# Optimizing the Short- and Long Term Regression for QRS Detection in Presence of Missing Data

Piotr Augustyniak

AGH-University of Science and Technology, Krakow, Poland

## Abstract

*Recently presented new QRS detection algorithm uses a detection function based on the value of angle between two regression segments adjacent in a given point. The optimization of these segments, representing long-term and short-term acceleration of heart activity, is presented in this paper.*

*The detection algorithm is based on four simple steps: (a) iterative linear regression based on samples selected in two windows of different length $\Delta t_L$ and $\Delta t_M$ (b) calculation of series of angle values between regression segments adjacent to each point, (3) expressing the synchronicity in both series as a detection function and (4) adaptive thresholding of the detection function.*

*The method was tested with original records from MITDB for different combinations of thresholding parameters and for $\Delta t_L$ and $\Delta t_M$ varying in ranges of 41.7-75 ms and 8.3-41.7 ms respectively. The quality of the detection was measured with false detection rates and represented by the area under the ROC curve. For the normal ECG, the values yielding the highest area of ROC for sensitivity and positive predictive value of QRS detection are respectively: $\Delta t_L$ = 58.3 ms and $\Delta t_M$ = 19.5 ms.*

## 1. Introduction

Although automatic detection of the heart beats has been investigated since the beginning of digital electrocardiography, current challenges consist in making it independent on signal quality. Unsupervised acquisitions are often made in home care conditions and interpretive applications require robust algorithms performing well independently on missing data, local distortions or uneven lead connection.

The analysis of the electrical activity of the heart and in particular the heart beats detection is based on time series of surface-recorded voltage. The ECG is then processed as a regular digital signal, where, in the absence of other requisites all samples are uniformly spaced and their values are equally accurate and reliable. This tacit assumption also justifies the common yet unwitting usage of filtering techniques in heartbeat detectors [1-3]. Some rare proposals consider representing the heart activity as temporal patterns of voltage instead of as time series [4-6]. Although the patterns are also uniform series of discretized voltage, the pattern matching process allows for certain dissimilarity of values (e.g. caused by a single outlying value) or for asynchronicity of samples. This approach integrates the local activity represented in a cloud of samples thus is favorable in case of distorted measurements (e.g. home care or stress-test ECG recordings) or intentional non-uniform sampling.

To this end we proposed very recently a new detection algorithm that uses a detection function calculated on the value of angle between two regression segments adjacent in a randomly selected point [7]. The algorithm is based on four simple steps:

1. iterative linear regression based on samples selected in two windows of different length $\Delta t_L$ and $\Delta t_M$,
2. calculation of series of angle values between regression segments adjacent to each point,
3. expressing the synchronicity in both series as a detection function and
4. adaptive thresholding of the detection function.

The algorithm uses two arbitrary selected time intervals around each given point in which the detection function is calculated. Optimization of these segments, representing long-term and short-term acceleration of heart activity, is presented in this paper.

## 2. Materials and methods

### 2.1. Background of detecting method

The proposed heart beat detector was based on two following assumptions:

1. the QRS complex is a coordinated electrical activity of prevalence of muscular fibers that manifests itself by a remarkable and consistent wave when projected to a segment in space corresponding to particular lead,
2. the QRS complex differs from other waves by sudden acceleration of electrical field resulting from steep surface enlargement of the depolarization front.

Consequently, the simultaneous occurrence of long-term and short-term activities are representative for a QRS, and these activities are to be detected from electrical measurements. A simple pattern of the local electrical activity proposed in [7] but also studied in [8] and [9] is an isosceles triangle with the top fit to the QRS peak and base aligned to the ECG's baseline. The angle measured at the triangle top changes from ca. 180 degrees for signal sections of low acceleration and deceleration (e.g. ECG baseline) to very small values for steep signal changes (as in the vicinity of the QRS). It gives a quantitative estimate of changes of signal trend that can be used for detection of QRS if the length of the triangle's sides is properly selected.

The selection of these sections length also has a practical consequence. Calculation of a single side of approximating triangle requires considering of at least two signal samples. Including more samples makes the method more robust to accidental outlying values.

As first approach we assume that detection of coordinated activity needs a triangle base length to be comparable with the total QRS duration, while detection of the sudden change corresponding to R peak needs a triangle base length to be comparable with shortest QR or RS section. Therefore initial values of these two sections were set intuitively as 75 ms and 41 ms respectively and their optimal values were studied throughout this paper. The optimal values are expected to meet the following detection criteria:

– yield a reliable detection peak unique per heart beat disregarding the beat morphology,
– separate two adjacent beats in fastest possible rhythm and tachycardia
– provide a fast response (i.e. values of detection function are not delayed with regard to the actual ECG signal),
– provide unaltered detection function in case of variable sampling of the source ECG.

## 2.2. Detecting scheme

The details of proposed detection algorithm was presented in [7] and will be briefly recalled here. The algorithm performs in two stages: calculation of the detection function and thresholding it for indicate the most probable location of R peaks. While the first stage is expected to provide distinct values for QRS and non-QRS events, the second has to indicate the time points of QRS occurrence in a reliable and repetitive way. The first stage performs in the following steps (fig. 1):

Calculation of triangle ascending and descending slopes at a given time point based on ECG signal samples in its surrounding. Two series of slopes are calculated for each point: long term marked with L and short term marked with M. The slope coefficients, respectively $m_{L-}(t)$, $m_{L+}(t)$, $m_{M-}(t)$ and $m_{M+}(t)$ are calculated by a linear regression fitting the $y = m \cdot t + b$ line to the subset of signal values preceding and to the subset of signal values following the given time point $t$.

$$m_t = \frac{\sum_{j=1}^{J} K_j \cdot (t_j - t) \cdot v_j}{\sum_{j=1}^{J} K_j \cdot (t_j - t)} \quad (1)$$

where $J$ is the total number of samples in the window and $j$ is the current sample number relative to the window onset. The subset of samples taken into account for consecutive values of $t$ may overlap or may be disjoint, thus the train of regression values can be calculated as frequently as necessary, making the detection function independent on the ECG signal.

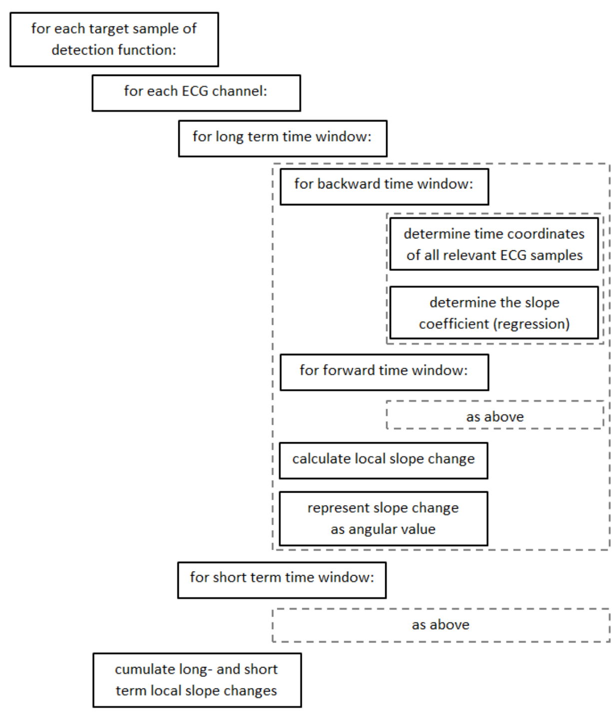


Fig. 1. Calculation of detection function

Calculation of the angle between the ascending and descending slope as the parameter representing the impulse-like activity approximated by the triangle. Two angle values $T_L(t)$ and $T_M(t)$ are calculated at each time point for two respective surroundings of the time point $t$.

$$T(t) = \tan^{-1} \frac{m_+(t) - m_-(t)}{\Delta t} \qquad (2)$$

To emphasize the simultaneity of the long- and short term activity we multiply the angle values at corresponding time points *t*. In case of a multichannel record, activities from several channels are also combined at this step.

$$p(t) = \sqrt{\sum_{c=1}^{N} (T_{Lc}(t) \cdot T_{Mc}(t))^2} \qquad (3)$$

where *N* is the total number of channels and *c* is the current channel number.

At the second stage, the detector determines the occurrence of heartbeats by detection of upward crossing of a specific threshold by the detection function determined at first stage. The value of threshold is initialized based on first two second of detection function and then adapted to its variability in specified adaptation sections. New threshold is calculated as:

$$H(t) = \propto \cdot H(t - N) + (1 - \propto) \cdot A(N) \qquad (4)$$

where adaptation inertia α = 0.9, *N* is the number of detection function samples in the section *d* and $A(N) = \frac{1}{N}(\sum(n_o = p(t): n_o > H) - \sum(n_u = p(t): n_u \leq H))$. The duration *d* of the adaptation section and thus the regulation responsiveness is adjusted to reach a compromise between required reaction time and suppression of overshoots.

## 2.3. Optimizing detection parameters

The detector scheme has four adjustment parameters: the lengths of long- and short time slopes ($\Delta t_L$ and $\Delta t_M$) at the first stage, and the adaptation section duration and inertia factor (*d* and α) at the second.

Both parameters of the first stage were set independently to favorite the QRS of any morphology and discriminate all other events (e.g. noise, spikes) possibly present in the ECG signal.

Thresholding of a detection function of any origin produces two kind of errors: false positive (*fp*), when non-existing events are detected and false negative (*fn*), when existing events are not detected. Based on the total number of events and assuming the sensitivity *Se* and positive predictive value *Pp* calculated accordingly to (5):

$$Se = \frac{tp}{tp + fn} \cdot 100\%$$
$$Pp = \frac{tp}{tp + fp} \cdot 100\% \qquad (5)$$

equally influence detection quality, the latter is commonly expressed as area (AUC) under the receiver operating curve (ROC). Both thresholding parameters: duration of the adaptation section *d* and adaptation inertia α have temporal coincidence and may be considered jointly. Fast responsiveness yields higher *fp* detection ratio and thus decreases *Pp*, whereas low responsiveness yields higher *fn* detection ratio and thus decreases *Se.*

Consequently, we performed a two-step optimization process:

- first looking for best threshold adaptation parameters *d* and *α* for each given slope duration (fig. 2), and
- then iteratively looking for the best value of AUC in two-dimensional space with variables $\Delta t_L$ and $\Delta t_M$ ($\Delta t_L > \Delta t_M$) corresponding respectively to long- and short term slope durations.

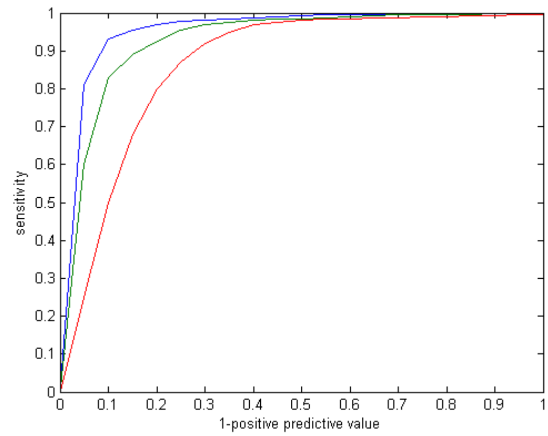


Fig. 2 Example of pursuit for best threshold adaptation parameters based on area under a receiver operating characteristics.

Considering that the MIT Arrhythmia Database (MITDB) [10] consists of samples of different QRS morphology in a true-to-live ratio, the method was tested with original records from MITDB. Although a non-integer sample number can be considered in both time windows, we only used a consecutive increment of the window length by one sample (i.e. 2.777 ms).

## 3. Results

Principal results of the reported research were optimal values of length of short- and long-time slopes. The maximum values of AUC achieved by combination of thirteen short time windows (ranging from 3 to 15 ms) and thirteen long time window (ranging from 15 to 27 ms) are presented in fig 3.

The combination of short- and long-time regression window yielding a global AUC maximum is: 7 and 21 samples (i.e. 19.5 and 58.3 ms). The resulting AUC is 0.9978 what suggests a simultaneously rare occurrence of false detection of both types (0.22 %). The achieved results in sensitivity (*Se* = 99.91%) and positive predictive value (*Pp* = 99.87%) make the proposed algorithm comparable with the most recent achievements in the domain (Tab. 1).

Detailed results on robustness to outliers, missing data and processing signals of sampling rates in the range of 200-500 Hz are presented in our previous work [7].

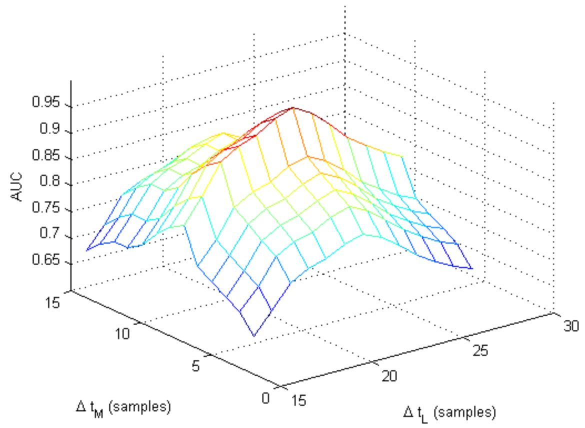


Fig. 3 Mesh of maximum AUC value for combination of different short- and long-time slopes

Table 1. Performance of the proposed algorithm compared to selected other work.

| algorithm | results | | |
|---|---|---|---|
| | ***Se* [%]** | ***Pp* [%]** | ***Fd* [%]** |
| Pan and Tompkins (1985) [1] | 99.75 | 99.54 | 0.71 |
| Martínez et al (2004) [11] | 99.80 | 99.86 | 0.34 |
| Martínez et al (2010) [8] | 99.71 | 99.97 | 0.32 |
| Song et al. (2015) [9] | 99.91 | 99.91 | 0.18 |
| proposed algorithm | 99.91 | 99.87 | 0.22 |

## 4. Discussion

Optimization of a novel QRS detector scheme was presented in this paper. This procedure was necessary to properly capture long-term and short-term variability of the signal as the only characteristic features of the QRS complex. The use of these features is motivated by electrical and mechanical processes typical for a heartbeat and makes the detection less prone to the signal quality. An interesting feature of the detection scheme is independence on signal sampling and tolerance for outliers and missing values. Therefore, the detection is robust against bad signal quality and allows various sampling frequencies at signal input.

The detection scheme is based on piecewise linear regression. It doesn't contain any form of digital filtering and the remapping of the signal to the detection function allows for asynchronous (and even non-uniform) sampling.

The preliminary optimization of threshold adaptation parameters $d$ and $\alpha$, is essential, as it yields a maximum value of AUC in ROC. The principal optimization, however, consists in finding slope duration parameters for appropriate representation of QRS-related phenomena.

Some limitation of the reported work comes from the use of sampling interval-dependent iterative adjustment of slope length. Therefore, further development of the proposed detection scheme needs verification of the results on databases with different sampling frequencies. An alternative approach assumes variable step iteration and respective resampling of the reference database.

## Acknowledgements

This scientific work is supported by the AGH University of Science and Technology in year 2015 as a research project No. 11.11.120.612.

## References

[1] Pan J, Tompkins WJ, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng., 1985;32:230-236
[2] Pahlm O, Sörnmo L. Software QRS detection in ambulatory monitoring – A review," Med. Biol. Eng. Comp., 1984;22:289–297
[3] Köhler B-U, Hennig C, Orglmeister R. The principles of software QRS detection," IEEE Eng. Med. Biol. Mag., 2002; 21:42–57
[4] Sörnmo L. A model-based approach to QRS delineation, Comput. Biomed. Res., 1987;20:526–542
[5] Gritzali F, Toward a generalized scheme for QRS detection in ECG waveforms, Signal Processing, 1988;15:183–192
[6] Andreão RV, Dorizzi B, Boudy J. ECG signal analysis through hidden Markov models, IEEE Trans. Biomed. Eng. 2006;53:1541-1549
[7] Augustyniak P, A robust heartbeat detector not depending on ECG sampling rate. in Proc. 37-th Annual Conference of IEEE EMBS 2015
[8] Martínez A, Alcaraz R, Rieta JJ. Application of the phasor transform for automatic delineation of single-lead ECG fiducial points, Physiol. Meas. 2010; 31:1467–1485.
[9] Song M-H, Cho S-P, Kim W, Lee K-J. New real-time heartbeat detection method using the angle of a single-lead electrocardiogram. Computers in Biology and Medicine 2015;59:73 – 79.
[10] Moody GB, Mark RG, The MIT-BIH arrhythmia database on CD-ROM and software for use with it, IEEE Comput. Cardiol. Proceed., 1990;185–188
[11] Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P, A wavelet-based ECG delineator: evaluation on standard databases. IEEE Trans. Biomed. Eng. 2004; 51:570–581

Address for correspondence.

Piotr Augustyniak
AGH University of Science and Technology
30 Mickiewicz Ave., 30-059 Kraków, Poland.
E-mail: august@agh.edu.pl