FEATURE EXTRACTION IN BODY SURFACE POTENTIAL MAPPING

Václav Chudáček*, Michal Huptych*, Lenka Lhotská*, Otomar Kittnar**, Mikuláš Mlček**

* Czech Technical University in Prague/Gerstner Laboratory, Technická 2, Prague, Czech Republic
** Charles University/Physiological Institute First School of Medicine, Albertov 5, Prague, Czech Republic

chudacv@fel.cvut.cz

Abstract: Multichannel electrocardiography is an extension of the conventional electrocardiography. It enables to acquire refined data for more complex analysis. Body surface mapping is a graphical presentation of cardiac activity as measured from the body surface. In this work we have focused on feature extraction from the measured signals and finding relations between them. Three groups of healthy women have been measured; one group with heart signal, as we assume, predictably distorted by pregnancy; second group of women after delivery, and third group of healthy not pregnant women. We have generated several decision trees from the extracted values, using different combinations of features. We have identified several features that may be more discriminative than the other ones.

Introduction

Electrocardiography (ECG) deals with the electrical activity of the heart. An ECG signal can provide us with a great deal of information on the normal and pathological physiology of heart activity. An ECG as an electrical manifestation of a human activity is composed of heartbeats that repeat periodically. In each heart beat several waves and interwave sections can be recognised. The shape and length of these waves and interwave sections characterise cardiovascular diseases, arrhythmia, ischemia and further heart diseases. Basic waves in ECG are denoted by P, Q, R, S, T, U. From these, the denotation (and length) of the intervals and segments is derived. In medical practice ECG analysis is performed nearly exclusively as a temporal analysis. When interpreting an ECG, physicians first locate the P waves, QRS complexes, T complexes and U waves. Then they interpret the shapes (morphology) of these waves and complexes; in addition they calculate the heights and the interval of each wave, such as the RR interval, PP interval, PR interval, QT interval, and ST segment. From the technical point of view, the assumption for ECG analysis is the existence of perfect ECG signals (i.e. signals with sufficient dynamics and a minimum of artefacts).

The standard 12-lead ECG is still the only universally accepted practical method used for diagnosing heart diseases. However, it is not optimal and has its limitations as many research and clinical studies show [1], [2]. Multichannel ECG (MECG) is an extension of the conventional electrocardiography that provides refined non-invasive characterisation of cardiac activity. The MECG systems may use 30-300 electrodes. Increased spatial sampling on the body surface provides more in-depth information on the cardiac generated potentials, and thus in many cases exhibits better diagnostic value [3]. However, the MECG has its drawbacks as well, e.g. more complicated measurement and electrode positioning. MECG is used in body surface potential mapping (BSPM). The most frequently used body surface maps are isopotential, giving a distribution of the potential at a specific moment, and isointegral, providing a distribution of the sum of potentials over a specified time interval.

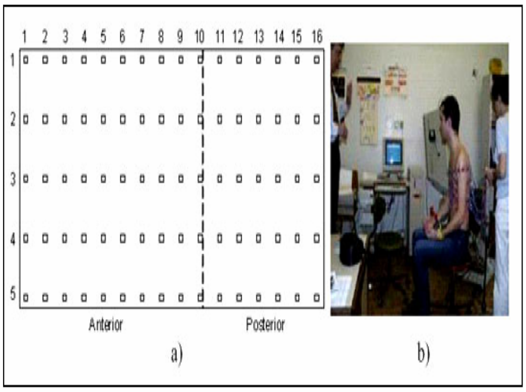

Figure 1: Cardiag 112.2 electrode matrix

The result of a MECG measurement is in fact a time series of consecutive vectors, encapsulating data collected by all the electrodes at the sampling time. These vectors can be visualised using interpolation, by different types of isovalued lines or regions, also called the body surface map. Electrodes placed at different positions result in different data. If we want to obtain normalised and comparable results some kind of data

ISSN: 1727-1983 © 2005 IFMBE

transformation has to be performed. The MECG measurements deliver large amount of electrocardiological data for further processing. However, not all the observed phenomena are fully understood. The aim of our research is discovery of possibly useful knowledge, so far not routinely analysed, from repolarisation phase of the heart cycle.

Measurement and Data

We have used the Cardiag 112.2 system (Czech MECG device) with 80 electrodes in 16 x 5 equidistant matrix as on Figure 1. The system allows recording of standard ECG, vectorcardiograph and MECG. The advantage of using more electrodes is that we acquire more data, although potentially redundant, and thus mapping is more exact. If lower number of electrodes (30) is used additional transformation of data has to be applied to correctly reproduce approximated equidistant maps.

In our study, electrocardiographic, vectorcardiographic and BSPM recordings have been obtained from three groups of women, pregnant women in 36th - 40th week of pregnancy, women in 3rd - 6th day after delivery, and reference set of young healthy not pregnant women. Total number of women has been 88, out of which 18 have not been pregnant, 43 have been pregnant, and 27 after delivery. In Table 1 selected parameters of measured data mean values are shown.

Table 1. Selected parameters of data mean values, SD – standard deviation, QT – length of QT interval, QTc – length of QT related to RR interval, Ta – amplitude of T-loop, Tw – angle (width) of T-loop

	QT[ms]	SD	QTc[ms]	SD	Ta	SD	Tw	SD
Not pregnant	342	13	382	35.5	597	141	8.1	4
Pregnant	322	12.8	419	39	413	151	11	8.5
After delivery	325	11.8	413	33	426	133	12.5	8.5

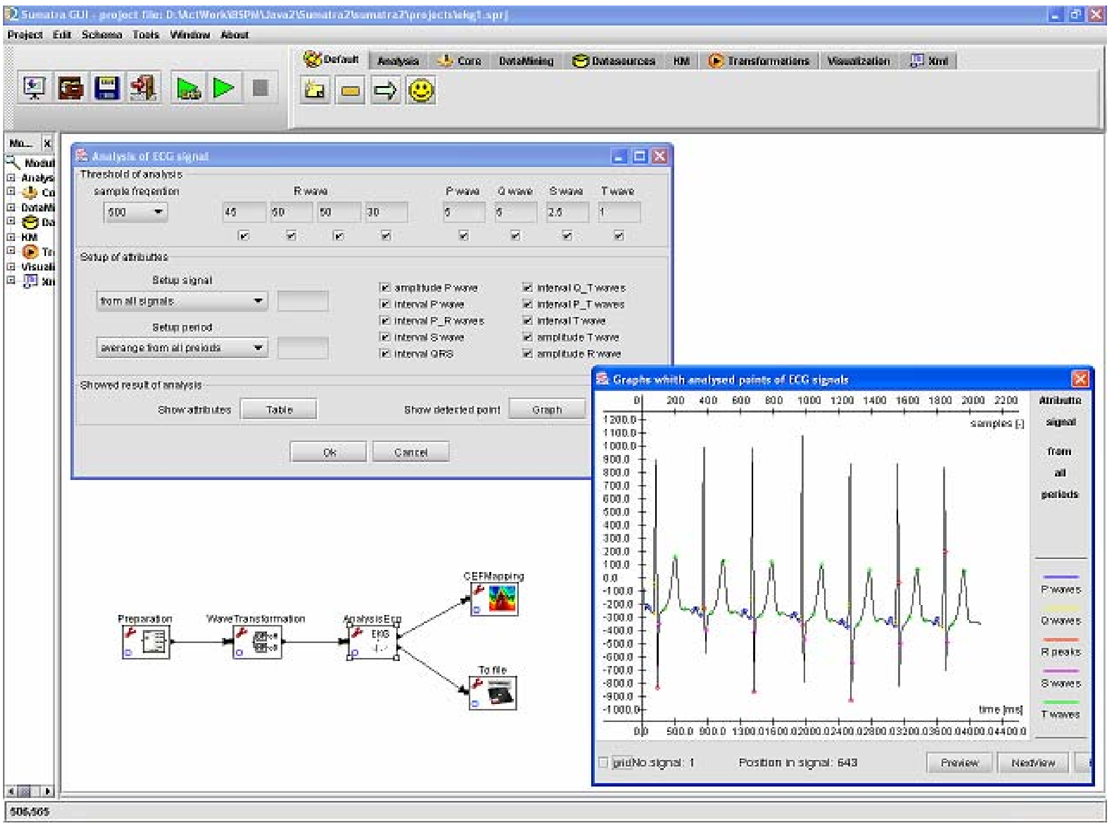

Figure 2: Pre-processing module with ECG analysis

ISSN: 1727-1983 © 2005 IFMBE

Feature Extraction

Feature extraction and feature selection are very important steps in the process of knowledge discovery and in classification systems. Proper selection of features may significantly influence the success rate of classification. Use of irrelevant and weakly relevant features can decrease the accuracy. The features can be selected either automatically or manually. Automatic selection can be viewed as a state space search where each state represents a single combination of features. The goal of the search is to find the state with the highest value of the evaluation function that characterises the success rate of classification with the corresponding features. It is obvious that such an evaluation function is only an estimation of the success rate of the classification, because the training set is limited. The transition operator is feature adding or deleting. The average accuracy of cross-validation is usually used as an evaluation function. Manual selection, on the other hand, is a more or less intuitive process based on experience.

One of the most important aspects of the ECG classification systems is reliable analysis of ECG records, which enables significant values to be identified on the measured signal. This analysis is a necessary condition for correct classification.

We have analysed a number of parameters that can be computed from the measured signals. Based on our previous experience with ECG signal preprocessing and classification [4], [5], and a number of experiments we have identified several parameters that may contain the significant information (see Table 1).From conventional ECG we have used the following parameters: length of QT segment, length of QRS complex, length of STend (length of segment from the end of S wave to the end of T wave), data referred to the heart rate, namely QT_rel, QRS_rel, STend_rel, using R-R interval.

From vectorcardiography we have used two parameters, namely amplitude of the T-loop Ta and angle (width) of the T-loop Tw. The amplitude of the T-loop is defined as the distance of the most distant point of the T-loop from the outset of the coordinates. The angle Tw is the angle of the vector connecting the outset with ¼ of the T-loop and the vector connecting the outset with ¾ of the T-loop.

From MECG we have used the following parameters:

- maximum of T-wave, in case of noisy data replaced with "?" as character representing missing value;
- position of maximum of T wave in time – relatively with respect to the length of STend segment;
- position of maximum and minimum of isopotential map in x,y coordinates;
- values of maximum and minimum of isopotential map;
- position of maximum and minimum of isointegral map in x,y coordinates;
- values of maximum and minimum of isointegral map, summed in the STend interval.

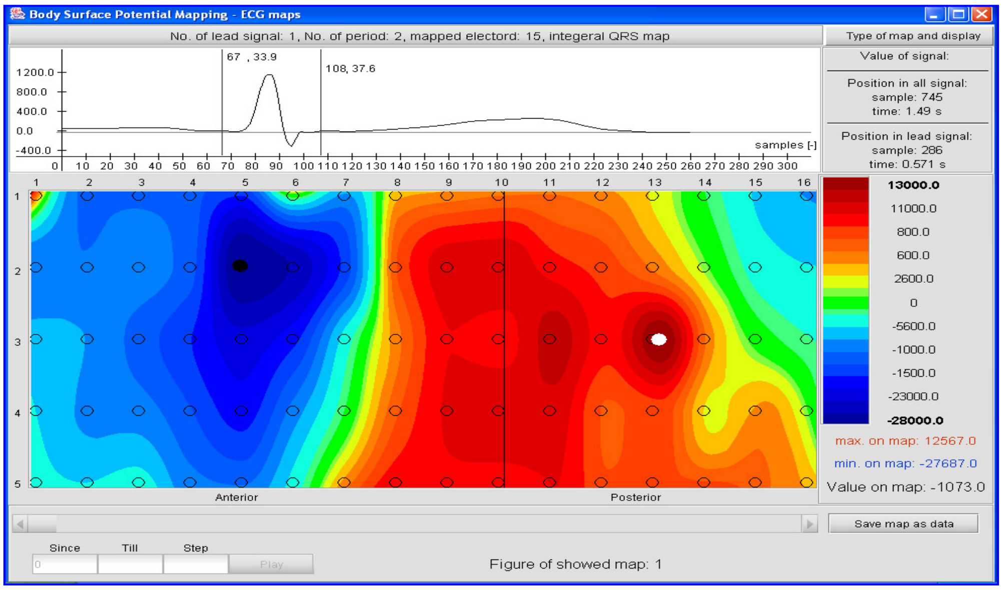

Figure 3: Example of isointegral map

ISSN: 1727-1983 © 2005 IFMBE

Additional information is state of the patients, namely healthy not pregnant women – reference group (class 1), pregnant women (class 2), and women after delivery (class 3).

Implementation

We have designed and developed a generally applicable software tool for data visualisation and feature extraction that is embedded in SumatraTT system [10]. The developed software tool for BSPM computation, visualisation and feature extraction is highly modular and enables to extract potentially interesting features from the data. The raw measured data can be shown, analysed, and presented graphically in time domain as standard ECG plots, different body surface maps, or animated maps (see Figure 2). In addition, different computations may be performed: finding R-R interval, QT interval, QTc interval, spatial maximum, Tmax-Toff parameter, etc. The extracted feature values can be evaluated using various machine learning algorithms. We have chosen two algorithms, namely See5 [5] and J48.PART from the WEKA open source [6] for generation of decision trees. Our aim has been to acquire results in an explicit and understandable form.

Analysis of Results

Although results suffer from dependency on small data set, and problems with medical data in general, such as non-heterogeneity of biological systems and noise induction during measurement period, some interesting results have been obtained. An example of a generated decision tree is shown in Table 2.

We have found following interesting (important) differentiating features using Weka software tool [6]:

- Description of vectorcardiographic T-loop Tw and Ta quite unambiguously differentiate the reference group from the pregnant and after-delivery groups.
- Position of spatial maximum of isopotential map differentiates the group of pregnant women from the reference group. Position of spatial maximum of pregnant women has been usually found caudally (low in the direction of y axis) than in the reference group
- Tmax-Toff parameter includes information about the angle of the heart axis, meaning heterogeneity of repolarization.
- The groups of pregnant and after-delivery women are merging partially. The group of pregnant women that is stronger has "pulled" nearly half of the after-delivery group. This can be seen not only in Table 1 but it was also observed in [4]. Our analysis is thus in correspondence with medical observations indicating that changes of electrical field of the heart caused by changed geometry of thorax are covered by other more significant changes caused by pregnancy and remaining for certain period after delivery; for example, changes of inner conditions of the body (hormonal changes, retaining fluids) cause changes of electric conductivity of tissues.

Table 2: Example of generated decision tree

Decision tree:
Ta <= 547.6988:
:...TmaxToff > 37:
: :...pozice_maxima_z_mapy03x <= 37: 1 (3)
: : pozice_maxima_z_mapy03x > 37: 2 (3)
: TmaxToff <= 37:
: :...pozice_maxima_z_mapy08x <= 59:
: :...T_max_poz82 <= 0.78626: 2 (32.6/11.5)
: : T_max_poz82 > 0.78626: 3 (15.4/2.9)
: pozice_maxima_z_mapy08x > 59:
: :...T_max_hod01 <= 378.8304: 1 (2)
: T_max_hod01 > 378.8304: 2 (8)
Ta > 547.6988:
:...Tw > 14.5955: 3 (2)
Tw <= 14.5955:
:...T_max_hod46 > 148.9341: 2 (5)
T_max_hod46 <= 148.9341:
:...T_max_poz75 <= 0.7622: 2 (2.1)
T_max_poz75 > 0.7622: 1 (14.9/1.9)

Size of decision tree : 10.
Accuracy: 79,5%
(a) (b) (c) <-classified as
---- ---- ----
18 (a): class 1
2 42 (b): class 2
16 10 (c): class 3

Rules:
Rule 1: (14/1, lift 4.3)
Ta > 547.6988
Tw <= 14.5955

However, we have not confirmed unambiguously several expected characteristics, namely extremes of isointegral maps, position of the extremes of isointegral maps. We have analysed the original data with respect to these results and have come to the following conclusion. Measurement of both extremes is burdened by:

- objective error of the measurement – drift of signal zero has higher influence on map minimum because most of pronounced T waves are in positive leads;
- error given by current skin conductivity of the measured patient, this error contributes more significantly to increase of maximum;
- error caused by incorrect contact electrode-skin.

Conclusion

At present body surface potential mapping is the most complete visualisation of the heart activity mapped on body surface. However its problem is the interpretation of the majority of measured data, with the exception of measurement of isopotential maps (measurement of differences of map values from healthy etalon). This study has tried to discover new relations in measured evaluated data that might contribute to deeper understanding of electrophysiological heart activity influenced by physiological changes of thorax geometry (heart position in thorax).

We have designed a new tool for visualization of body surface potential maps. We have developed several modules for further features extraction, and we have tried to acquire information hidden in body surface potential maps. Concluding from the above mentioned results we can state that heart changes due to pregnancy are covered by larger changes of conductivity of thorax caused by changes in the body. Further investigation that will provide more discriminative set of features is required for revealing relations between heart position and BSPM.

Acknowledgement

This work has been supported by the research program “Information Society” under Grant No. 1ET201210527 "Knowledge-based support of diagnostics and prediction in cardiology " and by grant No. FRV 3164/G3 “Body Surface Potential Mapping – pre-processing and visualization” from the Ministry of Education, Youth and Sports.

References

[1] KUBOTA, I., IKEDA, K., OHYAMA T., YAMAKI, M., KAWASHIMA, S., IGARASHI, A., TSUIKI, K., YASUI, S. (1985): ‘Body surface distributions of ST segment changes after exercise in effort angina pectoris without myocardial infarction’, American Heart Journal, **110**, pp. 949-955

[2] YANOWITZ, F., VINCENT, M., LUX, R.L., MERCHANT, M., GREEN, L.S., ABILDSKOV, J.A. (1982): ‘Application of body surface mapping to exercise testing: S-T80 isoarea maps in patients with coronary artery disease’, American Journal of Cardiology, **50**: pp. 1109-1113

[3] KORNREICH, F. (1997): ‘Appropriate electrode placement in evaluating varied cardiac pathology’ in Liebman, J., ‘Electrocardiology ‘96 From the cell to the body surface’, pp. 83-92.

[4] LHOTSKÁ, L., FEJTOVÁ, M., MACEK, J., NOVÁK, D., (2004): ‘Feature Extraction From Biological Signals: A Case Study’ in ‘IEEE 4th International Conference on Intelligent Systems Design and Application’, pp. 139-144

[5] MACEK, J., LHOTSKÁ, L., PERI, D. (2004): ‘Evaluation of ECG: Comparison of Decision Tree and Fuzzy Rules Induction’ in ‘Cybernetics and Systems 2004’ pp. 713-718

[6] See5 demo version 1.19. [online] <http://www.rulequest.com/see5-info.html>

[7] WEKA (Waikato Enviroment for Knowledge Analysis), University of Waikatom Dep. Of Computer science, New Zealand. [online] <http://www.cs.waikato.ac.nz/~ml>

[8] LECHMANOVÁ, Z., KITTNAR, O., MLČEK, M., SLAVÍČEK, J., DOHNALOVÁ, A., HAVRÁNEK, J., KOLAŘÍK, J., PAŘÍZEK, A. (2002): ‘QT Dispersion and T-Loop Morphology in Late Pregnancy and After Delivery’, Physiol. Res. **51** pp. 121-129

[9] WITTEN, I., FRANK, E. (2000): ‘Data mining: practical machine learning tools and techniques with Java implementations’

[10] AUBRECHT, P., ŽELEZNÝ, F., MIKŠOVSKÝ, P., ŠTEPÁNKOVÁ, O. (2002): ‘SumatraTT: Towards a Universal Data Preprocessor’ in: ‘Cybernetics and Systems 2002’ Vienna Austria, pp. 818-823

ISSN: 1727-1983 © 2005 IFMBE