ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2011. No. 4(110)
ELEKTRONIKA IR ELEKTROTECHNIKA

T 121 SIGNAL TECHNOLOGY
SIGNALŲ TECHNOLOGIJA

Efficient ECG Modeling using Polynomial Functions

S. Jokic, V. Delic
Faculty of Technical Sciences, University of Novi Sad,
Dositej Obradovic Square: 6, 21000 Novi Sad, Serbia, phone +381 21 475 29 99, e-mails: stevan.jokic@gmail.com, vdelic@uns.ac.rs

Z. Peric
Faculty of Electronic Engineering, University of Niš,
Beogradska: 14, 18000 Niš, Serbia; e-mail: zoran.peric@elfak.ni.ac.rs

S. Krco
Faculty of Organizational Sciences, University of Belgrade,
Jove Ilica: 154, 11000 Belgrade, phone +381 11 3950893, e-mail: srdjan.krco@fon.rs

D. Sakac
Medical Faculty, Novi Sad,
Hajduk Veljkova: 3, 21000 Novi Sad, Serbia; e-mail: dejsakac@open.telekom.rs

Introduction

ECG signal represents electrical changes on the skin that are caused by the heart muscles, and is usually measured by the electrodes placed on body surface. ECG analysis is one of the most common procedures in the heart diagnostic domain. After digitalization of ECG signal, next step in computer abstraction of ECG signal is appropriate modeling. Due to high importance of accurate modeling several ECG modeling approaches applicable for different purposes like heartbeat synthesis, analysis, compression, and filtering were introduced. In [1], a realistic ECG model designed for ECG synthesis based on Gausses curves modeling of appropriate waves is presented. Fitting this model to a real ECG signal is possible using iterative nonlinear optimization algorithms. An algorithm for ECG compression by alignment of heart beats in a matrix and fitting polynomial curves to columns is used by Riad B. [2]. Instantaneous module and the instantaneous phase of Hilbert Transform are used for ECG modeling in [3]. Mealy and Moore automata model is used in [4] for ECG complexity analysis.

The main motivation for development of our ECG model was to build an efficient feature extraction algorithm for automated ECG analysis that will be possible to implement on devices with limited computing capabilities like mobile phones [6].

Polynomial Functions estimation

Let y represents ECG segment vector length N samples, $y(n)$: $y(1)$, $y(2)$, ... , $y(N)$. Polynomial Function (PF) $\hat{y}$ of degree m is an approximation of y, and can be written as follows

$$\hat{y}(n) = a_0 + a_1 n + ... + a_m n^m, n = 1, ..., N. \quad (1)$$

The PF coefficients are calculated by minimizing the least square error

$$e^2 = \left\| y(n) - \hat{y}(n) \right\|_2^2. \ \frac{\partial e^2}{\partial a_i} = 0. \quad (2)$$

Error minimization is performed by solving quadratic system of equations introduced in (2). The solution can be computed explicitly

$$\hat{a} = \left[C^t C \right]^{-1} C^t y, \ \ C = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & N & \cdots & N^{m-1} \end{bmatrix}. \quad (3)$$

A special case of the described fitting for straight line estimation through point with coordinates (0, 0) with equation $\hat{y} = k * n$ has very simple solutions for its only parameter, the slope

$$k = \sum_i y_i \Big/ \sum_i n_i \, . \quad (4)$$

Analysis of the slope estimated over the ECG signal segments will be used for the ECG fiducial point detection.

Method

The ECG modeling proposed in this paper introduces segmentation of ECG signal and fitting polynomial curves with appropriate degrees to the specific segments. Heartbeats are modeled using PF up to the fourth order. The first step in modeling is the R peak detection as it is the largest deflection away from the baseline. All segment intervals and peak positions are then expressed relative to the R peaks.

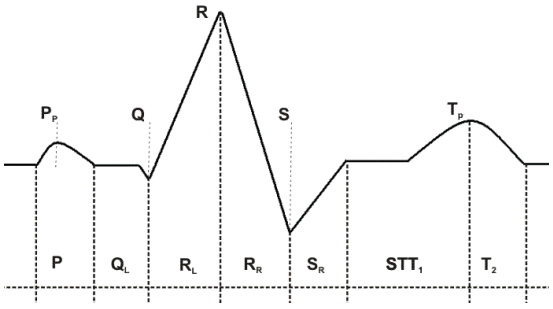

Fig. 1. Heartbeat model segments

Each heartbeat is divided into seven segments, as shown in Fig. 1. The R peak is modeled using two third order PFs. The PFs are estimated from the Q peak to the R peak, fitted over the R_L segment and from the R peak to the S peak, fitted over the R_R segment. The R peaks boundaries, Q and S peaks, are detected by analysis of changes in the slope of a straight line fitted on R_L and R_R segment respectively. The straight line is fitted over a moving three samples long window using (4). Average slopes over three successive moving window positions are calculated then. Also, the extreme values of the average slopes are updated and stored. A point is marked as a S peak candidate if the slope of the fitted straight line change sign or change values for more than 85% of the average slope or 90% of the extreme slope. To avoid potential classification errors due to noise and jaggy peaks in the signal, the S peak is detected if changes in the signal are above the mentioned thresholds for six successive samples after a S peak candidate. During this procedure, average and extreme values of the slope are not updated. The selected number of samples over which the thresholds were calculated was determined empirically as the value that provides the best accuracy in the presence of noise and artifacts in a ECG signal. The S peak could have several morphologies and could carry the most significant amount of information useful for heartbeat classification, like for Premature Ventricular Contractions (PVC). Analog procedure is applied over R_L segment for finding Q peaks.

The ECG signal preceding a Q peak, the so called PR interval, is analyzed in order to determine the isoelectric level. This value is useful in detection of ST depressions, elevation analysis and for S peak modeling. To determine this level, first, starting from a Q peak and using (4), a slope is calculated based on the three preceding samples. Then, the closest three points to a Q with the smallest slope are marked as isoelectric points. The isoelectric level is an average value of isoelectric points.

The end of a S peak is searched in the ECG samples after the S peak as the point of a significant change in the slope or points in which ECG crosses the isoelectric level. The slope change is analyzed using similar approach as for detection of Q and S peaks.

A high diversity of T wave morphology makes its modeling difficult. In this paper, the ST segments and the T waves are modeled by two third order PFs. The first one models the ST segment, from the end of S peak over the ST segment and a part of the T wave up to the T wave extreme, T_P. The second polynomial function models the rest of the T wave. The first step in a T wave modeling is T wave detection. A T wave is searched in the area between the end of a S peak and the Q peak of the next QRS complex. A T wave extreme value is located using analysis of a fitted squared function, estimated on a 120 ms long moving window. The 120 ms window length is chosen because it contains the most significant parts of a T wave. If a processed T wave is wider than 120 ms that information will be in successfully fitted in the squared PF coefficients. The square fitting function is chosen because of simplicity in the shape analysis of a fitted function. Fitting PFs of higher order have tendency to oscillate in segments where the original signal has tendency to be constant. For the following square function ax^2+bx+c, an extreme position and the curve width are given by following equation

$$x_{extr} = -b/2a\,,\ w_h = \frac{\sqrt{b^2-4ah}}{a}\,. \qquad (5)$$

In this formula *h* is the distance from vertex of square function parabola on axis of symmetry. The curve extreme point represents relative positions of the square function vertex in a moving window. The window is moved over a ST segment to the T wave while estimated square function on moving window does not fit according to the following conditions

$$x_{extr} \in (0.45t_w, 0.55t_w) \text{ and } w_h \in (0.5t_w, 2.5t_w), \quad (6)$$

where t_w is the width of a moving window, initial value 120 ms, *h* is empirically set to the value $0.6D$, where D is dispersion calculated over the ST segment. Location of T wave extreme *Tp* is equal to position of the first moving window sample increased for x_{extr}. If a T wave could not be detected in between successive R peaks, the *h* value is reduced by 10% and t_w is increased by 30%. Corrections and initialed values for parameters *h* and t_w are chosen based on intensive testing. If a T wave extreme could not be located, the middle point is returned for T wave extreme.

Due to small amplitude, a P wave detection is one of the hardest tasks in ECG processing. P wave can be absent, inverted or sharpen due to different arrhythmias. Often, a P wave absence occurs in supra SVT arrhythmias, Atrial Fibrillation (AF). A P wave is modeled using fourth order polynomial function fitted on ECG signal window in which a P wave is detected. Similarly procedure to the T wave detection procedure is adopted for P wave locating.

Moving window of width t_w for P wave locating is 40 ms. The analysis is performed on an ECG segment between the preceding heartbeat *Tp* and the current heartbeat's isoelectric points from PR interval. ECG segment that connects a P wave and a Q peak is modeled using a third order PF.

Table 1. PRD and Execution time (EXC) of modeling ECG signals from MIT BIH ECG base for Realistic and Polynomial model

MIT BIH Signal		106	107	116	119	200	205	208	209	210	213
Realistic	PRD [%]	16.8	7.7	15.2	14.5	17.1	15.7	17.9	18.8	13.1	18.8
	EXC [ms]	13500	15093	17812	11812	20844	20547	22375	20953	21375	26703
Polynomial	PRD [%]	6.6	3.6	3.5	5.3	3.5	6.1	6.3	5.9	10.8	4.9
	EXC [ms]	15	15	16	31	31	32	32	31	31	31

Fig. 2. Modeling performance of MIT BIH signal *233* (top panel), Polynomial model with error below, Realistic model with error (bottom panel).

Results

Model quality is usually expressed in terms of the Percentage of the Root mean square Difference (PRD) given by the following equation

$$PRD = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{y_i^2}}. \qquad (7)$$

PRD is an overall measure of model quality across the data points. Objective measure of quality is in the amount of diagnostic information retained in the model features after ECG raw data modeling. In this paper PRD is measured on several signals from the MIT BIH ECG base and compared with the realistic ECG model. The realistic ECG model is based on fitting six Gausses to real ECG [1], [5]. Model error defined in (2) is minimized using Levenberg–Marquardt algorithm with 13 iterations. The results of modeling with PRD and execution time are listed in Table 1. Both models are implemented in JAVA and executed on standard PC. Realistic ECG model has a greater execution time due to iterative approach for solving (2). Furthermore, the realistic model requires ECG data to be sampled at frequency above 500 Hz in order for the interpolation to be performed on the original signal and without segmentation fitting is performed over the whole heartbeat. Benefits of proposed polynomial modeling are noticed in modeling ECG with high morphological diversity heartbeats because proposed polynomial modeling does not require template initialization. Example of modeling of ECG signal with PVC is shown in Fig. 2.

Clinical information persistence of the model is evaluated using ST-T analysis applied on model features. Features chosen for analysis are isoelectric level, STT_1 segment duration and polynomial coefficients and T_2 polynomial coefficients. ST-T analysis evaluation is performed on annotated signals *e0108* and *e0113* from MIT BIH European STT data base [9]. ST segment classes are *N* (normal), *ST+* (elevated) and *ST-* (depressed). T wave is classified as one of the following classes: *N* (normal), *T+* (augmented) and *T-* (inverted). The most intuitive T wave feature is amplitude [7]. In proposed modeling approach T weave amplitude is available as PF value at first and last point over T_2 and STT_1 segments respectively. Classification is done using Feed Forward Artificial Neural Networks (FFANN) with ten inputs, ten nodes in a hidden layer and six outputs. ANN shows good results in a both classification and prediction [8]. Performances of ST-T analysis are shown as confusion matrix in Tables 2. and 3.

Table 2. Confusion matrix for ST segment analysis

Actual / Predicted	N	ST+	ST-
N	1050	112	0
ST+	136	3792	12
ST-	15	5	118

Table 3. Confusion matrix for T wave analysis

Actual / Predicted	N	T+	T-
N	3992	73	53
T+	40	822	9
T-	17	23	121

Conclusions

An efficient ECG modeling approach is presented in this paper. The algorithm achieves high accuracy in the ECG data modeling and does not require significant computing resources. Furthermore, estimated model features persistently retain important medical information such as ST-T level and shape. Future work will be focused on algorithms development dedicated to ECG heartbeat analysis using feature set provided by the model with potential compression and filtering applicability.

References

1. **McSharry P. E., Clifford G. D., Tarassenko L., Smith L. A.** A Dynamical Model for Generating Synthetic Electrocardiogram Signals // IEEE Trans. On Biomed. Eng. – Publishing House, 2003. – Vol. 50. – No. 3. – P. 289–294.
2. **Riad B., Amine N. A., Jacques L.** ECG Compression Using an Ensemble Polynomial Modeling: Comparison With the DCT Based Technique // Cardiovascular Engineering. – Publishing House, 2004. – Vol. 4. – No. 3. – P. 237–244.
3. **Jean C. N., Amine N. A.** Hilbert Transform Based ECG Modeling // Biomedical Engineering. – Springer Link, 2005. – Vol. 39. – No. 3. – P.36–40.
4. **Martusevičienė A., Navickas Z., Vainoras A.** ECG Data Analysis Using the Convolution of Mealy and Moore Automata // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 4(100). – P. 103–106.
5. **Clifford G. D., Shoeb A., McSharry P. E., Janz B. A.** Model–based filtering, compression and classification of the ECG, International // Journal of Bioelectromagnetism, 2005. – Vol. 7. – No. 1. – P. 158–161.
6. **Krčo S., Kostić S., Sakač D., Lukić Z.** mSens Mobile Health Monitoring System // Proc. of EUROCON 2005. – P. 80–83.
7. **Smidtaite R., Navickas Z., Bikulciene L., Vainoras A., Poškaitis V.** Evaluation of Coherence of T wave in Different Leads // Electronics and Electrical Engineering. – Kaunas: Technologija, 2009. – No. 5(93). – P. 113–116.
8. **Rutka G.** Prediction Accuracy of Neural Network Models // Electronics and Electrical Engineering. – Kaunas: Technologija, 2008. – No. 3(83). – P. 29–32.
9. MIT–BIH ECG Database. Online: http://www.physionet.org–/physiobank/database/mitdb/.

Received 2011 02 14

S. Jokic, V. Delic, Z. Peric, S. Krco, D. Sakac. Efficient ECG Modeling using Polynomial Functions // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 4(110). – P. 121–124.

An efficient ECG modeling algorithm is presented in this paper. The model is based on fitting polynomial functions to real ECG. This algorithm describes a segmentation of a heartbeat and fitting an appropriate polynomial function to the segments. The model performances are evaluated in terms of PRD, preserve of ST-T segment clinical information and the execution time. When comparing this model with the existing one, the PRD improvements can be seen, especially in those signals with high morphological diversity heartbeats. Moreover, the computing time is significantly reduced. Using appropriate model features, ST-T analysis achieves an average accuracy of more than 94%. The obtained data shows that this model is applicable to other ECG processing like: heartbeat analysis, compression and filtering. Ill. 2, bibl. 9, tabl. 3 (in English; abstracts in English and Lithuanian).

S. Jokic, V. Delic, Z. Peric, S. Krco, D. Sakac. EKG efektyvumo modeliavimas taikant polinomus // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2011. – Nr. 4(110). – P. 121–124.

Pateiktas EKG efektyvaus modeliavimo algoritmas, paremtas polinomo funkcijų atitiktimi realiai EKG. Analizuojamas širdies plakimas segmentuojant ir aprašant polinomais. Atliktas siūlomo modelio našumo patikrinimas klinikinės informacijos turinčiame ST-T segmente. Nustatyta, kad skaičiavimas paspartėja. Remiantis gautais rezultatais galima teigti, kad algoritmas gali būti naudojamas EKG analizei. Il. 2, bibl. 9, lent. 3 (anglų kalba; santraukos anglų ir lietuvių k.).