SVM Detection of Premature Ectopic Excitations Based on Modified PCA

Stanisław Jankowski[1], Jacek J. Dusza[1], Mariusz Wierzbowski[3], and Artur Oręziak[2]

[1] Warsaw University of Technology, Institute of Electronic Systems
ul. Nowowiejska 15/19, 00-660 Warsaw, Poland
{sjank,jdusza}@ise.pw.edu.pl

[2] I[st] Department of Cardiology, Medical University of Warsaw,
ul. Banacha 1 A, 02-097 Warsaw, Poland
artur.oreziak@amwaw.edu.pl

[3] Military Academy of Technology, Institute of Electronic Fundamentals
ul.Kaliskiego 2, 00-908 Warsaw 49, Poland

Abstract. The paper presents a modified version of principal component analysis of 3-channel Holter recordings that enables to construct one SVM linear classifier for the selected group of patients with arrhythmias. Our classifier has perfect generalization properties. We studied the discrimination of premature ventricular excitation from normal ones. The high score of correct classification (95 %) is due to the orientation of the system of coordinates along the largest eigenvector of the normal heart action of every patient under study.

1 Introduction

The morphological analysis of ECG signal is one of basic non-invasive diagnostic methods [7]. It allows making an assessment of myocardium state. The examination of arrhythmia and sporadically episode detection requires the analysis of long sequences of heartbeats. It corresponds to approximately 100 000 of ECG cycles. Therefore, automatic morphological analysis of Holter ECG recordings can be considered as a useful diagnostic tool.

In [7] we reported the results of neural network and SVM classifiers that enabled fast and efficient detection of premature ventricular and supraventricular excitations with a high score of successful classification. Prior to automatic classification the ECG Holter recordings were preprocessed: filtered, segmented into separated heartbeats. Then we applied to each heartbeat segment the principal component analysis of its covariance matrix. Hence, the description of the ECG signal shape was reduced to two angles of the corresponding principal eigenvector. Consequently, the resulting classifiers were just linear and the number of support vectors were minimal. The advantage of this approach is the graphical presentation of classification on the plane and clear interpretation of results. However, this method of automatic shape recognition required to design a special classifier for each considered patient.

In this work automatic classifiers based on support vector machine (SVM) is presented. The statistical classifiers, as e.g. neural networks and SVMs, require large enough learning set of labelled examples The power of learning set, according to the Cover theorem [2], must be greater than $(2N + 1)$, where N is the dimension of the input space Therefore it is reasonable to apply the principal component analysis[1, 3, 5] for the dimensionality reduction [4]. The computation speed of classification and

J.L. Oliveira et al. (Eds.): ISBMDA 2005, LNBI 3745, pp. 173–183, 2005.

its effectiveness is obtained due to signal compression and particular parameterisation method. For each heartbeat description only two parameters were used. Such a small number of descriptors allow us to apply a training set containing not too many patterns.

In this paper we introduce the modification of PCA approach to ECG morphological classification that enables to design one SVM classifier for the group of patients suffering from the same disease. We attempt to obtain one classifier for all group of selected patients that can perfectly discriminate pathological excitations from normal ones. The classifier efficiency is defined as a quotient of correctly classified patterns to total number of testing patterns. It reaches 95% for the classes of normal as well as pathological heartbeats. Our study is based on the Holter recordings from the I[st] Department of Cardiology, Medical University of Warsaw for a group of patients with arrhythmia caused by premature ventricular excitations.

2 PCA Parameterisation of ECG Holter Recordings

We studied the 3-channel 24-hours Holter ECG signals measured by magnetic type recorder in the I[st] Department of Cardiology of the Medical University of Warsaw. The signals are sampled by specialized hardware system at 128 Hz with 8-bit accuracy and preprocessed by the Oxford MEDILOG Excel 2 software package. The data are stored in the *fdb* format.

Fig. 1. Filtered Holter recordings of 3-lead ECG signal.

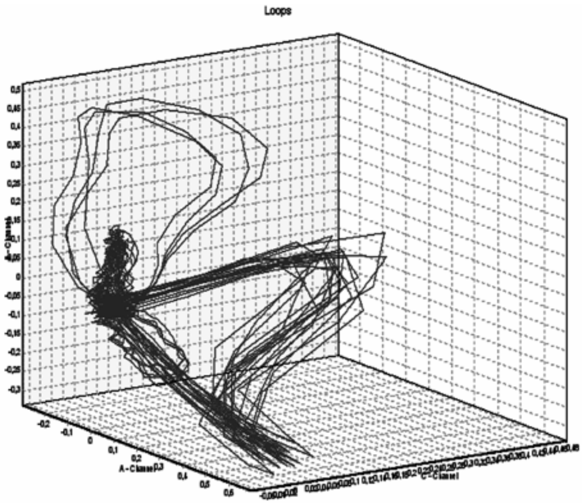

Fig. 2. Trajectory of ECG signal in 3-dimensional phase space reconstructed from signals x(t), y(t), z(t) shown in Fig. 1

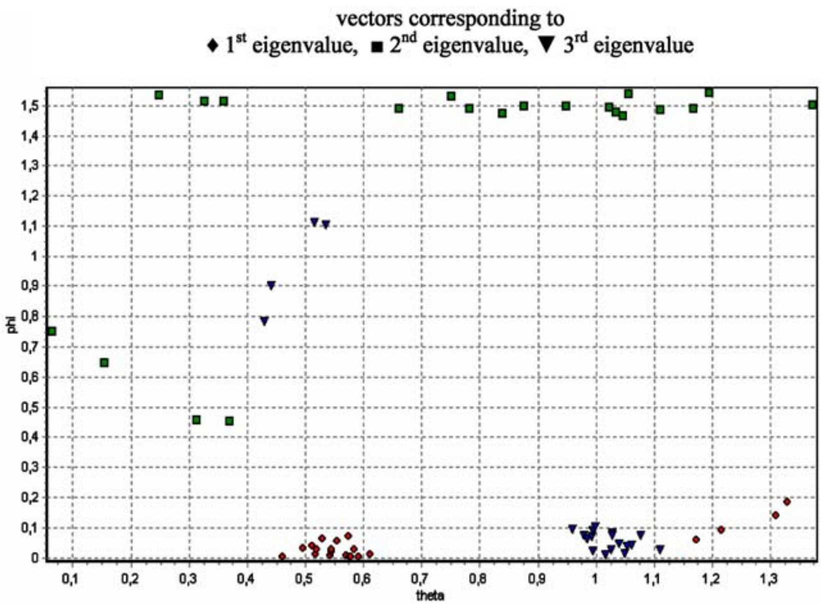

Fig. 3. Angles of eigenvectors in plane-spherical system of coordinates

The signals x(t), y(t) and z(t), shown in Fig. 1 can be treated as quasi-orthogonal components of 3-dimensional trajectory of a hypothetical dynamic system. The corresponding trajectory is reconstructed in phase space as shown in Fig. 2. As can be stated, the trajectory corresponding to a single heartbeat is set up of 3 loops.

The basic steps of our approach are as follows: ECG signal segmentation, calculation of covariance matrix of each signal segment and corresponding eigenvalues and eigenvectors. Thus each segment is represented by 9 numbers: 3 eigenvectors in 3-dimensional space. Taking into account only the direction of eigenvectors in spherical coordinate system reduces the number of shape descriptors to 6 angles in plane-spherical system of coordinates (θ, φ), as shown in Fig. 2. The method described in [7] used the orientation of the sum of 3 eigenvectors, therefore two angles in spherical coordinates were sufficient descriptors of the signal shape.

3 Modified PCA Parameterisation of ECG Signal

Our approach deals with the 3-channel Holter monitoring performing quasi-orthogonal system of coordinates that enables to register the electric potential in myocardium. The idea of presented modification of PCA parameterisation is to introduce a special system of coordinates that is oriented natural for the average normal heart action of every patient under study. The Ox axis of this orthogonal system is determined by the largest eigenvector of the average normal excitation. Hence, in plane-spherical system of coordinates the angles of the largest eigenvectors of normally excited beats are concentrated near the origin of the system coordinates (0, 0).

Suppose that each signal is a function of time x(t), y(t) and z(t). An example of these recordings is presented in Fig. 1. At any moment t_0 three coordinates of points in 3-dimensional space can be calculated. In this way we can reconstruct the trace of the electric vector in 3-dimensional space as a trajectory, as shown in Fig 2. For any single normal heartbeat the trajectory consists of 3 loops. The largest loop corresponds to QRS wave and two small loops represent P and T waves respectively. Single trajectory as a 3- dimensional object can be placed into rectangular prism. The lengths of the edges are proportional to eigenvalues and orientation of rectangular prism depends on eigenvectors of covariance matrix.

We attempt to reconstruct the trajectory of heart electric potential in three dimensional phase space by using signals from each channel as ($x(t)$, $y(t)$, $z(t)$) components. The 3-channel Holter ECG signal can be described by the matrix **S**

$$\mathbf{S}_{3\times N} = \begin{bmatrix} x_1, x_2, \ldots, x_N \\ y_1, y_2, \ldots, y_N \\ z_1, z_2, \ldots, z_N \end{bmatrix} \tag{1}$$

where: N – number of sample points of a given segment of the signal, x_i, y_i, z_i – values of sampled signals from 3 channels.

The evaluating person selects the interval that contains a reasonable number of subsequent normal heartbeats(in practice 20 to 30). This selected part of ECG recording is described by $F_{3\times k}$ matrix (where k is the number of points of a given signal part). Its covariance matrix **K** is equal to:

$$\mathbf{K} = \mathbf{F} \cdot \mathbf{F}^T \tag{2}$$

The eigenvalues λ_i and eigenvectors $\mathbf{w}_i$ of matrix **F** define a new matrix:

$$\mathbf{S'} = \mathbf{WS} \tag{3}$$

where matrix **W** is set up of rows equal to eigenvectors of matrix **K**.

This operation is equivalent to projection of a given trajectory into the coordinate system that is oriented along the largest eigenvector of the average normal excitation of the heart.

The orientation of the largest eigenvector (called principal component) is relevant to the shape of each beat. In order to improve the classification we calculate the orientation of each principal component with respect to the average orientation of the principal components corresponding to normal beats. The SVM classifier is used to discriminate the normal and abnormal beats upon the relative orientation of principal components.

The modification of the classical approach is aimed to define a new orthogonal system of coordinates for the average trajectory of normal beats so that the principal component (the largest eigenvector) is parallel to Ox axis or to (0, 0) point in plane-spherical system.

Thus we obtain compressed information about the signal energy (the elements on a diagonal are proportional to the square of RMS values of each channel) and correlation between the pairs of signal (the rest of elements of matrix **C** are dot product of every pair of signals) is not lost. The elements of covariance matrix are averaged over time interval.

The projected components are shown as x'(t), y'(t) and z'(t) and the corresponding trajectory is shown in Figures 4 and 5. Fig. 6 presents the eigenvector in plane-spherical coordinate system.

The product describes the projection of the ECG signal trajectory onto axes determined by eigenvectors of the selected interval of normal heat action. Thus we obtain a new 3 orthogonal components x'(t), y'(t) and z'(t) of the signal S, shown in Fig. 1. These signals and the trace of transformed 3-dimensional trajectory are presented in Fig. 4 and in Fig.5.

The ECG Holter recording is subjected to segmentation into intervals corresponding to single heartbeats. Every cycle is related to R wave of ECG and contains 30% of $R_{n-1}R_n$ interval and 70% of R_nR_{n+1} interval. The covariance matrix of *k*-th heartbeat is equal:

$$\mathbf{C}^k = \frac{1}{N}\mathbf{S}'^k \cdot \left(\mathbf{S}'^k\right)^T \tag{4}$$

where N – number of sample points.

Then we calculate eigenvectors of these covariance matrices corresponding to subsequent heartbeats. The results are shown in plane-spherical coordinate system. As can be seen from Fig. 6, the angles of the largest eigenvector of normal heartbeats are concentrated in the vicinity of point (0, 0) while those corresponding to pathological ones are significantly distanced from the point (0, 0).

We applied the linear soft margin support vector machine classifier. The input of the classifier has the form of feature vector $\mathbf{x} = (x_1,...,x_n)$ (column vector) and its output is real-valued function f: $X \subseteq \mathrm{R}^n \to \mathrm{R}$. If $f(\mathbf{x}) \geq 0$ the input $\mathbf{x}$ is assigned to the positive class and otherwise to the negative class.

Linear classifier in general form can be expressed as

$$f(\boldsymbol{x}) = (\boldsymbol{w} \cdot \boldsymbol{x}) + b = \sum_{i=1}^{n} w_i x_i + b \tag{5}$$

where: $(\mathbf{w}, \mathrm{b}) \in \mathrm{R}^n \times \mathrm{R}$ - parameters that control the function f. $\mathbf{w}$ - the weight vector, b - the bias (threshold).

The learning paradigm says these parameters must be learned from the data. The decision rule of classification is

$$h(\mathbf{x}) = \mathrm{sgn}(f(\mathbf{x})) \tag{6}$$

This classifier has the natural geometric interpretation. The input space X is divided into two parts by the hyperplane defined by the equation:

$$(\mathbf{w} \cdot \mathbf{x}) + b = 0 \tag{7}$$

which divides the space into two half spaces which correspond to the inputs of the two distinct classes. The vector **w** defines a direction perpendicular to the hyperplane, the value of *b* is the distance of the hyperplane from the origin.

In order to separate a training set with a minimal number of errors we introduce some non-negative variables $\xi_i \geq 0$ (slack variables). The Lagrangian of the data set is equal

$$L(\mathbf{w}, b, \xi, \boldsymbol{\alpha}) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + \frac{C}{2}\sum_{i=1}^{l} \xi_i^2 - \sum_{i=1}^{l} \alpha_i [y_i((\mathbf{w}_i \cdot \mathbf{x}_i) + b) - 1 + \xi_i] \tag{8}$$
$$\alpha_i \geq 0$$

where: **w** – weight vector, b – bias, C –regularisation term, ξ -slack variable, α - Lagrange multipliers, l – number of examples

The Lagrangian L has to be minimised with respect to the primal variables **w** and *b* and maximised with respect to the dual variables α_i - a saddle point has to be found.

Then the weight vector $\mathbf{w}^*$:

$$\mathbf{w}^* = \sum_{i=1}^{l} y_i \alpha_i^* \mathbf{x}_\mathbf{i} \tag{9}$$

realises the maximal margin hyperplane with geometric margin $\gamma = 1/\|\mathbf{w}\|_2$.

In this expression only these points are involved that lie closest to the hyperplane because corresponding Lagrange multipliers are non-zero. These points are called support vectors.

Usually there are only few support vectors in the training set hence, the information compression property.

The fact that only a subset of the Lagrange multipliers is non-zero is referred to as sparseness and means that support vectors contain all the information necessary to construct the optimal separating hyperplane. The fewer number of support vectors the better generalisation can be expected. This property does not depend on the dimension of the feature space.

In our case the input space is just a plane and the linear classifier takes the form of an optimal separating line.

4 Experimental Results

Holter electrocardiography are clinical routine examinations that produce a large amount of data. Monitoring of the electrocardiogram during normal activity using Holter devices has become standard procedure for detection of cardiac arrhythmias.

Ambulatory electrocardiography was carried out using the Oxford Medilog MR 45 ECG recorder. The subjects with the history of myocardial infarction and heart failure post were encouraged and advised to undertake their usual daily activities except bathing. They were also advised to note the time and details of any symptoms perceived (the event diary).

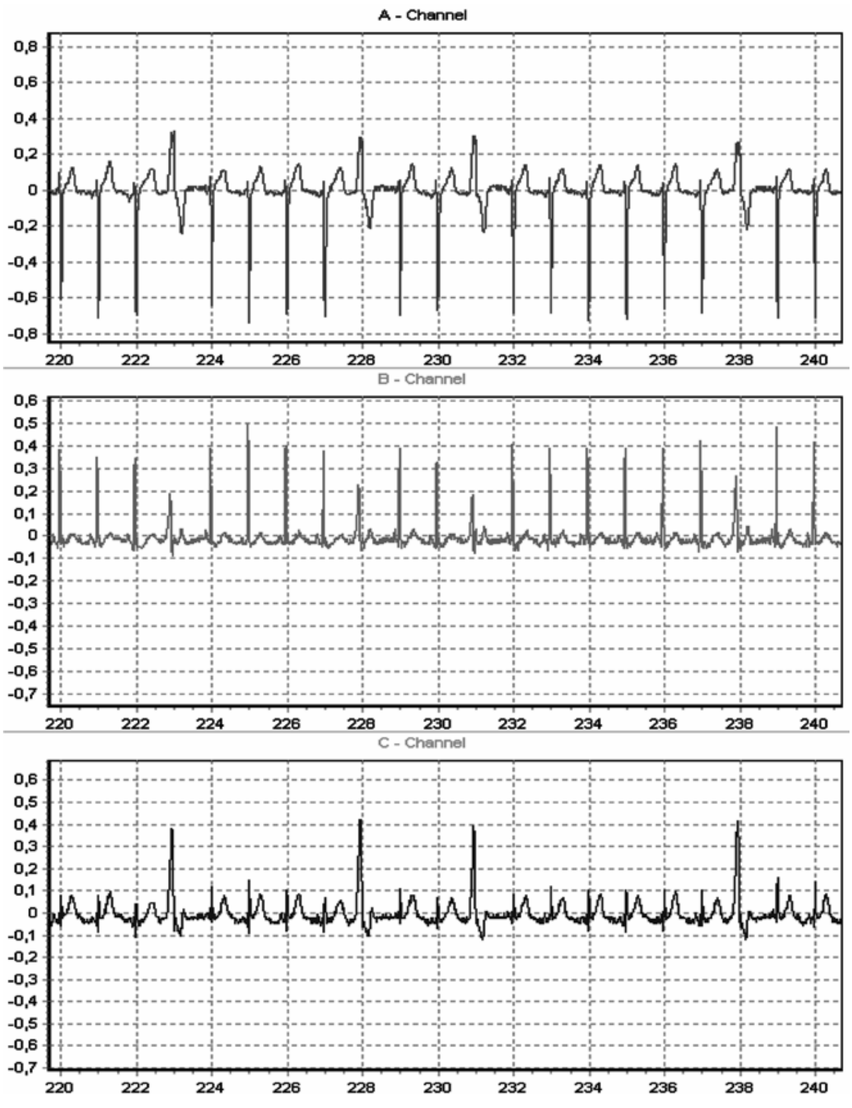

Fig. 4. Three components of the ECG trajectory projected into orthogonal system of coordinates

The 24-hour data was analysed by Medilog Excel 2 Holter Management System that divided the cardiac arrhythmias in supraventricular arrhythmias (supraventricular extasystoles, supraventricular couplets, supraventricular triplets, supraventricular bigeminy, supraventricular trigeminy, supraventricular tachycardia) and ventricular arrhythmias (ventricular extrasystoles, ventricular couplets, ventricular triplets, ventricular bigeminy, ventricular trigeminy, ventricular tachycardia and R on T phenomenon). The decisions of Medilog Excel 2 system were verified by the physician-cardiologist.

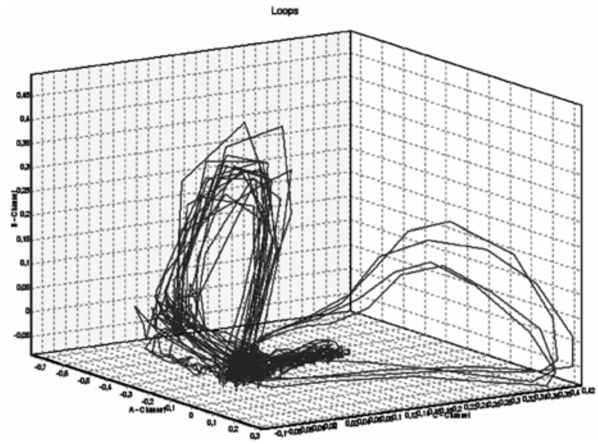

Fig. 5. ECG signal trajectory in orthogonal system of coordinates

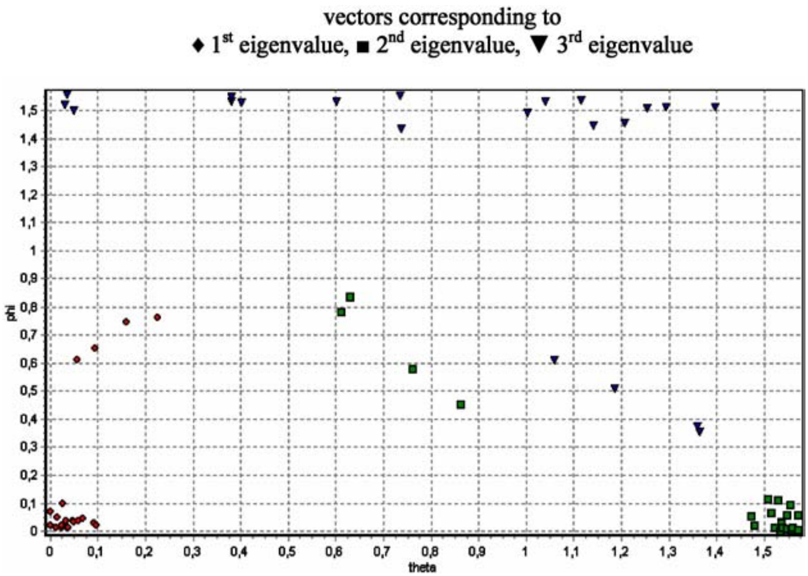

Fig. 6. Distribution of eigenvectors after transformation

The data from the Chair and Clinic of Cardiology, Medical University of Warsaw of 5 selected patients suffering from arrhythmia caused by premature ventricular excitations from various foci were examined. The signal files of 2000 heartbeats for each patient were analysed. The training set consisted of 200 cycles for each patient. The training set for SVM classifier consisted of 200 heartbeats for each patient, hence total number of examples was 1000 excitations. The linear SVM classifier trained for one patient is able to detect normal beats of another patient recording. The score of successful recognition of normal and pathological excitations is greater than 95%, in 4 considered cases, as listed in Table 1. The functionality of SVM classifier is illustrated in Figures 7 and 8.

Table 1. Results of support vector machine classification

Patient	P003	P011	P018	P101	P103	
Correct classification	**1302**	**1344**	**1581**	**740**	**1775**	Normal ECG beats
Wrong classification	**15**	**56**	**0**	**2**	**7**	
Number of patterns	**1317**	**1400**	**1581**	**742**	**1782**	
Score	**0.987**	**0.960**	**1.000**	**0.997**	**0.996**	
Correct classification	**384**	**325**	**417**	**329**	**202**	Pathological ECG beats
Wrong classification	**15**	**6**	**22**	**8**	**13**	
Number of patterns	**399**	**331**	**439**	**337**	**215**	
Score	**0.962**	**0.982**	**0.959**	**0.976**	**0.940**	
Total number of patterns	**1716**	**1731**	**2030**	**1079**	**1997**	

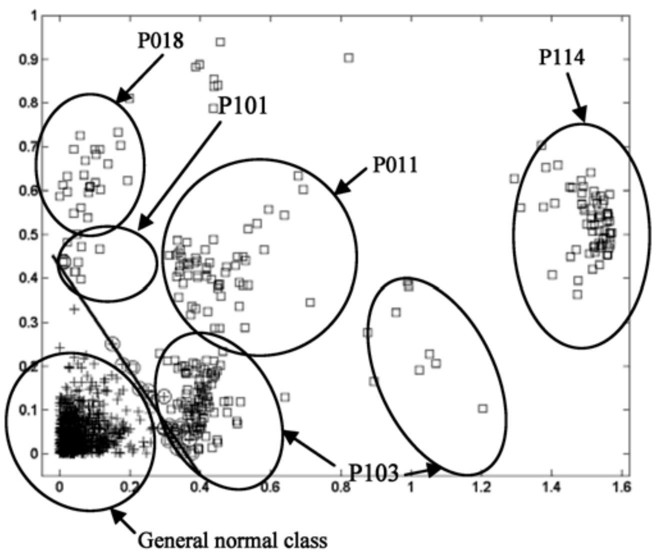

Fig. 7. Clusters of heartbeats representing the normal and the premature ventricular excitations for 5 patients (P011, P018, P101, P103, P114) represented by modified PCA descriptors and the linear SVM classifier

5 Conclusions

The modified PCA representation of single heartbeats obtained from 3-channel Holter monitoring enables efficient automatic classification of normal and pathological cases. Application of modified principal component analysis to data parameterisation allowed us to design simple and efficient classifier based on support vector machine. We emphasize that SVM trained on data set of one patient is able to classify the heart beats of other patients with approximately equal probability.

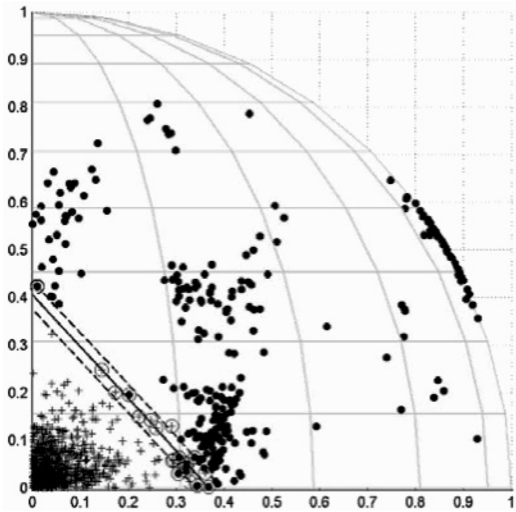

Fig. 8. Linear SVM classifier and margin of data from Fig. 5 presented in spherical coordinate system

Modified principal component analysis enables to perform one general classifier for all selected patients. The shape of clusters in plane-spherical system of coordinates can be used for unsupervised classification of large data sets.

The successful recognition is due to combination of PCA data suppression, as reported in [4] and to natural orientation of systems of coordinates for each patient along the largest eigenvector corresponding to individual normal heart action.

References

1. Endo T., Yamaki M., Ikeda H., Kubota I., Tomoike H.: Relation of principal components of ECG maps to loci of wall abnormality in old myocardial infarction, *Am. J. Physiol.* 266 (*Heart Circ. Physiol.* 35), pp. H1604-H1609, 1994
2. Cover T. M.: Geometrical and Statistical Properties of Systems of Linear Inequalities with Application in Pattern Recognition, *IEEE Transaction on Electronic Computers*, June 1965, pp.326-334
3. Dubois, R.: Application des nouvelles méthodes d'apprentissage à la détection précoce d'anomalies en électrocardiographie, Ph.D. thesis, Université Paris 6, 2003
4. Cao, L. J., Chua, K. S., Chong, W. K., Lee, H. P., Gu, Q. M.: A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine, *Neurocomputing* **55** (2003), pp. 321-336
5. Jankowski, S. Tijink, J., Vumbaca, G., Balsi, M., Karpiński, G.: Morphological analysis of ECG Holter recordings by support vector machines, Medical Data Analysis, Third International Symposium ISMDA 2002, Rome, Italy, October 2002, Lecture Notes in Computer Science LNCS 2526 (eds. A. Colosimo, A. Giuliani, P. Sirabella) Springer-Verlag, Berlin Heidelberg New York 2002, pp. 134-143
6. Jankowski, S., Oręziak, A., Skorupski, A., Kowalski, H., Szymański, Z., Piątkowska-Janko, E.: Computer-aided Morphological Analysis of Holter ECG Recordings Based on Support Vector Learning System, *Proc. IEEE International Conference Computers in Cardiology* 2003, Tessaloniki, pp. 597-600

7. Oxford Medilog Holter Management Systems - http://www.scanmed.co.uk/oxford.htm
8. Jankowski S., Dusza J. J., Wierzbowski M. and Oręziak A. "PCA representation as a useful tool for detection of premature ventricular beats in 3-channel Holter recording by neural network and support vector machine classifier", International Symposium on Biomedical Data Analysis ISBMDA 2004, Barcelona, Lecture Notes in Computer Sciences LNCS 3337, Springer Verlag, Berlin Heidelberg New York 2004, pp. 259-268.
9. Wagner, G. S.: Marriot's Practical Electrocardiography, Williams & Wilkins, 1994
10. Vapnik, N. V.: Statistical Learning Theory, John Wiley & Sons Inc., New York, 1998