

A New Method for ECG Signal Feature Extraction

Adam Szczepański¹, Khalid Saeed¹, and Alois Ferscha²

¹ AGH University of Science and Technology
Faculty of Physics and Applied Computer Science
Al. Mickiewicza 30, PL-30059 Cracow, Poland

{aszczepa,saeed}@agh.edu.pl

² Johannes Kepler University Linz, Institute for Pervasive Computing
Altenberger Strasse 69, 4040 Linz, Austria
ferscha@pervasive.jku.at

Abstract. The paper concentrates on electrocardiographic (ECG) signal analysis for the sake of acquiring proper heartbeat datasets for further detection and recognition of anomalies in an ECG signal. The analysis is based on the distribution of the voltage extreme values of the signal and the time distribution of proper extreme values. The main assumption of the work is creation of as efficient and fast algorithm as possible. The main part of the analysis is carried out on the smoothened signal with the original signal used as the reference for the purpose of keeping desired precision of proper heartbeat identification in the situations when during smoothening necessary data is lost. The proposed method of feature extraction will be used to create the distinctive normal heartbeat samples of patients and the average pattern of the patients' heartbeat for reference purposes in analysis of the real time signal obtained from 1-lead holter carried by a patient constantly.

Keywords: ECG signal, signal analysis, 1-lead holter, ECG feature extraction.

1 Introduction

First clear electrocardiographic (ECG) signal was obtained over a century ago in 1895 by Willem Einthoven [1]. Since then ECG analysis became a standard and a trusted technique of detecting cardiac defects. ECG measurement is now used both for monitoring and diagnosis as a detection method of vast range of anomalies and the examination lasts 24 hours and 12-lead holter is used as a measuring device [2].

In contrast to hospital diagnostics, studies described in this paper are intended to lead to a development of 1-lead holter which will be carried by people with high risk of myocardial infarction or other rapid life threatening cardiac injury to indicate alert immediately when the life threatening situation occurs. Therefore, during the first stage of the project, algorithms for isolating single heartbeats

and obtaining reference sequences of heartbeats to create the average pattern of ECG of the patient and the heartbeat profile of the patient from signal are proposed.

2 ECG Signal Representation

The schematic of a single heartbeat in ECG signal and the actual recording are presented in Fig. 1. The actual signal has a large number of local extreme values.

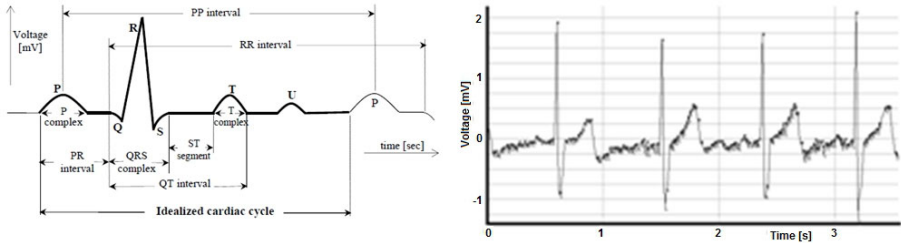


Fig. 1. Schematic (left) [3] and actual (right) ECG signals. On the actual signals extreme values are highlighted.

Points presented on the signal as P, Q, R, S and T are the main reference points in interpretation of ECG signal [4,5]. The U Wave which may appear in some people's heartbeats after T Wave may be skipped in further analysis [6].

What Could be Measured? The most important values to be measured are regularity of the signal, the distances between characteristic points and interferences in the shape of single heartbeat [4,7].

Signal may also contain artifacts which are generated by external factors such as the movement of electrodes and the chest muscles' activity. This should be detected and excluded before the analysis in order to identify the exact proper ECG signal position [2].

In the case of regularity of the signal R-R intervals, P-P intervals and bits per minute (bpm) are usually measured [4,8] and in the case of signal shape PR, QRS and QT interval could be measured and the extreme values of single heartbeat are checked for any changes in the average heartbeat. In many cases of cardiac diseases some parts of the signal, especially P Wave, is by-passed and is covered by T Wave from previous heartbeat. Some parts of the heart could also inversely polarize, leading to inverted parts of the signal [9,10]. Therefore the main goal of the real time heartbeat analysis shall be the detection of the anomalies in the normal rhythm by the detection of the characteristic signal points and calculation of their correlations. Parameters of ECG signal should be treated individually for every patient. Normal ECG values are described in [4,11]. In the system that will detect rapid cardiac malfunctions the main goal is to detect any rhythm of the heart and the shape of the signal that departs from average and to report the need of immediate medical assistance of the person.

3 Analysis of the Signal

This paragraph describes the algorithm based on particular sample analysis. The algorithm was tested on 15 recordings from [5] and 54 recordings (2 signals each) taken from [12]. Both the concept and the implementation of the algorithm are new and presented for the first time. The ECG signal feature extraction algorithm consists of 4 steps (Fig. 2).

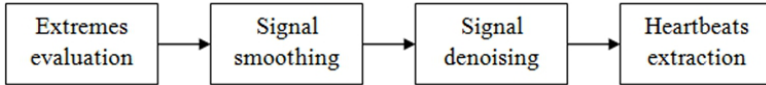


Fig. 2. The flowchart of the algorithm

If any constant time value is used in the algorithm as a constraint, it is calculated experimentally on the basis of the typical average ECG signal values [4,10]. All the millivolt values depend on the mV span of the input signal.

The signal used for demonstration purposes in this paper lasts 43 minutes and has the frequency of 200 samples per second. The signal has both regular and irregular heartbeats and some artifacts [5].

3.1 Smoothing of the Sample

Table 1 shows the distribution of the digital minima and maxima in the exemplary signal. The split of the signal into not more than 15 ranges was deduced experimentally. R and S points of heartbeats are clearly distinguishable but high density of extremes, previously shown in Fig. 1, leads to difficulties in detecting P and T. It is also noticeable that the position of Q point is not always obvious. That is why the smoothing of the signal is necessary.

Elimination of Strongest Artifacts. The range with most minima is chosen as the starting range of the true signal. Then the most probable span containing proper ECG signal is calculated using the method presented in Fig. 3.

The counter of the regions that does not meet the criteria was necessary because if the signal is clear, often the R and S points are located after about 2 range gap of regions without or with insignificant number of extremes so the algorithm should not stop the search when it approaches the first range that does not meet the criteria of number of extremes inside it.

The next step is the elimination of all the minima that are above minima range and all of the maxima that are below the maxima range. Then the ranges of minima and maxima are merged into one range and it is widened by 1/3 of the single range span.

The signal is analyzed again and all the samples that are not extremes and are not in the range are marked as deleted and the extremes are checked whether their values are in -3 mV to 3 mV range and, if so, another check is made to

Table 1. Distribution of minima and maxima of the exemplary ECG signal in 15 ranges. Bottom and Top describe the value range to which extremes belong, Mean is the mean value of the extremes in the range and Count is the number of extremes. Highlighted ranges are the most probable location of the proper signal. The 1 mV span has the range of 4927 digital units [13].

Minima total: 125555				Maxima total: 125545				
No	Bottom	Top	Mean	Count	Bottom	Top	Mean	Count
1	-25471	-22077	-25270	43	-25471	-22077	-25234	43
2	-22076	-18682	-20224	14	-22076	-18682	-20249	10
3	-18681	-15287	-16792	19	-18681	-15287	-16733	17
4	-15286	-11892	-13380	19	-15286	-11892	-13416	10
5	-11891	-8497	-9960	87	-11891	-8497	-10019	57
6	-8496	-5102	-5721	1909	-8496	-5102	-6142	174
7	-5101	-1707	-2592	8335	-5101	-1707	-2448	4740
8	-1706	1688	-343	105802	-1706	1688	-191	105662
9	1689	5083	2306	9076	1689	5083	2454	12032
10	5084	8478	6321	188	5084	8478	7341	664
11	8479	11873	9335	44	8479	11873	9515	2086
12	11874	15268	12946	9	11874	15268	12924	33
13	15269	18663	16811	7	15269	18663	16941	8
14	18664	22058	0	0	18664	22058	20660	3
15	22059	25455	23996	3	22059	25455	24283	6

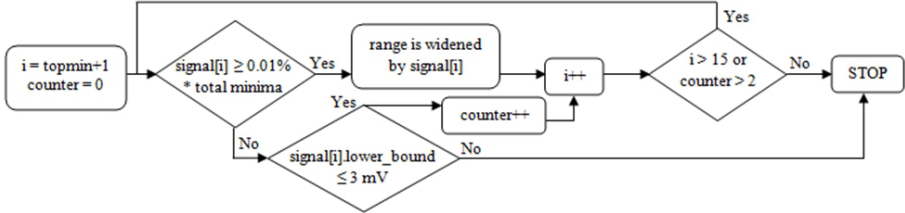


Fig. 3. Flowchart showing the appending of the signal range for the regions upper than the region with most minima. The same method, but without counter, is used for the regions lower than the region with most minima. For the maxima ranges the methods are the same. Topmin is the index of the region with most minima (8 in Table 1).

find out if the nearest neighbors of them are in the calculated range. If any of the nearest neighbors is not in the range, extreme is marked as deleted. Also whenever a point is marked as deleted, all the points in $(signal\ frequency)/5$ range on both sides of this point are marked as deleted. The reason of this is that high voltage artifacts have irregular shape or are often surrounded by lower voltage anomalies which are left in the signal if only points outside the range are eliminated, as shown in Fig. 4.

The Selected Method of Smoothing. When the signal true range is calculated and high voltage artifacts are removed, the signal is smoothed to lower the

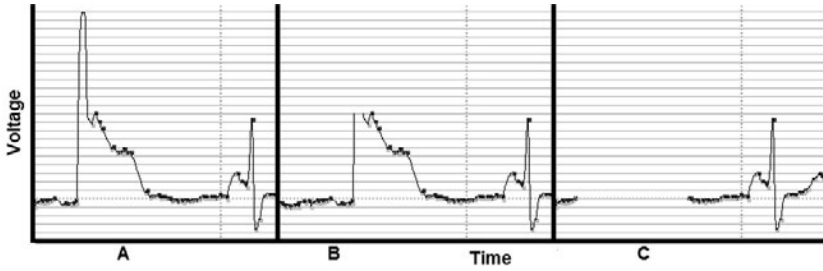


Fig. 4. The original signal (A), the signal after filtering of points that are not in range (B) and the signal after filtering also the neighbors (C)

number of extremes especially in low voltage part of the signal. For that purpose the variation of the method of bidirectional moving average [14] is chosen. The new sample value is calculated from (1):

$$x(i) = \frac{\sum_{k=-t}^t y(k)}{2 * t + 1}. \tag{1}$$

In (1) y is the original signal, x is the estimated signal and t is 1% of the signal frequency rounded up. The proposed method of estimating how many smoothing cycles should be applied is to smooth the sample as long as the change in the number of extremes before and after smoothing is less than 1% of the original number of extremes plus two more cycles. Also in the exemplary sample if the minimum value is lower than -0.5 mV or the maximum value is greater than 0.7 mV the extreme is not smoothed. These voltage constraints depend on the voltage span of the signal.

Table 2 shows the ranges left in the exemplary sample after the described procedure.

Table 2. Distribution of minima and maxima of the smoothed ECG signal

Minima total: 13262					Maxima total: 13407			
No	Bottom	Top	Mean	Count	Bottom	Top	Mean	Count
6	-8496	-5102	-5691	1743	-8496	-5102	-5972	39
7	-5101	-1707	-3264	2342	-5101	-1707	-2939	567
8	-1706	1688	-606	8988	-1706	1688	-218	7875
9	1689	5083	3309	138	1689	5083	2503	2382
10	5084	8478	6094	39	5084	8478	7705	497
11	8479	11873	0	0	8479	11873	9516	2033
12	11874	15268	0	0	11874	15268	12415	14

3.2 Searching for Heartbeats

After smoothing in the possible location of S and R points there may be more than one minimum and maximum (Fig. 5).

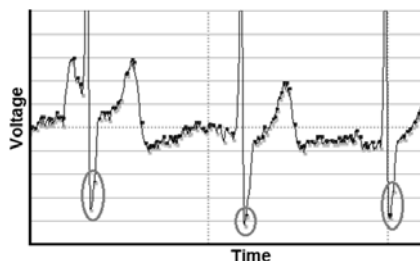


Fig. 5. Distribution of maxima and minima in the lower part of the signal. Ovals on the picture highlight the problem.

This problem is solved during the acquisition of R and S points by searching the top minimum value for S and maximum for R in the short span around the obtained point.

At first the algorithm first finds R-S pairs of the heartbeats. This is acquired by searching the signal for the pairs maximum-minimum which are not more than $1/8$ seconds away and the mV span of the pair is greater than 50% of the current signal range. After that the right side of the S point is checked for T point. T point is considered to be the biggest larger than 0 maximum which is not further than $1/3$ of the second away from R point. The left side of the R point is searched for the existence of Q point. Q point is considered as the lowest minimum in the range of $1/6$ second from R point but the mV span of Q-R pair should be not less than 60% of the span of the positive part of ECG signal. The P point is the largest maximum in the $1/4$ second span on the left side of the R point as the highest maximum before Q point.

3.3 Obtaining the Reference Signal

The signal is searched for the sequences of regular heartbeats. The average length of the heartbeat is calculated at first and then for all heartbeats the nearest neighbors from left and right are found. If there are any deleted earlier points between two heartbeats, this site of heartbeat is marked as not having the neighbor. Also all the distances greater than twice the signal frequency are eliminated. The shorter distances are checked whether between the end of the one signal and P point of the other signal there are any values less than the end point value or greater than P point value. If yes, the neighboring heartbeats are marked as not having neighbors on that side. Next the values of the heartbeat points are retrieved from the original signal. This is done by searching for the nearest maximum for P, R and T points and nearest minimum for Q and S points.

4 Results

Resulting number of properly acquired heartbeats is around 99.43% of total heartbeats in the signals. Results for exemplary samples as long as the results for all samples are shown in Table 3.

Table 3. The results

Name	Bits total	Detected beats	Rate of detection (%)
Worst 200 Hz sample	1322	1254	94.86%
Best 200 Hz sample	1976	1976	100%
Worst 1000 Hz sample	227	247	91.90%
Best 1000 Hz sample	3148	3148	100%
15 200 Hz samples	23454	23710	98.92%
108 1000 Hz samples	90013	89613	99.56%
Total (123 signals)	113072	90013	99.43%

The obtained results are comparable to the results of other popular algorithms like in [15] where 97.65% rate of detection was acquired, [16] with 99.85% rate of detection and [17] with 99.76% rate of detection.

On the contrary to those algorithms, the proposed algorithm does not use any filtering, wavelet transforms or slope detection techniques which leads to the acquisition of fast calculations' speed.

5 Conclusions and Future Work

The obtained efficiency of the algorithm is comparable to other commonly used algorithms. The main advantage of the proposed algorithm is that it does not need any other prefiltering than what may be provided by the signal measurement hardware. Also the lack of any software filtering means that when the system is working in real time it needs only a short (around 1 second length) buffer before it can start the analysis.

The next step of the studies shall lead to determining how long the ECG signal should be to extract the proper average of the heartbeats of the patient. Also fragments of the ECG signal should be connected with patient activities to determine correct patterns for all the possible stress and effort conditions.

This will eventually lead to the development of a 1-lead holter that would monitor the patient state constantly.

Acknowledgement

The research is supported within the FP7 project SOCIONICAL, No. 231288.

References

1. ECG Library, <http://www.ecglibrary.com/ecghist.html>
2. Adamec, J., Adamec, R.: ECG Holter. Springer Science+Business Media, LLC, New York (2008)
3. Gawroski, P., Saeed, K., Kuakowski, K.: Early warning of cardiac problems in crowd. In: Jędrzejowicz, P., Nguyen, N.T., Howlet, R.J., Jain, L.C. (eds.) KES-AMSTA 2010. LNCS (LNAI), vol. 6071, pp. 220–229. Springer, Heidelberg (2010)
4. Jones, S.A.: ECG Notes: Interpretation and Management Guide. F. A. Davies, Philadelphia (2005)
5. Riener, A., Ferscha, A., Aly, M.: Heart on the road: HRV analysis for monitoring a driver's affective state. In: Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2009), pp. 99–106 (2009)
6. Stouffer, G.A.: Practical ECG Interpretation: Clues to Heart Disease in Young Adults. Wiley-Blackwell, UK (2009)
7. Van Mieghem, C., Sabbe, M., Knockaert, D.: The Clinical Value of the ECG in Noncardiac Conditions. *Chest* 125, 1561–1576 (2004)
8. Trahanias, P., Skordalakis, E.: Syntactic pattern recognition of the ECG. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 648–657 (1990)
9. Hampton, J.R.: The ECG Made Easy, 6th edn. Churchill Livingstone, New York (2003)
10. Khan, M.G.: Rapid ECG Interpretation, 3rd edn. Humana Press, New Jersey (2008)
11. Hung-Chi, L.: ECG in the Child and Adolescent. Blackwell Publishing, Massachusetts (2006)
12. Non-Invasive Fetal Electrocardiogram Database, mothers' part of the signal was used for analysis, <http://www.physionet.org>
13. Kemp, B., Vrri, A., Rosa, A.C., Nielsen, K.D., Gade, J.: A simple format for exchange of digitized polygraphic recordings. *Electroencephalography and Clinical Neurophysiology* 82, 391–393 (1992)
14. Chou, Y.: Statistical Analysis. Holt Reinhart Winston, New York (1975)
15. Haque, M.A., Rahman, M.E., Al Sayeed, C.A., Uddin, B.M.Z.: A Fast Algorithm in Detecting ECG Characteristic Points. In: Proceedings of Second International Conference on Electrical and Computer Engineering (ICECE 2002), Dhaka, Bangladesh, December 26–28, pp. 160–163 (2002)
16. Cuiwei, L., Chongxun, Z., Changfeng, T.: Detection of ECG Characteristic Points Using Wavelet Transforms. *IEEE Transactions on Biomedical Engineering* 42, 21–28 (1995)
17. Chiarugi, F., Sakkalis, V., Emmanouilidou, D., Krontiris, T., Varanini, M., Tollis, I.: Adaptive Threshold QRS Detector with Best Channel Selection Based on a Noise Rating System. In: Computers in Cardiology 2007, pp. 157–160 (2007)