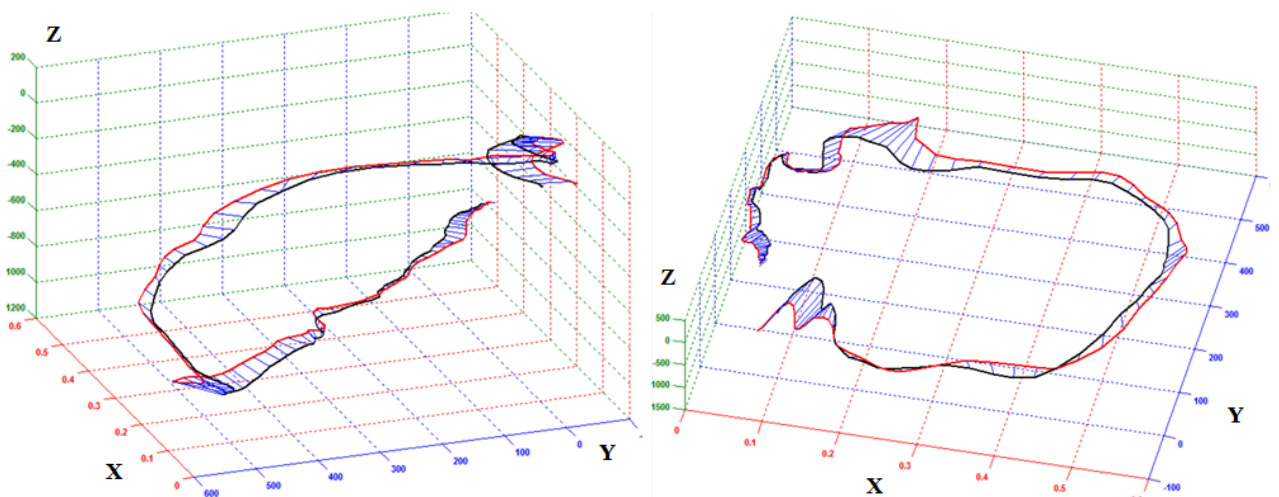

Figure 5.18 - Visualizing using DTW algorithm from different angles

Thus, we obtain additional method of visualization and calculation the distance between beats, which is useful later in analysis part in this work.

5.3.5 Identification and visualization of clusters peaks

The task is to indicate position in three-dimensional space needed peaks graphs selected beats. Based on further analysis resulting cloud of points and their coordinates, it is possible for example, to predict ventricular cardiac arrhythmias.

5.3.5.1 Mathematical description method of determining peaks

As can be seen from the graph (Fig. 5.19) of heart muscle beats cycle is divided into intervals: PR, QRS, QT, and others. Major peaks is PQRST. Based on this determination task placed peak positions in three-dimensional space. For each

coordinate is quite difficult to determine PST peaks in this case it was decided to determine the peaks based on coordinate transformations in total vector magnitude (Fig. 5.20):
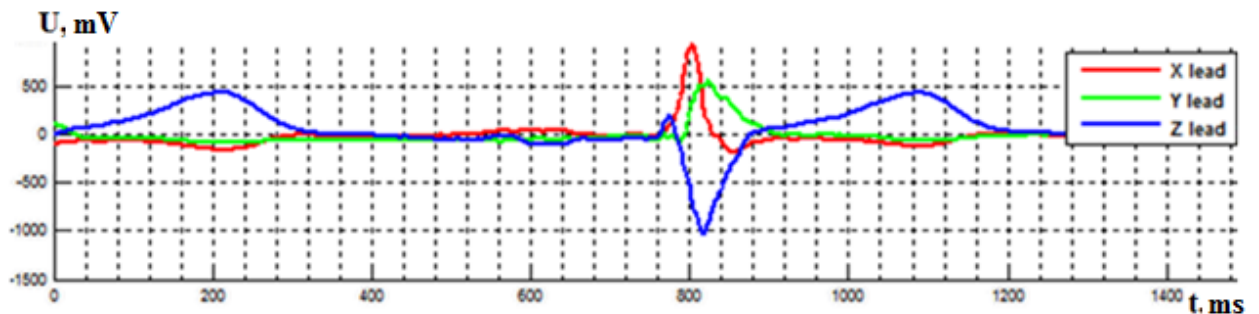
$$M = \sqrt[2]{X^2 + Y^2 + Z^2} \quad (13)$$

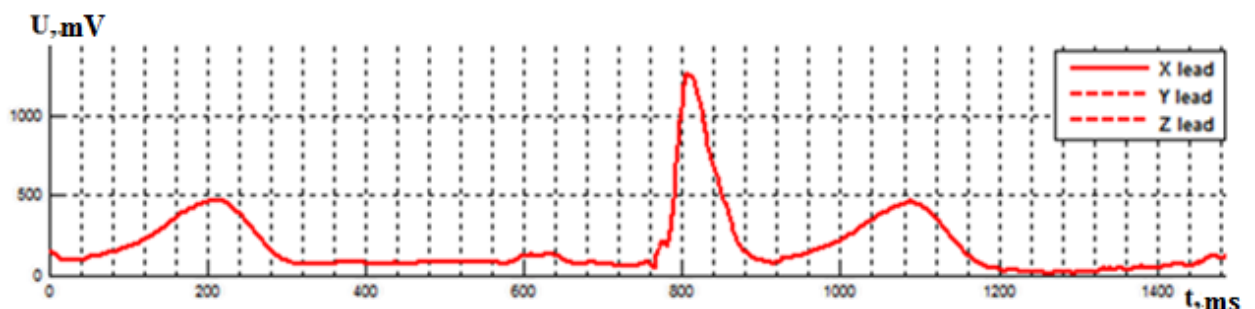Figure 5.19 - The graphs input ECG signals

Figure 5.20- The graph magnitude of ECG signal

5.3.5.2 Description of algorithm peaks search in MatLab

To find peaks in MatLab environment was developed function:

function [pks,locs] = func_find_extremum (XYZ, beat_n, start_point, end_point, axes, peaks_count, peaks_dist)

Input parameters:

1) XYZ - a set of points in three-dimensional space;

2) beat_n - beats number;

3) start_point, end_point - start and end point;

4) axes - takes the value 'X', 'Y', 'Z', 'M', graph in which peaks are searched;

5) peaks_count - number of returned peaks;

6) peaks_dist - the distance between the peaks.

Output variables:

1) pks - the magnitude of peak;

2) locs - peak position.

Since is quite difficult to identify peaks in the noisy signal, it was decided to apply a digital filter (Figure 5.21):

magnitude_f = filter(b,a,magnitude);

where filter parameters:

windowSize = 10;

b = (1/windowSize)*ones(1,windowSize); %Nomination coefficients

a = 1; % Denomination coefficients
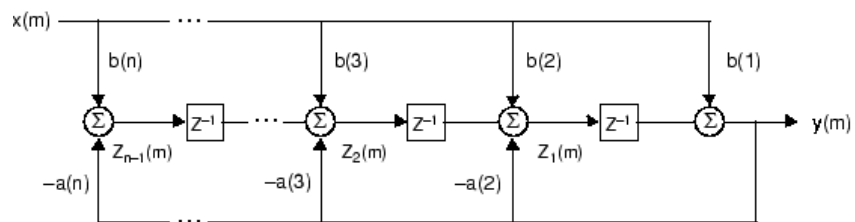

Figure 5.21 - Scheme of the digital filter in MatLab

Next, using a standard search function peaks:

[pks,locs] =

findpeaks(magnitude_f,'SORTSTR','descend','MINPEAKDISTANCE',peaks_dist,'NPEAKS',

peaks_count);

At the output we get coordinate the peaks based on these construct a point in three-dimensional space (Fig. 5.22) the right side shows at which area is searched peaks.
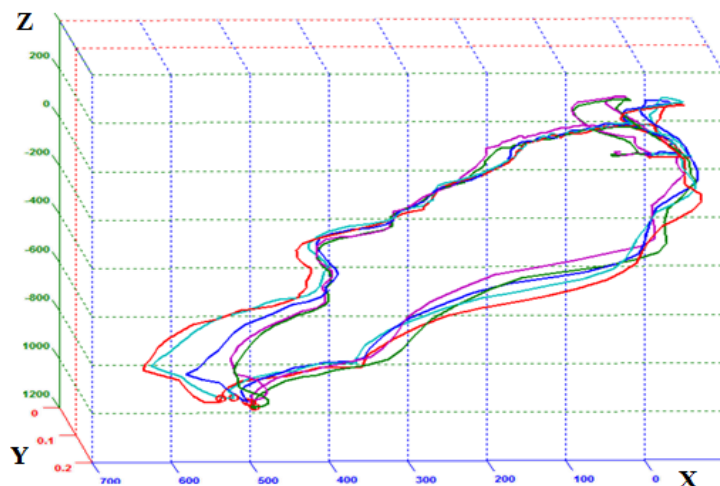

Figure 5.22 - Visualizing of cluster Q peaks in the EKG graph

Thus, around a point Q can be seen inflection. This algorithm allows to fine tune find peaks on the graph by varying the input parameters, such as number of peaks, the distance between peaks and peaks search section. For example, based on their dispersion center peaks can predict cardiac arrhythmia.

In this section, methods of visualization additional data signal and 3D EKG graphically were obtained. Based on results visualization necessary further remove additional ECG parameters and present them in the form of data for the purpose of data mining.

6. Determination of signal patterns in three-dimensional representation of ECG signal

Pattern is a regularity in graph, repeated from time to time in a particular area. Analyzing the frequency and deviations can be detected certain abnormalities of the heart, such as ventricular tachycardia; arrhythmia; myocardium; etc. [16]. Figure 6.1 shows the patterns classified model developed in this work. As can be seen from the graph occurs beat offset relative to each other, however the patterns are arranged close to each other. On the graph can be seen that one of the clusters has gone beyond the region. Concerning this can assume that recovery from the S peak of one of the beats signals it is not typical behavior. With such approach can to recognize most thin changes in graphics behavior both in time domain and in multidimensional that will be described further.
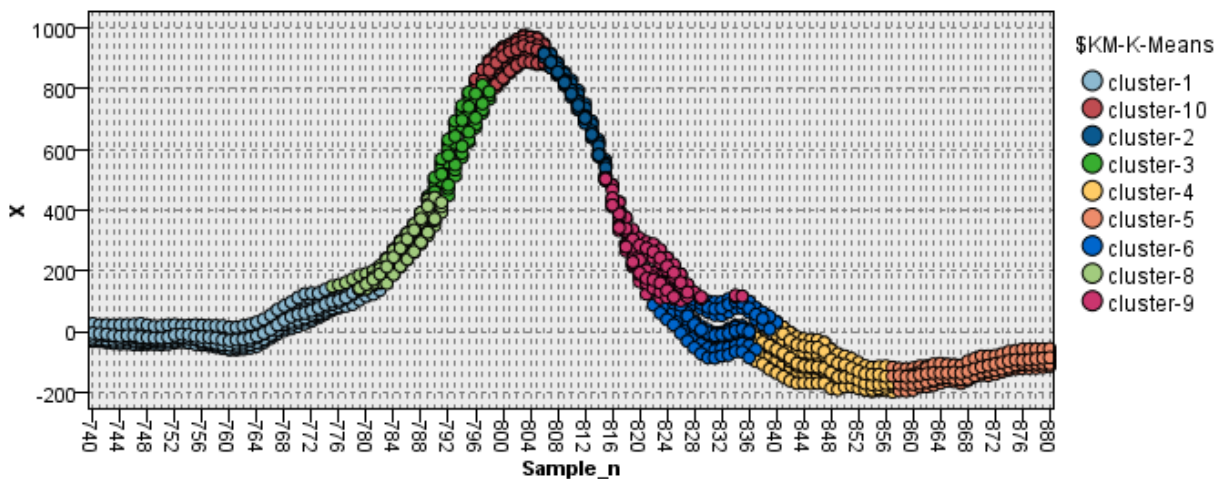

Figure 6.1 – Visualization of patterns in a QRS complex

6.1 Methods for determining patterns in electrocardiographic signals

There are many methods and software for finding patterns in ECG they are increasingly distributed to the medicine and easier recognition of a diagnosis and improve its accuracy. However, all the methods solve the problem of data mining using a particular model [3, 4].

For example, the finding patterns based on wavelet transformation, neural networks, decision trees and others. An example of a methodology with the use of decision tree is shown in Figure 6.2.
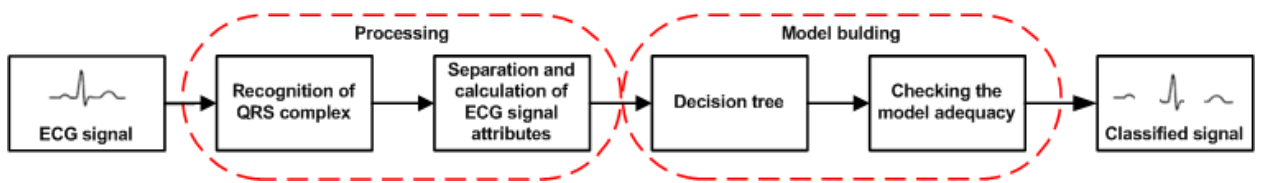
Figure 6.2 - Block diagram of find patterns using decision tree

The first step is to immediately identify the beginning and end of the QRS complex and based on which performed classification. The second and the most difficult is the allocation and calculation of the beats attributes, these include the amplitude, offset, width, etc. of the complex. (To be described later). The third step is to train a decision tree, which will be use in classification task. And the final step - assessment of model adequacy it is based on data obtained from the teacher. This can be realized through use methods active learning with teacher.

Figure 6.3 shows an example of patterns recognition in the time series using algorithm of hierarchical classification. The first graph given normal human eye visible signal change. On the second graph change invisible but with the increase becomes apparent. Data mining methods allow working with such tiny changes which one of the theories are indicators of subsequent abnormalities of the heart and its chemical composition [5].
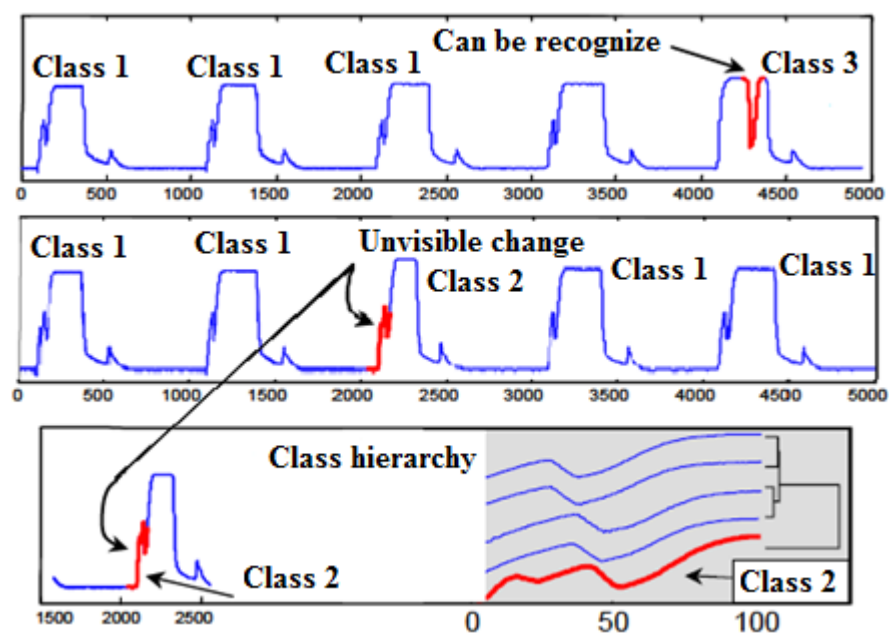

Figure 6.3 - Example of finding clusters in ECG graph

6.2 Model description patterns definition

Based on these examples necessary to develop a find model EEG patterns in three-dimensional domain [5, 6].

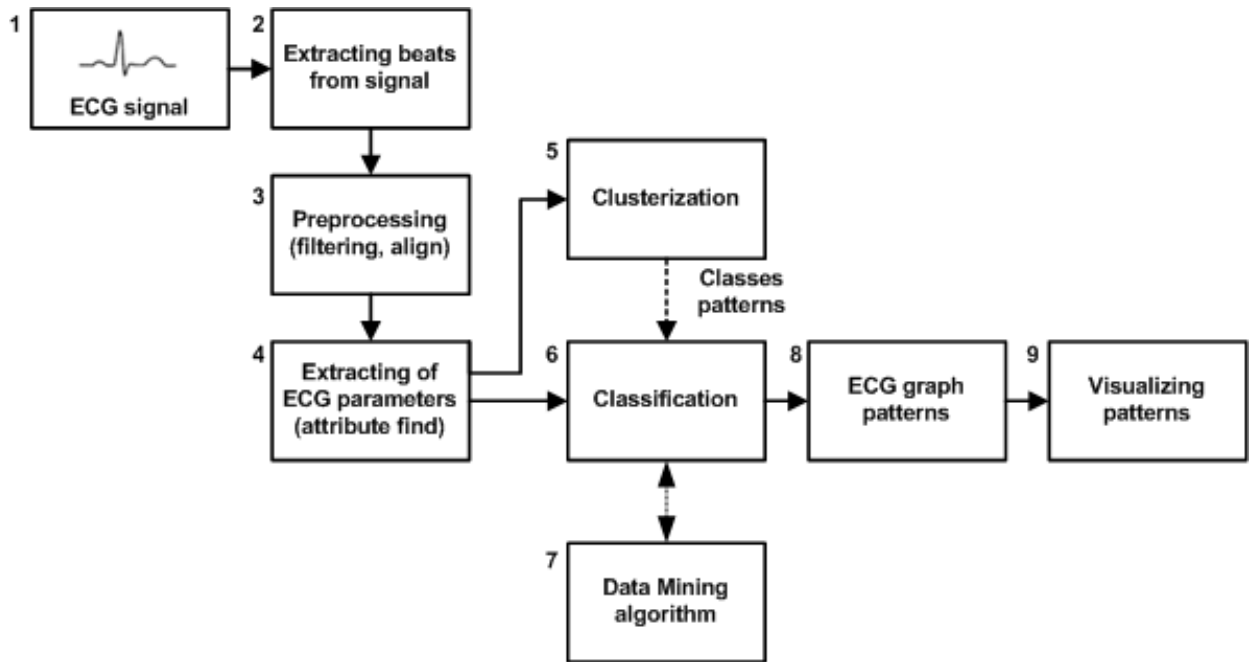Figure 6.4 shows a block diagram of find patterns.

Figure 6.4 - Block diagram of find patterns

The block diagram consists of the following steps:

1) Recording ECG directly using a device, this step is described above;

2) Extraction of beats and segmentation.

There are many methods of segmentation which have their accuracy such as:

- Hidden Markov model;

- Wavelet decomposition;

- Envelogramm;

- Algorithm of dynamic transformation.

3) Data preprocessing

Preprocessing consists of several phases: filtering and alignment.

- Filtering step performed by filtering low frequency (2 Hz) and high frequency component (20 Hz). The low-frequency distortion caused by uneven of contact with skin, changes in capacitance change of HEA position. High-frequency body movements caused by muscle spasms and others.

Filtering performed bandpass filters with constant parameters, the transfer function that does not depend on the parameters of the input signal and the signal changes over time [2].
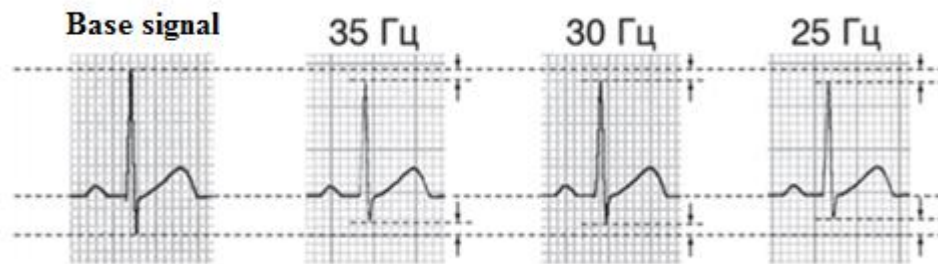

Figure 6.5 - Influence filters on ECG signal

Any filter makes an error in ECG signal. Figure 6.5 shows the graphs with different cutoff frequency parameters the graphs show a reduction in cutoff frequency, the amplitude decreases which may lead to incorrect decoding of ECG and incorrect diagnosis. In these cases should be applied adaptive filtering where filter parameters are dependent from input signal based on the least squares algorithm and recursive least squares algorithm [6, 7, 8].

Primary filtration is already built into ECG recording device for industrial frequency of 50 Hz and 60 Hz and also additional filters HF and LF.

4) Extraction parameters and search attributes

At this stage collect greatest amount of information about object. Here there is a separation in data mining task of time series and multidimensional arrays. The first method uses a time signal attributes and additional attributes. The second case - is abstraction from time series and representation of the signal as a multidimensional array of data with the main attributes - coordinates of points and additional attributes obtained during 3D EKG analysis that will be described under the following part.

5) Clusterization

Step clustering needed to train the model without a teacher. Since initially unknown classes and their number, it is necessary to apply the clustering, thus it is

necessary identify the number of classes having the greatest distance between clusters.

6) Classification

It is the following after clustering algorithms based on data mining. Its task to predict with a certain probability based on the input attribute class assigned to the previous clustering stage after studying a particular mining model.

7) Data mining Algorithm

At this stage make selection of necessary algorithm, its control and training. Most the best results as practice showed, at this task give trees of decisions. The best results give decision trees.

8) Graph Patterns

At the output get ready to file with information about class definition ECG sections.

9) Visualizing patterns

In purpose determining practical application of patterns necessary to produce their visualization. The most different areas in the graphs visible through clusters color codification.

6.3 Description and attributes preparation for data mining

The main attributes of the model are directly values of the signals, sample number, beats number and patient number. To improve algorithms accuracy and comparing work with time and additional multidimensional data series necessary attributes characterizing signal with other side.

Look at each additional attribute separately:

1) The Euclidean distance:

The most common attribute - this measure similarity between graphs. Relatively middle graphics, necessary to calculate the similarity of each beat to the middle beat. Calculation of similarity - it is calculation of the total distance between graphs [6].

The Euclidean distance is measured as the distance from points (Fig. 6.8) which are relative to each other and get array of distances with reference to the number of points:

$$D(C,Q) = L_p(C,Q) \quad (14)$$

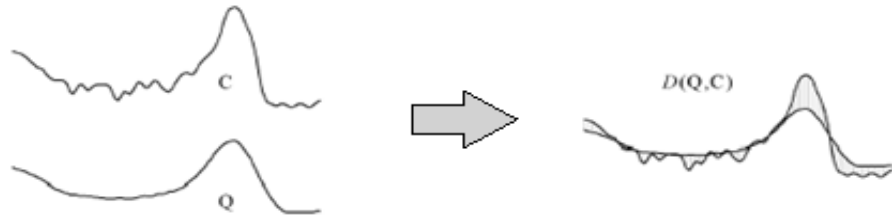where L_p - normal line connecting two points

Figure 6.8- calculation of Euclidean distance

This representation of the difference between graphs easy to understand, but a number of difficulties associated with the shifting of graphs (Fig.6.9) with respect to each other.

In the first case this offset by axis Y for algorithm application is necessary to make the alignment operation but there are situations when curves shape also differ significantly and alignment is not possible to produce.

The second case is when the graphics are shifted along axis X and it is necessary to make the normalization graphs by shifting relative to each other:

$$c_i' = \frac{c_i - \mu(C)}{\sigma(C)} \quad (15)$$

where $\mu(C)$ - the average deviation; $\sigma(C)$ - the current deviation.

Then the number of elements of first graph $C = \{c_1, \ldots, c_n\}$ is replaced by the graph c_i'.
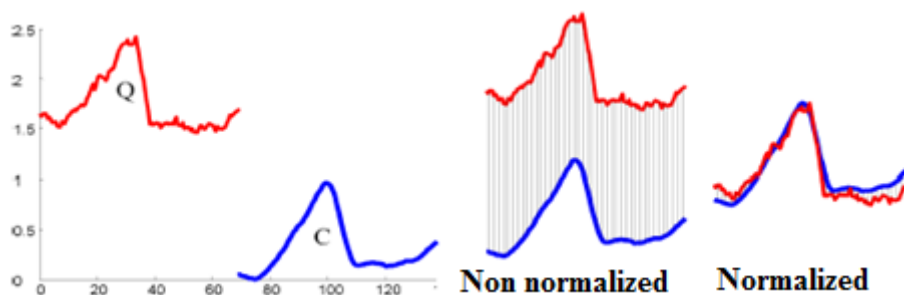

Figure 6.9 - Influence of bias on the calculation of Euclidean distance

The most appropriate and which takes into account all of above Euclidean distance disadvantages (Fig. 6.10, a) is an algorithm dynamic time warping (DTW) (Fig. 6.10, b).

After algorithm operation obtain the matrix matching points of one graphic to another and on basis of which it is possible to calculate the distance between the corresponding points. This algorithm allows to take into account the extension and compression graphics its shifting and over gain.
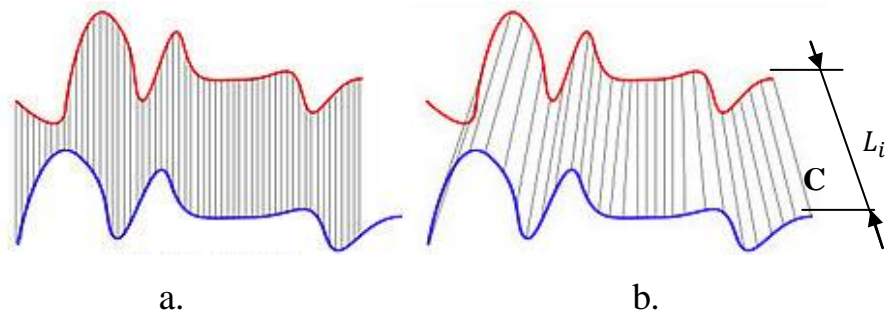

a.   b.

Figure 6.10 - Comparison of approaches for calculating distances: a - Euclidean distance; b - algorithm DTW

To use this of algorithm necessary decide with respect to which beats calculate deviations. The most appropriate option is to calculate the deviation relative to the mean graph.

Which calculation is carried out according to:

```
for (i = 1:len(2))
    sum_x = 0;
    sum_y = 0;
    sum_z = 0;
    for (j = 1:len(1))
        sum_x = sum_x + XYZ(j,i,1);
        sum_y = sum_y + XYZ(j,i,2);
        sum_z = sum_z + XYZ(j,i,3);
    end;
    middle_loop(i,1) = sum_x/len(1);
    middle_loop(i,2) = sum_y/len(1);
    middle_loop(i,3) = sum_z/len(1);
end;
```

where len (1,2) - the number and length of beats.

After applying calculation function DTW get all previously described data.

Further for use the obtained values of matrix w in data mining problem is necessary to transform the ratio beats points in two variables: the connection point and the number of connections, according to:

```
for (i = 1:len(2))
    count_samles = 0;
    for (k = 1:w_len(1))
       if (i == w(k,1))
          count_samles = count_samles + 1;
          DTW_count_3D(i,j) = count_samles;
          if (count_samles == 1)
             DTW_point_3D(i,j) = w(k,2);
          end;   end;    end;   end;
```

2) The distance between samples

This attribute is most simple to calculate and is a distance between points in space:

$$Distance(j,i) = ((XYZ(j,i-1,1) - XYZ(j,i,1))^2+(XYZ(j,i-1,2) - XYZ(j,i,2))^2+(XYZ(j,i-1,3) - XYZ(j,i,3))^2)^0.5;$$

3) Calculation of deviations

This attribute is the Euclidean distance in three-dimensional space and calculated according to:

```
for (j = 1:len(1))
   for (i = 1:len(2))
      Deviation_x = middle_loop(i,1) - XYZ(j,i,1);
      Deviation_y = middle_loop(i,2) - XYZ(j,i,2);
      Deviation_z = middle_loop(i,3) - XYZ(j,i,3);
      Deviation(j,i) = (Deviation_x^2+Deviation_y^2+Deviation_z^2)^0.5;
   end;
end;
```

4) The position of the dynamic vector HEA

This attribute has two variables that determine the exact vector position. According to the vector representation in polar coordinates the vector described azimuth and normal angles calculation of which is as follows:

```
for (j = 1:len(1))
   [normal,Err] = func_normal_vector(XYZ, j, step, win_size, start_point, end_point);
   for (i = 1:(len(2)-win_size-2))
     [az,el,r] = cart2sph(normal(1,i),normal(2,i),normal(3,i));
     n_vectros(j,i,:) = [az,el,r];
   end;
end;
```

5) This is near point of graph to current coordinate XYZ. This attributes consist of 9 values shifted by 4 point to right and left. This attributes needs for estimation of wave form of signal.

All data sets are written to a CSV file, the structure of which is shown below.

6.4 Description of the data mining model

6.4.1 Description of clustering and classification algorithms

For tasks learning without a teacher must firstly obtain the patterns classes. In this case it is necessary to solve the problem of clustering. The simplest algorithm is a clustering is k-means [10].

1) k-means algorithm is an improved EM algorithm consists of two stages Expectation (calculation and expectation of results) and Maximization (selection of results that minimize error). The first step in the algorithm - is the choice of centers. Number of centers established parameters of algorithm (by user). Position of center can be selected random as well as using additional techniques such as constrained clustering or active learning (Fig. 6.11, a).

Then the whole area of points divides between centers according to minimum distances between points and the centers (Fig. 6.11, b).

The next step - counting mass centers where based on an average distance of all points calculated new cluster position (Fig. 6.11 in). The process continues as long as the position of centers will be a constant.
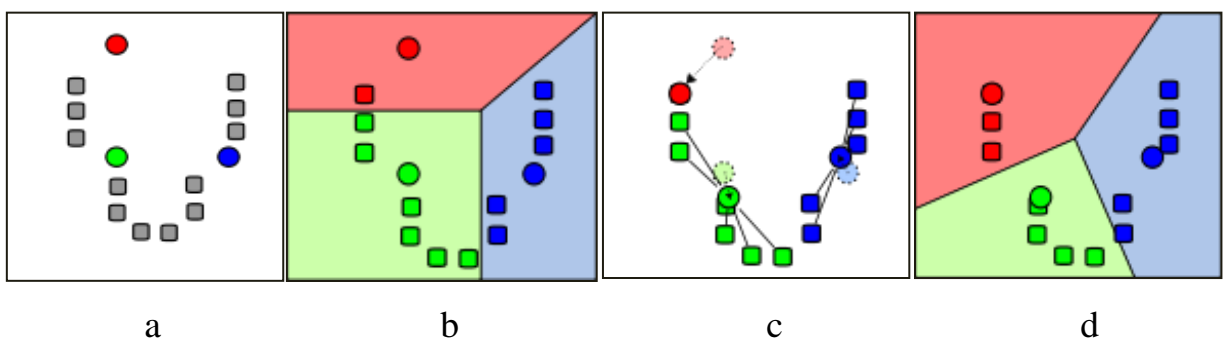

| a | b | c | d |

Figure 6.11 - Operation k-means algorithm

The main disadvantage of this algorithm is: there are several options for splitting points in clusters, the number of clusters is necessary to know in advance.

2) The following clustering algorithm - Kohonen self-organizing maps.

It is a neural network learning without a teacher, is a means of projecting many dimensional data to a lower dimension. Also perform visualization task.

The network structure represents two input and output layers, output is a two-dimensional array. This structure learns the method of successive approximation where is tuned internal network weight coefficients. After several iterations, the neurons become sensitive to certain input data. Thus after clustering defined the so-called neurons - the winner is the cluster number. Learning begins with a large neighborhood from the winner neuron and learning as this neighborhood is narrowed forming a map.

Figure 6.12 - The resulting self-organizing Kohonen maps

Figure 6.12 shows the Kohonen maps two learning iteration. The sensitivity of the neurons shown by hot-map color.

For further classification of data necessary describe work of classification algorithm. Experience has shown that the most accurate results are obtained by making a decision tree C5.0. Decision trees look like a graph that has a hierarchical structure in which peaks are the conditions on which make decisions to move to lower-levels. Figure 6.13 shows a part of graph trained decision tree.

As can be seen from the graph decision based on a comparison attribute with number received during training model. The tree structure may include large amount components that characterize a certain probability of data ownership to a particular cluster.
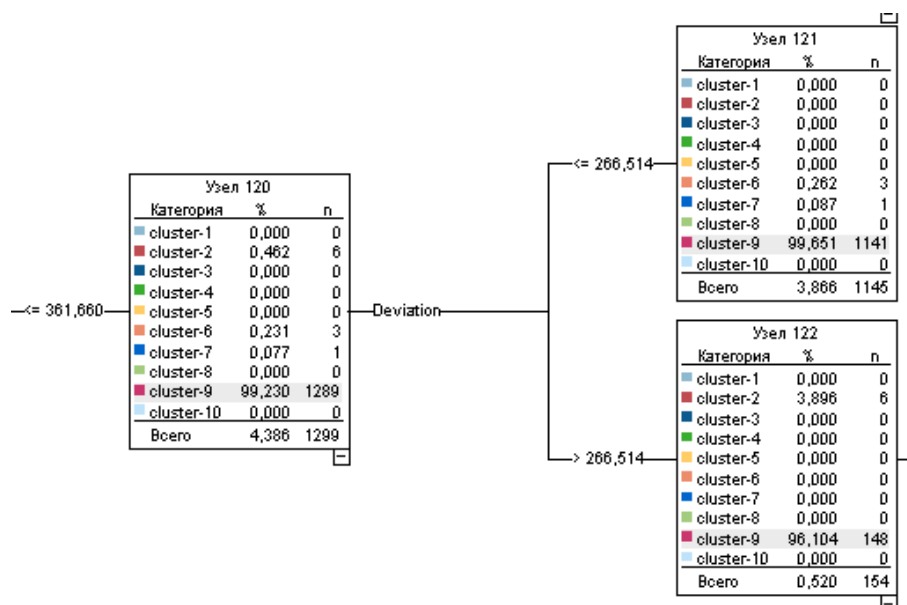
Figure 6.13 - Part of graph decision tree

Feature C5.0 algorithm is working only with categorical data, a high stability degree and the problems of missing data and a large number of them. Usually C5.0 model does not require much time to learn and provides a powerful opportunity for boosting data.

6.4.2 Application of algorithms improving the models accuracy

To improve accuracy of classification models necessary use special techniques to create a powerful classifier. These include cross-validation, boosting, weighted vote.

1) Cross-validation or cross-validation. The principle operation based on splitting the input data sample into subsets for training and control. For example, a first is test subset but second performed training then shifting with each step number of a test subset obtain the average estimate of the error probability. Thus obtain method of comparing trained models.

2) The weighted voting - is easiest algorithm to improve model accuracy. Based on different models and weights is carried out a simple vote in which select a most likely answer.

3) The Busting – method allow to compensate errors of the previous model based on creation set of models. Algorithm of building a model of a powerful

classifier from weak is iterative. The basic idea is to training model from different data parts and with different model parameters. Busting on decision trees is the most effective for classification.

One of boosting modifications is adaptive boosting. A feature of which is training followed model based on misclassified data a previous model.

These approaches to improving the accuracy of the model are not the only ones, but experience has shown that they have the greatest influence on result.

6.4.3 Building of ECG signal analysis models in IBM Modeler

6.4.3.1 Description of the principle of modeling in IBM Modeler

Software environment modeling and predictive analysis IBM Modeler is an extension project Clementine. This environment allows build models of predictive analytics with specialized charts blocks without knowledge of specialized programming languages. The program contains a set of common algorithms, allowing making data mining of time-series to solve the problem of optimizing and analyzing entities.

IBM Modeler contains many elements for building models.

These elements include:

1) Input data blocks - usually a CSV file of a certain format, but also supports the static structure of files, XML files, databases, data collection, web sources, manual entry, and others.

2) Blocks to work with records (rows) - all information in the program represents specific structure of the table, to work with row data, specialized units, these include components: sampling, data aggregation, sorting, merging and adding data exceptions.

3) Blocks to work with attributes (columns) - include blocks that allow work with data columns. These include: automatic data preparation, typing attributes, he exception of attributes, attribute addition, filtering attributes, reclassification, restructuring data, transposition tables and others.

4) Blocks to work with graphics - here can build a combined window graphics, render graphics in two-dimensional and three-dimensional coordinates, histograms and others.

5) Simulation slocks - this section is the largest and represents a set blocks containing the classification and clustering model. clustering models include: k-averages, Kohonen maps, two-step clustering. Classification models include neural networks, linear regression, decision trees, support vector method, the method of k-nearest and others.

6) Blocks for output data - help to evaluate the data at each stage of data mining. Output can be shown as tables, matrices, analysts, and others report.

7) Blocks of data export - data blocks keep and distribute resulting after data mining information. This can be a database, excel file, XML file and others.

6.4.3.2 Building data mining model of ECG signal

Based on blocks necessary make ECG and 3D EKG data mining model.

After processing and preparation information in MatLab program obtain CSV file such structure:

Pacient_n;Sample_n;Beat;Deviation_2D;Deviation_3D;Distance_3D;X0;Y0;Z0;n_vector_normal;n_vector_zenith;DTW_start_point_3D;DTW_count_3D;DTW_start_point_2D;DTW_count_2D;X-5;Y-5;Z-5;X-4;Y-4;Z-4;X-3;Y-3;Z-3;X-2;Y-2;Z-2;X-1;Y-1;Z-1;X+1;Y+1;Z+1;X+2;Y+2;Z+2;X+3;Y+3;Z+3;X+4;Y+4;Z+4;X+5;Y+5;Z+5

CSV file contains the following attributes are described detail in previous part:

1) Pacient_n - patient number;

2) Sample_n - sample number;

3) Beat - beats number;

4) Distance - the distance between points;

5) X - coordinate x;

6) Y - coordinate y;

7) Z - coordinate z;

8) n_vector_normal - angle of the normal vector of the dynamic situation;

9) n_vector_zenith - azimuth angle dynamic position vector;

10) DTW_start_point_3D - point to another point which defines the schedule for DTW TEKG;

11) DTW_count_3D - DTW number of points;

12) DTW_start_point_2D;

13) DTW_count_2D;

14) Neighbor 8 points to central point.

To enter this file you must use the var File block ⬚ and change the type of separator. Then obtain the form table (Fig. 6.14).

Pacient_n	Sample_n	Beat	Deviation	Distance	X	Y	Z	n_vector_normal	n_vector_zenith	DTW_start_point_3D	DTW_count_3D	DTW_start_point_2D	DTW_count_2D
14	1	1	126.591	0.000	-100.377	112.235	-10.222	0.694	0.232	1	12	1	1
14	2	1	123.852	10.536	-97.240	105.153	-3.080	0.495	0.001	13	1	1	1
14	3	1	118.278	7.286	-96.561	99.436	1.386	0.394	-0.177	14	1	1	1
14	4	1	109.774	3.094	-96.647	97.542	3.831	0.362	-0.292	15	1	1	1
14	5	1	106.495	7.740	-96.842	93.409	10.372	0.396	-0.424	16	1	1	1
14	6	1	100.249	4.576	-95.617	92.225	14.619	0.378	-0.407	17	1	2	1
14	7	1	91.512	3.045	-93.191	90.386	14.552	0.358	-0.223	18	1	3	1
14	8	1	87.452	5.757	-90.655	89.366	19.618	0.323	-0.137	19	1	4	1
14	9	1	84.464	6.281	-87.956	85.670	23.920	0.297	-0.062	20	1	5	1
14	10	1	79.697	6.492	-84.546	80.172	24.454	0.353	0.110	21	1	6	1

Figure 6.14- structure of CSV file

Further necessary determine the types of attributes in a file for this need to use a Type block ⬚ and each attribute to match a specific type of data depending on its content (Fig. 6.15).

Also in this section must specify the role attribute: an input that is not used or the target. Target is attribute that must be to predict in data mining problem. In this case the target attribute is not present because solve the problem of learning without a teacher.
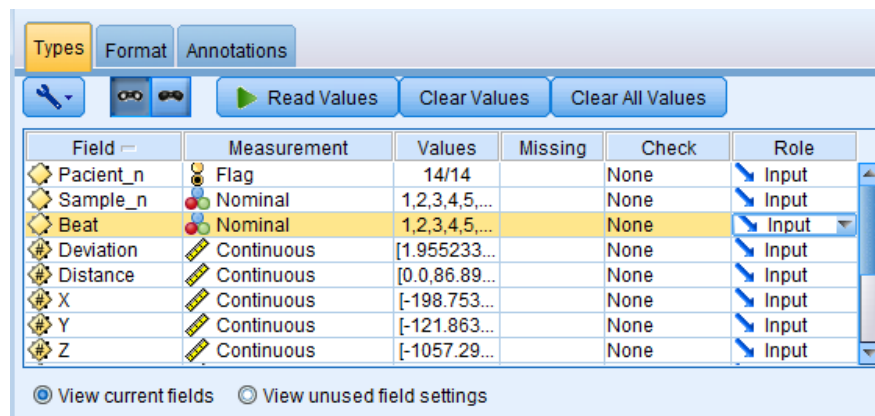
Figure 6.15 - Interface of type block and role attributes

As the clustering algorithm in this project will be used the k-means algorithm and the Kohonen maps. For clustering algorithm needs to know the number of clusters. But initially unknown how many clusters are actually. Thus necessary to use a stable clustering method based on parameters such as division ratio of clusters and size of the largest and smallest clusters.

For this carry out a comparative clustering analysis for different clusters numbers. As input values used XYZ coordinate values and beats number. To exclude other attributes applied block Filter (Fig. 6.16).
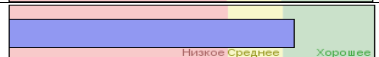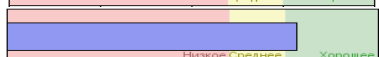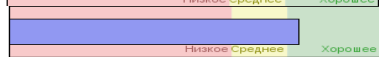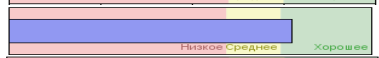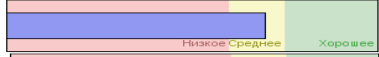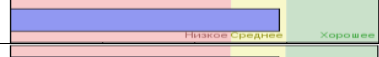
Figure 6.16 - Block attribute filtering

Table 6.1 shows comparative analysis of clusters number according to the size of smallest and the largest cluster as well as the degree of clusters separation.

Table 6.1 Comparison of separation for clustering problem

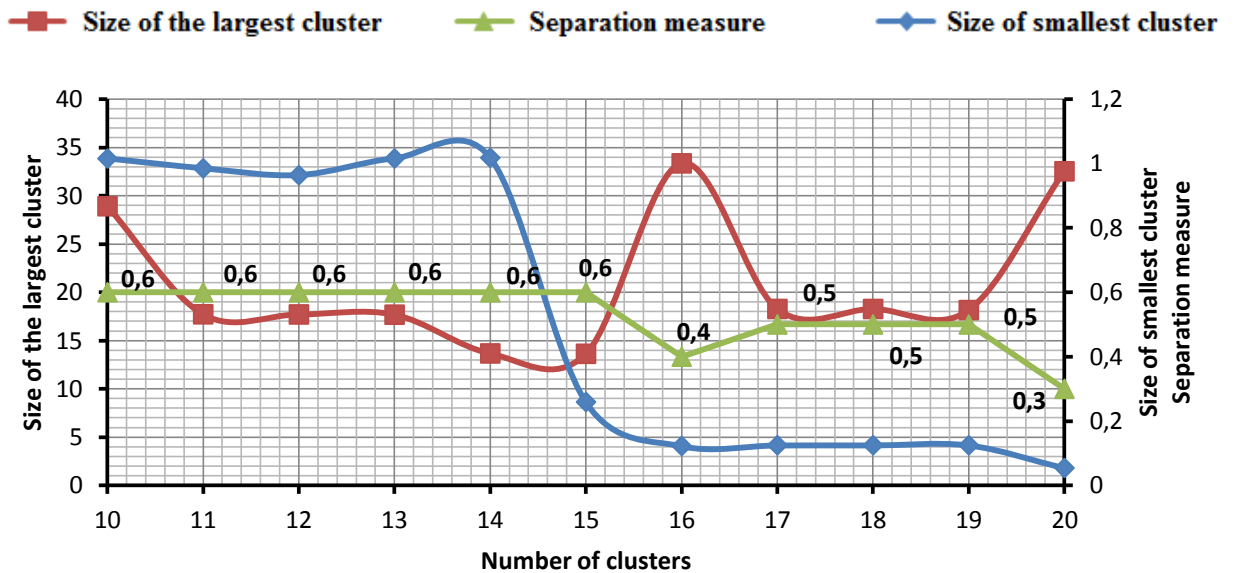| Number of clusters	Size of smallest cluster	Size of the largest cluster	Separation measure
10	431 (1%)	12271 (28,9%)	
11	418 (1%)	7521 (17,7%)	
12	409 (1%)	7515 (17,7%)	
13	431 (1%)	7504 (17,7%)	
14	432 (1%)	5801 (13,7%)	
15	110 (0,3%)	5784 (13,6%)	
16	52 (0,1%)	14154 (33,3%)	
17	53 (0,1%)	7756 (18.3%)	
18	53 (0,1%)	7755 (18.3%)	
19	53 (0,1%)	7705 (18.2%)	
20	23 (0,1%)	13806 (32,5%)	

Figure 6.17 - Graph comparing the number of clusters from the values obtained after clustering

The graph (Fig. 6.17) shows the dependence given in the table. On graph can be seen a sharp decline in the size of smallest cluster graphics and almost simultaneous rise of graphics largest cluster and the degree of separation of clusters same sharply falls down. According to stable clustering method is necessary to choose the number of clusters of 15.

After clustering obtain distribution of clusters as shown in Figure 6.18. Application of Kohonen maps gives a similar result, thus further work will be applied k-means algorithm.

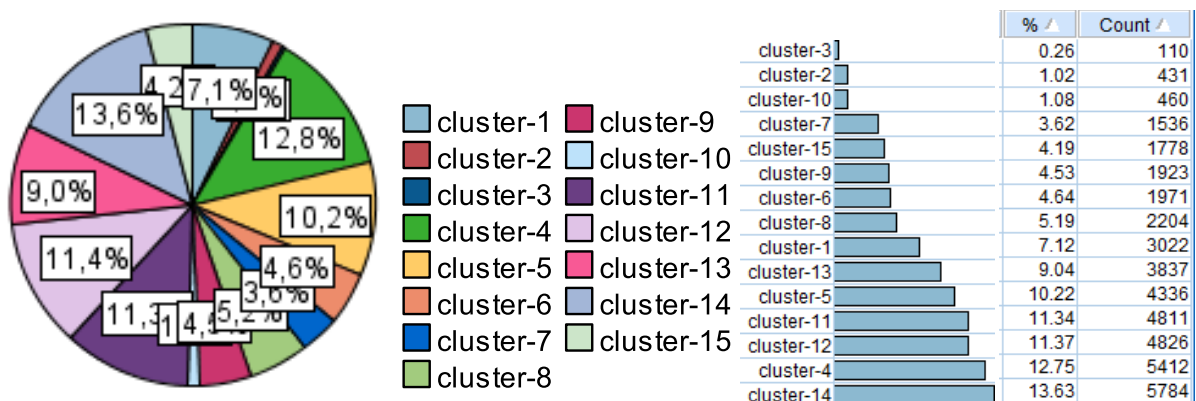|	%	Count
cluster-3	0.26	110
cluster-2	1.02	431
cluster-10	1.08	460
cluster-7	3.62	1536
cluster-15	4.19	1778
cluster-9	4.53	1923
cluster-6	4.64	1971
cluster-8	5.19	2204
cluster-1	7.12	3022
cluster-13	9.04	3837
cluster-5	10.22	4336
cluster-11	11.34	4811
cluster-12	11.37	4826
cluster-4	12.75	5412
cluster-14	13.63	5784

Figure 6.18 - Distribution received clusters

After clustering obtain additional attribute characterizing this cluster. Further the display position of clusters necessary to produce their visualization. For this

54

task necessary to use Plot block. For a two-dimensional visualization necessary specify the series of X and Y and also the color codification according to clusters. Figure 6.19 and 6.20 obtained visualization of graphics.
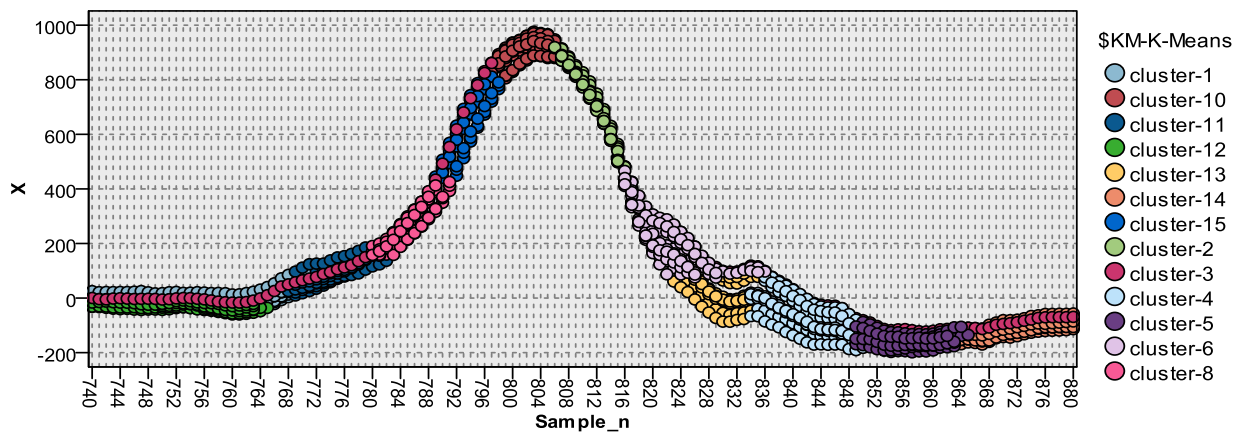

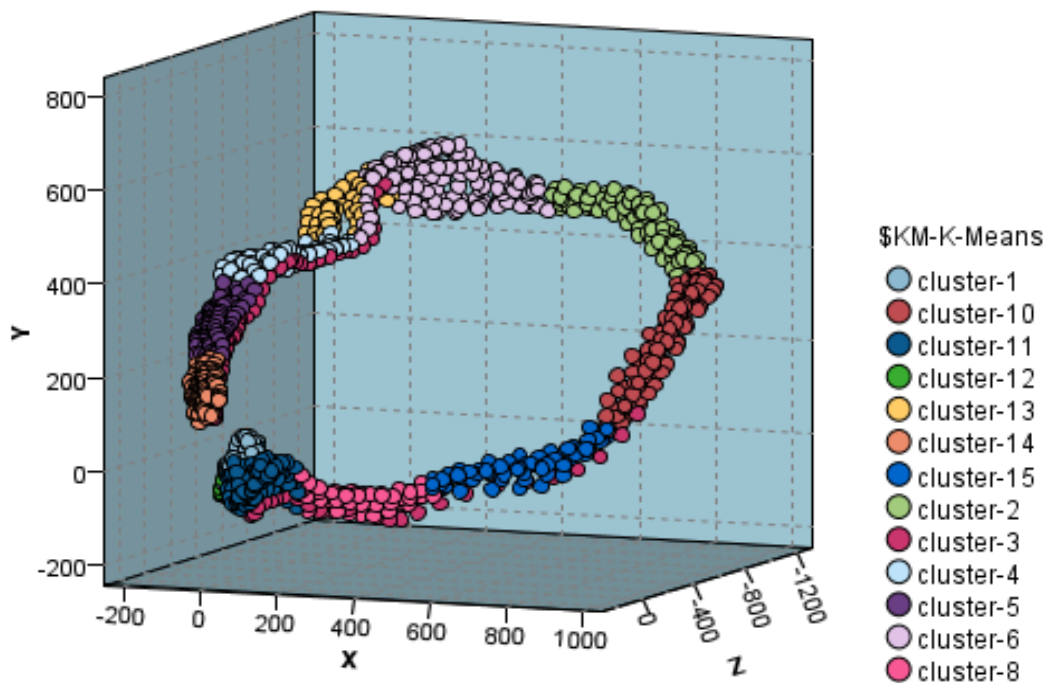Figure 6.19 - Graph X-axis with division by patterns

Figure 6.20 - Visualizing 3D ECG patterns

Further it is necessary to make the training of classification models but is not known what model and what options give the best results. In this case necessary use a block of auto classifier . This block allows to train selected models and to produce their comparison according to predicted values of clusters. Before using this block, all data should be divided into test and training sets. The training data is

used directly for purpose of learning models and test set to test the model adequacy. Optimal distribution between test and training subsets 30% to 70% respectively. Figure 6.21 displays the stage of separation. Also this block contains distribution parameter which can be tune.

Data mining model consist of two parts where use the comparison attributes for analysis of ECG and 3D EKG.
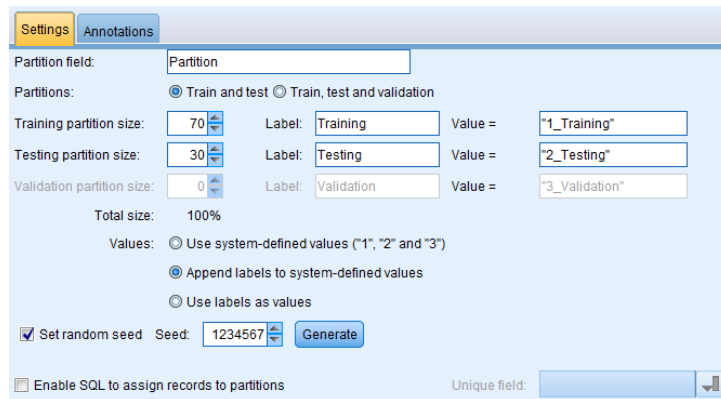

Figure 6.21- interface of distribution block

Further in the block of automatic classification must specify trained models (Fig. 6.22). Select all available models.

Figure 6.22 - Selecting trained models

After training obtain a table with a comparison of models. Automatic classification block automatically selects the best model.

The most precise models are decision tree C5.0 and KNN algorithm. The histograms (Fig. 6.24) shows the distribution right and wrong predictions of these

models. As can be seen most of them was incorrectly predicted cluster 1 and 12, corresponding areas of rear edges after the R and S peaks. However, better result on C5.0 model.

Use?	Graph	Model	Build Time (mins)	Overall Accuracy (%)
☑		C5 1	< 1	97,661
☑		KNN Algor...	16	95,115
☑		Logistic re...	< 1	91,282

Figure 6.23 - A list of trained models

a. b.
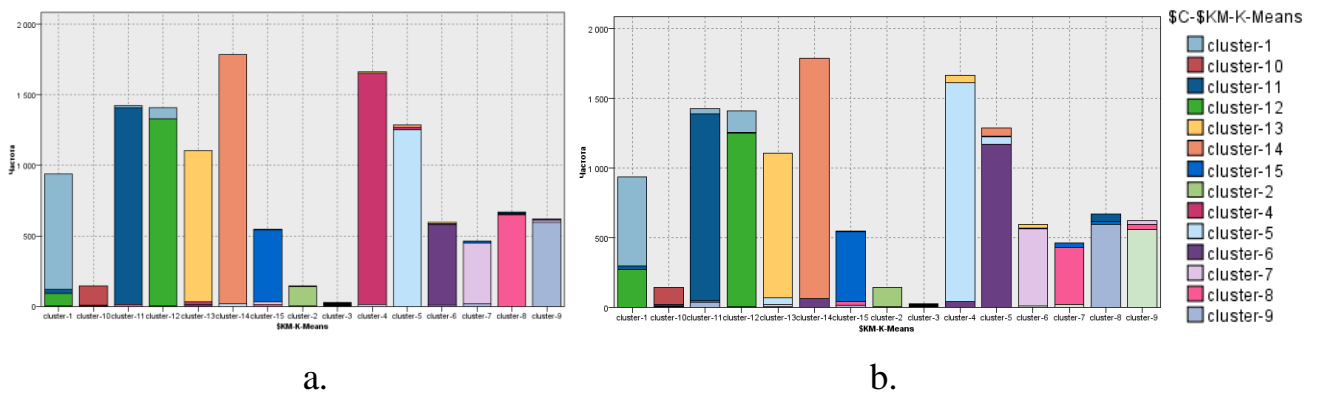
Figure 6.24 - histogram of distribution of clusters: a - model C5.0; b - KNN

For more fine-tuning necessary use a block of the model which was chosen by block of automatically classification for purpose of setting the model parameters. For this instead auto classifier necessary establish model and configure the settings (Fig. 6.25).

To increase accuracy of parameters include cross-validation and boosting, configure parameters and run the model for training.
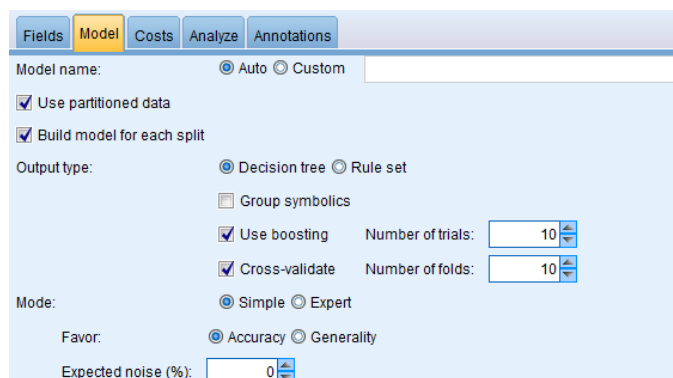
Figure 6.25- C5.0 model block interface

After training obtain a model with the calculated weight coefficients and a noticeable influence on clustering algorithm boosting accuracy (Figure 6.26).

Figure 6.26 - The weighting coefficients of models for boosting algorithm

The most important attribute of this model is relationship attribute TEKG points, namely number of a point to which call is to other points on second chart. For a task of data mining obtain model in figure 6.27.

Figure 6.27 – The model of data mining of ECG signals

After training and classifying obtain the result shown in Table 6.2.

Table 6.2 The result of the classification of 3D ECG and ECG

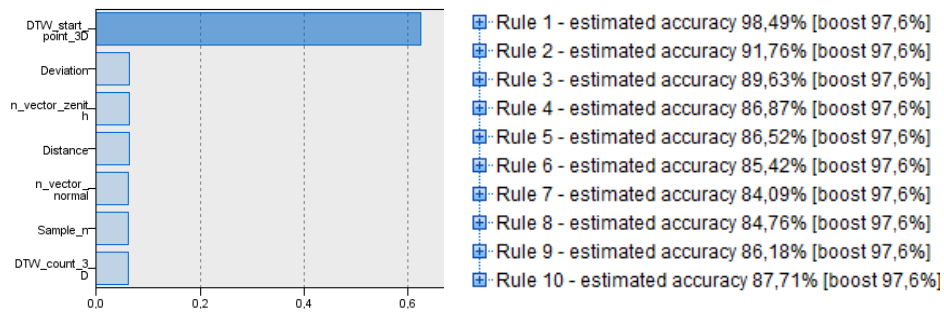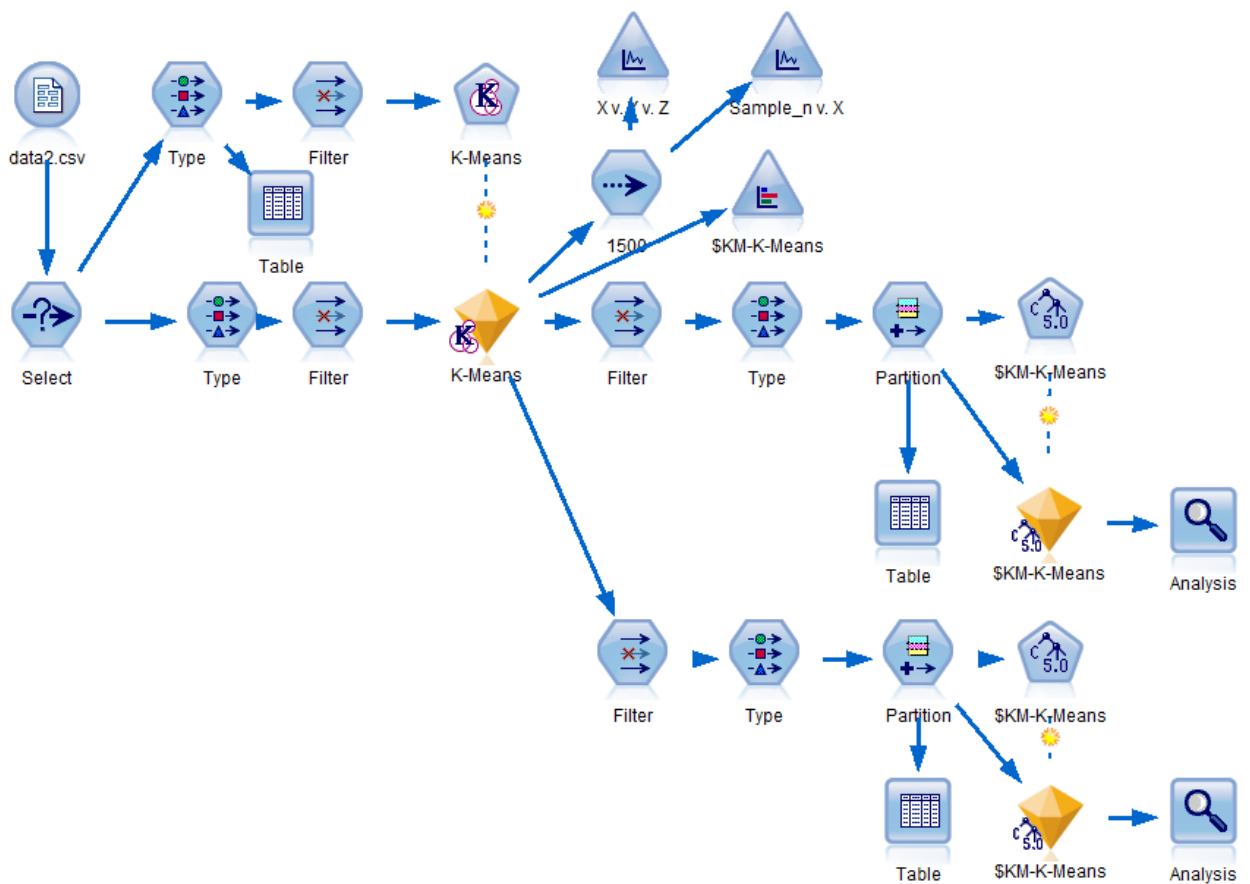| 'Partition'	1_Training		2_Testing	
Correct	29 323	99,01%	12 357	96,35%
Wrong	293	0,99%	468	3,65%
Total	29 616		12 825	

'Partition'	1_Training		2_Testing	
Correct	25 384	85,71%	10 960	85,46%
Wrong	4 232	14,29%	1 865	14,54%
Total	29 616		12 825	

As can be seen from results use of additional 3D EKG attributes compared to attributes used in ECG analyzes improves classification accuracy.

In this section make data mining of 3D ECG and ECG data. The first step is to prepare the data for analysis, the second phase is clustering of data and determine the optimal number of clusters. The last step is to classify data and comparing the results. At the end of this section is a table comparing accuracy of the data mining models.

7. Description of visualization and analysis software

To solve the problem visualization and data analysis necessary to develop software. The most appropriate package is MatLab on which can build a GUI with prepared functions. MatLab software package has a number of benefits required for quality data processing and visualization. One of these features is the wide range of data visualization ways MatLab allows flexible and quite simple to implement this functionality.

7.1 Description of the principles of software development in MatLab

Language programming in MatLab is high-level interpreted language. Programs written in MatLab divided into two types - it functions and scripts. Scripts run in the general workspace and use the function input and output variables and work in the local workspace [9].

Developing software in MatLab is two-stage - a code debug script code and further integration in the function. MatLab function allows save in individual files and drag them from code in code, need only that the files are in the same directory. One of parts of software in MatLab is GUI - separate files which are shaped to interface elements (Fig. 7.1).
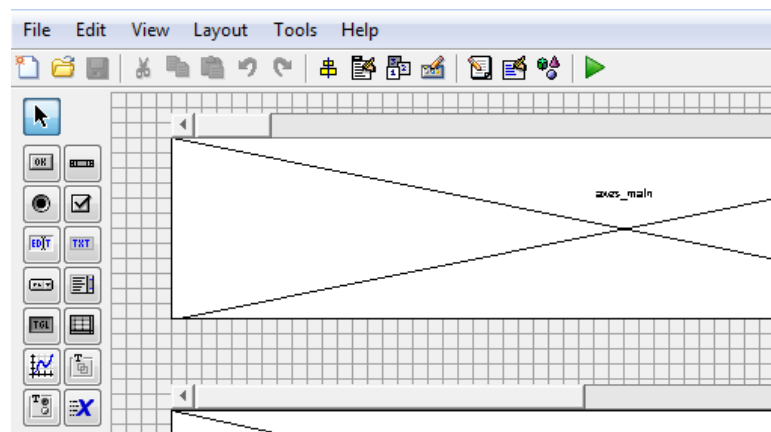

Figure 7.1 - GUI Interface Editor MatLab

To work with these elements used function - response to the action. Work with source code similar to an object-oriented language, but there are features in which the global variables is necessary to use.

7.2 Description of functions visualization and analysis software

Based on tasks and review of existing software were determined software features:

1) rendering of graphs three leads;

2) imaging leads individually selectable;

3) visualization of the magnitude of the signal is determined based on the vector sum of the three coordinates of the vectors;

4) determination of the boundaries of imaging signal;

5) definition of an interactive visualization of the signal window;

6) volume signal display graph with the release of the signal line style;

7) styling lines and color in accordance with the number of beats;

8) display of three-dimensional domain signal of the synchronization rotation axes with the graph of the entire signal;

9) signal mapping projections on the plane;

10) selection of beats visualization table;

11) interactive zooming graph visualization 3D EKG signals;

12) creating the animation of building a signal from the first value on the graph of the 3D EKG signal.

7.3 Description of the program work

The software consists of a main window (Fig. 7.2) where proceed to load file from the original variables and the selection of the graph for its subsequent displaying in three-dimensional space.

The main window consists of a graph visualization ECG signals to select the signal type display element is represented by list which lists the possible types of the XYZ graphs and the magnitude of signals on each axis.

Before opening file user must adjust the boundaries of the signal visibility in the elements. After file opening with a variable XYZ displays some parameters which include the number beats in a variable and a short description of beats that is assigned after process data clustering.
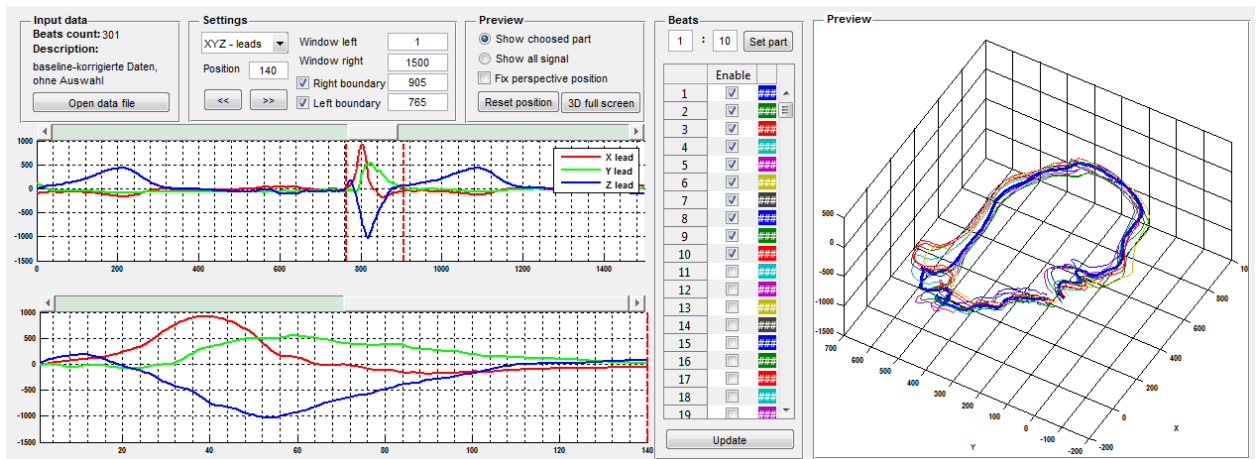
Figure 7.2 - Interface main window of visualization program

After a comfortable displaying graph user must select left and right border of window. What is done by clicking on graph. After that shows below the selected part of ECG and 3D EKG. Then the user can adjust the position of window by moving the scroll bar at top of graph. This visualization is needed for precise choice the position window. On figure 7.3 is selected the QRS complex. To fix the scale of the graph necessary to establish the appropriate check box.

For quick viewing plotting direction on window provided slider displayed in the middle graph. To control the position of slider provides a scroll bar. By dragging this slider to graphs 3D EKG pre-visualization shows the gradual graphing parts of beats (Fig. 7.3).
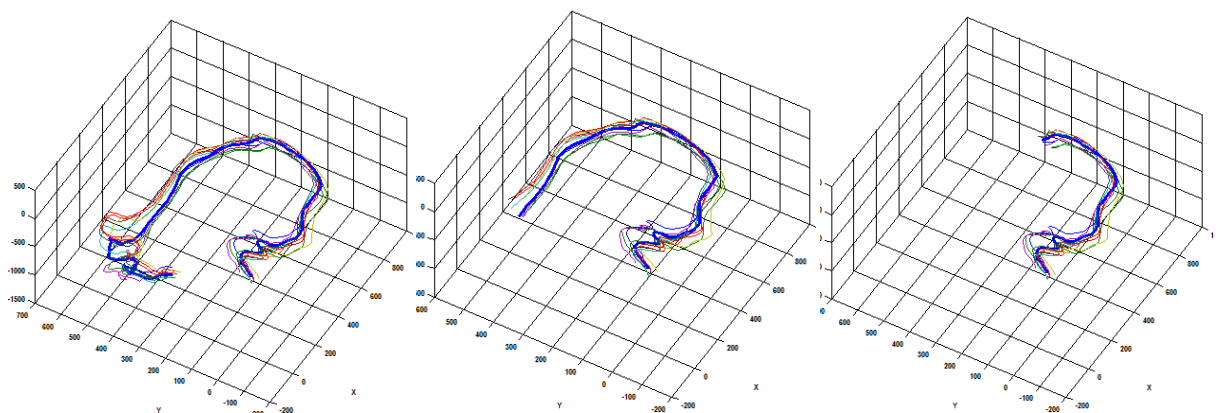

Figure 7.3 - Visualization of 3D ECG graph

To select and verifying necessary amount beats as well as visualization of each beats provided table that displays beats number and its setting by color

displaying. To set array of beats included is enough to indicate number of initial and final beats (Figure 7.4).
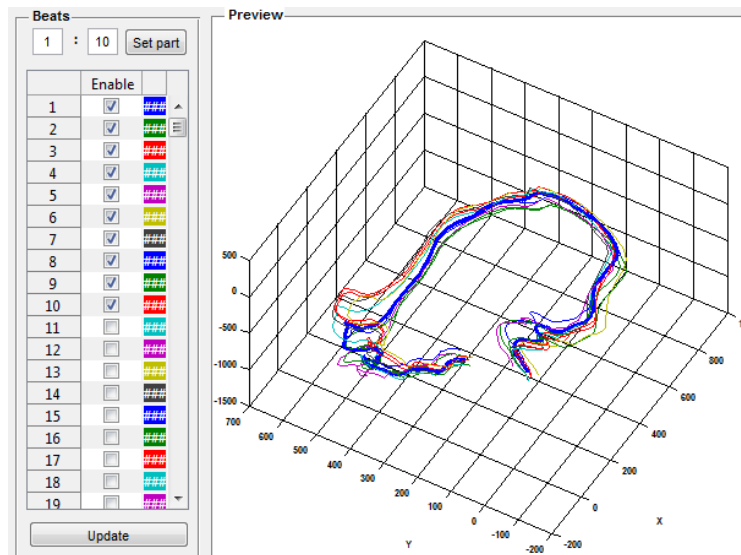

Figure 7.4 - Display list beats and their visualization

After user must specified the position and size window signal necessary to carry out a three-dimensional visualization in convenient format.

7.4 Description of program interface for three-dimensional visualization

For a comfortable visualization was decided to develop a window with automatic opening at full screen and full filling useful space monitor to zoom all the elements regardless of screen resolution.
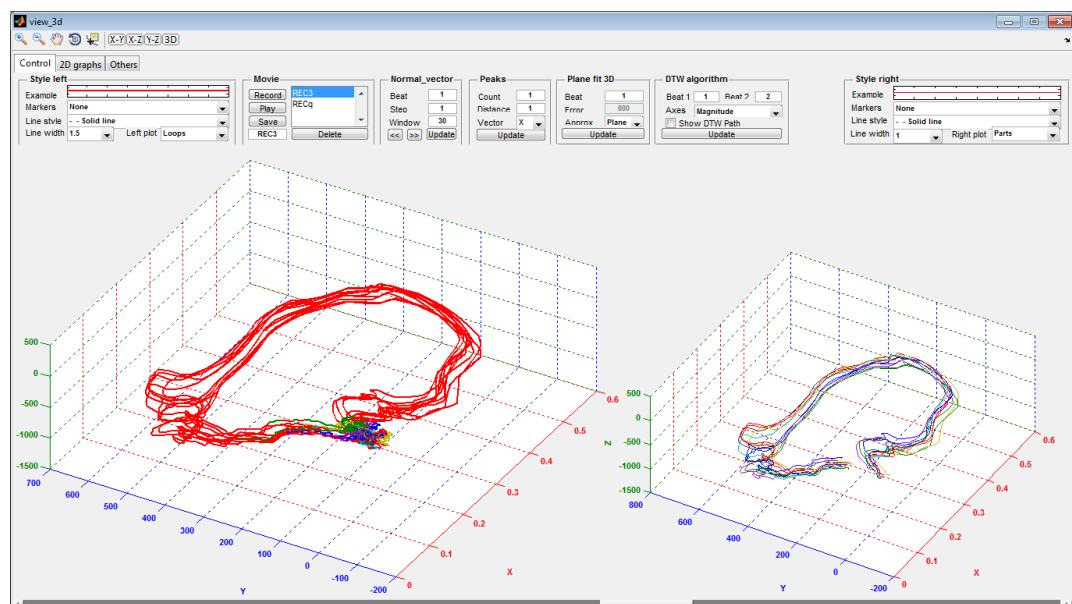

Figure 7.5 - Window 3D ECG signal visualization

Visualization window consists of two main graphs in three-dimensional format. On the left shows a complete picture of the signal which is loop with different kinds of inclusions and noise. To the right is displayed the selected part of the graph (Fig. 7.5).

For convenience, was enabled synchronization of rotation graphs around their axes. For convenience scaling graphs at the bottom of the scroll element is added which allows to change the scale of both charts and fill the space in proportion to relative position of the scroll element.

At the top of window is a panel which contains elements zooming, rotation axes, display positions of points and unusual elements such as projections on the display plane and alignment axes in the initial state (Fig. 7.6).
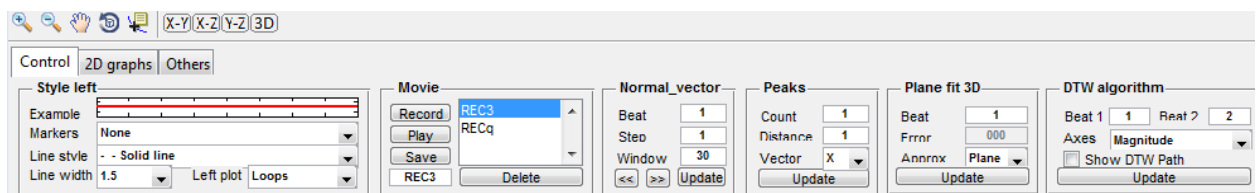

Figure 7.6 – Tools panel of window visualization

To implement function of selection area on the graph in three-dimensional space provided the change line style. On tab bar menu items are located in which is possible to change the type of line - straight, dashed, dotted with a point, change the line color by selecting the color box, change the points markers, and also thickness.

The task of visualization of change of graphs in time is useful for user. User can see tiny changes in signal and visualize speed of magnitude change. This problem is solved by using move function that allows you to record changes in graph when you build it and then play it like a video. In application, this function is presented on panel (Fig. 7.7).
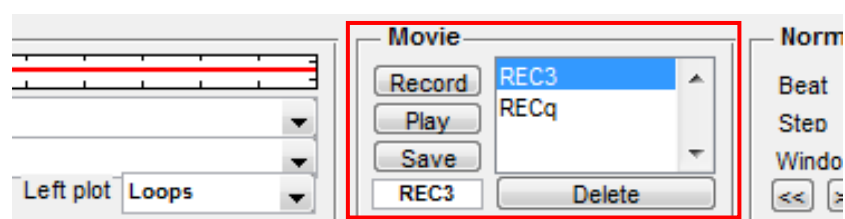

Figure 7.7 - panel of recording and playback

To display other types of visualization provided a list of panels and visualization settings. To visualize the dynamic HEA position vector available settings beats numbers, length step and width of the window (Fig. 7.8)
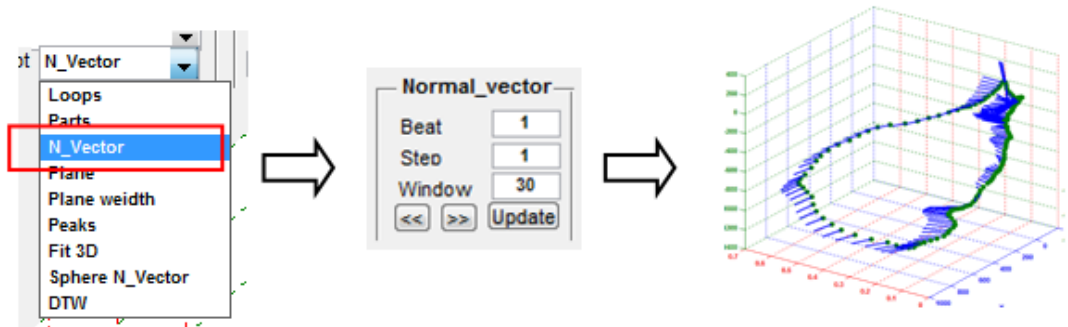

Figure 7.8 - Setting of visualization of dynamic vector HEA
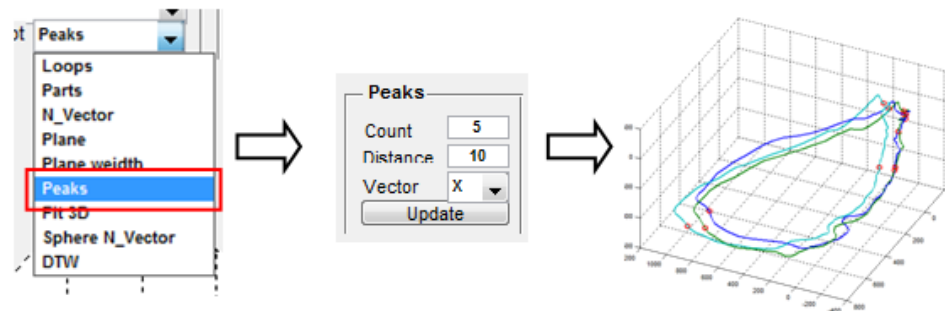

Figure 7.9 - Setting visualization of peaks

For visualization of peaks on the panel are set the number peaks, the minimum distance between peaks and a vector by which to find peaks (Fig. 7.9).
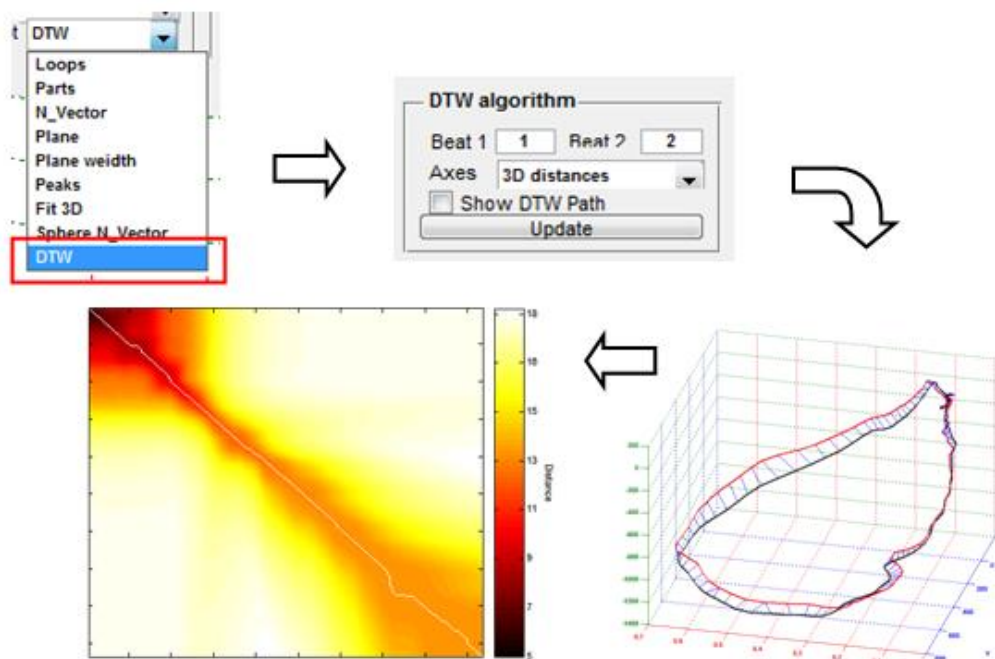

Figure 7.10 - configuration and visualization DTW

For of visualization operation of the algorithm DTW must be configured on the panel (see Figure 7.10) compared numbers beats, the type of calculation of the distances and the ability to visualize optimal path.

The program allows you to visualize well as other previously described visualization techniques. The source code given in Annex A.

Developing GUI applications in MatLab environment - this is the simplest way to implement user interaction with MatLab functions but as the experience of programming in this environment most difficult and awkward to implement large projects with lots of elements and functions. Since GUI development environment in MatLab has a limited functional elements. The main advantage of MatLab is an intuitive language to write applications and squeaks quite a functional graphical visualization of charts and work with them which is a significant advantage in working with this software.

In this section describes the development and use of visualization programs 3D EKG written in MatLab package. Considered examples of visualization properties and parameters of the algorithms. The main methods of work in the MatLab package.

Conclusion

In master's thesis have been resolved of visualization tasks and data mining 3D EKG. Namely are the methods to extract additional information from TEKG signal, developed methods for of visualization signals and additional data in the form of graphics and carried out data mining task to find a repeating signal - patterns.

The target was achieved, ECG analysis is to increase efficiency because the user is given an additional and complete information on which to base ECG specified areas with deviations from normal.

List of references

1) Carolina Varon Mining the ECG// Algorithms and Applications– 2005. - P. 209;

2) K.O. Gupta, Dr. P.N. Chatur, ECG Signal Analysisand Classification using Data Mining and Artificial Neural Networks// International Journal of Emerging Technology and Advanced Engineering – 2012. - P. 56-60;

3) Eamonn Keogh, Jessica Lin, Ada Fu, HOT SAX: Finding the Most Unusual Time Series Subsequence: Algorithms and Applications// ICDM – 2005. - P. 50-60;

4) Hongyan Liu, Jiawei Han, Dong Xin, Zheng Shao, Mining Frequent Patterns from Very High Dimensional Data: A Top-Down Row Enumeration Approach// The Society for Industrial and Applied Mathematics - 2014. –P. 280-291;

5) M. Kotas, Robust projective filtering of time-warped ECG beats// computer methods and programs in biomedicine – 2008. - P. 161-172;

6) Marian Kotas, Projective Filtering of Time-Aligned ECG Beats// IEEE transactions on biomedical engineering/ VOL. 51, №. 7 – 2004. – P. 1129 – 1140;

7) E. Mozhenina, D. Nikonorova, I. Antonenko, filtering in electrocardiography // Modern equipment and technologies - 2012. C. 399-400;

8) S. Porshnev MATLAB 7. Fundamentals and programming. Textbook. ISBN: 5-9518-0137-0. M.: Bean. Knowledge Laboratory, 2006 p.320;

9) Michael Steinbach, Levent Ertöz, Vipin Kumar, The Challenges of Clustering High Dimensional Data// New Directions in Statistical Physics- 2004. - P. 273-309;

10) Larisa G.Tereshchenko, Lichy Han, Alan Cheng, Beat-to-beat three-dimensional ECG variability predicts ventricular arrhythmia in ICD recipient // Heart Rhythm Society – 2010. - P. 1606 – 1613;

11) Guenter Breithardt, Michael E.Cain, Nancy C.Flowers, Standart for Analysis of Ventricular Late Potentials Using High-resolution or Signal-Averaged Electrocardiography//ACC Policy Statement /Vol.17,№5 – 2010. - P. 999 – 1006;

12) Michael Steinbach, Levent Ertöz, Vipin Kumar, The Challenges of Clustering High Dimensional Data// New Directions in Statistical Physics- 2004. - P. 273-309;

13) Nikola Bogunovic and Alan Jovic, Processing and Analyisis of Biomedical Nonlinear Signals by Data Mining Methods// IWSSIP – 2010. - P. 276 – 279;

14) Paolo Buono, Aleks Arisb, Catherine Plaisant, Amir Khellab, Ben Shneidermanb, Interactive Pattern Search in Time Series// VDA – 2005. – P. 1-11;

15) Philippe Esling,Carlos Agon, Time-Series Data Mining// ACM Computing Surveys /vol.45 №1– 2012. - P. 12:1 – 12:34.

16) M. Kotas, Robust projective filtering of time-warped ECG beats// computer methods and programs in biomedicine – 2008. - P. 161-172;

17) S. Kovalev, A. Sukhanov, Detection of specific types of patterns in time series based on a hybrid stochastic model // Bulletin of the SFU. Engineering - 2010. - S. 142 - 149;

18) A. Kolomeitseva, G. Mishugova, A. Mule, G. Ryabykh, Application of wavelet transform and Prony method for identifying nutrient signals // Vestnik DSTU / №4 - 2010. pp 455-465;

19) Kychkin A.V., Artemov S.A. Software and hardware solution for energy monitoring system//Proceedings of 6th International conference of young scientists on solutions of applied problems in control and communications, data processing and data analysis. – Perm, PNRPU, 2015. pp. 3-11.

20) S.Artemov, A.Kichkin Energomonitoring moving objects. // Proceedings of the edge scientific and technical conference. Automated control systems and information technology. - 2014. -S.126-130.