

An Approach for ECG Feature Extraction using Daubechies 4 (DB4) Wavelet

Muhidin A. Mohamed
Faculty of Electrical Engineering
Jamhuriya University - JUST
Mogadishu, Somalia

Mohamed A. Deriche
Department of Electrical Engineering
King Fahd University – KFUPM
Dhahran, Saudi Arabia

ABSTRACT

An Electrocardiogram (ECG) signal describes the electrical activity of the heart recorded by electrodes placed on the surface of human body. It summarizes an important electrical activity used for the primary diagnosis of heart abnormalities such as Tachycardia, Bradycardia, Normalcy, Regularity and Heart Rate Variation. The most clinically useful information of the ECG signal is found in the time intervals between its consecutive waves and amplitudes defined by its features. In this paper, an ECG feature extraction algorithm based on Daubechies Wavelet Transform is presented. DB4 Wavelet is selected due to the similarity of its scaling function to the shape of the ECG signal. R peaks detection is the core of this algorithm's feature extraction. All other primary peaks are extracted with respect to the location of R peaks through creating windows proportional to their normal intervals. The proposed extraction algorithm is evaluated on MIT-BIH Arrhythmia Database. Experimental results indicate that the algorithm can successfully detect and extract all the primary features with a deviation error of less than 10%.

General Terms

Cardiac Signal Processing, Digital filters, Digital signals

Keywords

ECG feature extraction, Daubechies Wavelets, cardiac signal

1. INTRODUCTION

The study of electrocardiogram (ECG) signals has been a growing research topic due to its role in diagnosing many cardiac diseases [1, 2]. The ECG signal is a "graphic record of the direction and magnitude of the electrical activity that is generated by the depolarization and repolarization of the atria and ventricles" [3, 4]. Each full cardiac cycle is represented by an ECG signal that consists of P-QRS-T waves. Figure 1 shows a normal ECG signal highlighting its main features in terms of its primary waves and time intervals. The majority of the medically useful information in the ECG is originated from the intervals and amplitudes defined by its features (characteristic wave peaks and time durations) [5]. Typical normal values for ECG amplitudes and durations of a healthy normal person are summarized in Table 1 and Table 2 respectively [6]. The improvement of a precise and rapid methods for automatic ECG feature extraction is of great importance, particularly for the examination of long recordings [3]. There is a significant research effort paid to the investigation of methods for extracting useful features carrying medical information from ECG signals. Mazomenos, Evangelos B. et al.[5] presented a low complexity algorithm for extracting fiducial points from ECG signals using Discrete Wavelet Transform (DWT). Vaneghi et al [7] compared six distinct approaches for an ECG signal feature extraction. Their findings revealed that the Eigenvector method

outperformed the other 5 presented schemes. One of the closest studies to ours is explored by Espiritu-Santo-Rincon et al [8]. Similar to this work, they have devised a mechanism for detecting R peaks followed by the rest of the signal features. However their method is totally based on Wave segmentation. A number of other techniques have been proposed in [9] for the detection of ECG features.

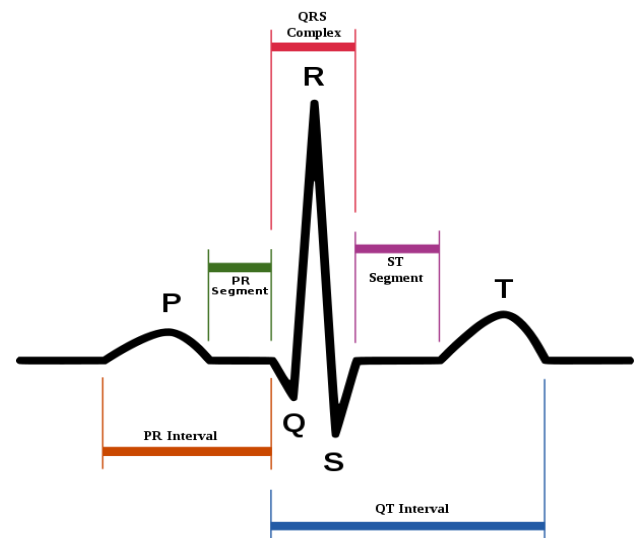


Figure 1: Typical ECG Signal¹

Most of the aforementioned methods for ECG signal analysis were based on the time domain. But it is not always adequate to study all the features of ECG signals in the time domain. Therefore, the frequency representation of a signal is equally important and its processing is more appealing. Current study is inspired by the ECG feature extraction algorithm proposed by Mahmoodabadi et al. [3]. Daubechies family of wavelet is used for the analysis and decomposition of the signal [10]. Current study's main objective is the extraction of primary features of the denoised and decomposed ECG signals. By extracting these primary ECG features, it is thought that some fundamental parameters such as the amplitudes of the waves and their durations such as RR, QRS, and PR intervals can be easily obtained which is then used for subsequent automatic analysis.

¹ <http://a-fib.com/treatments-for-atrial-fibrillation/diagnostic-testing/the-ekg-signal/>

Table 1: Normal ECG Wave Amplitudes

Wave	Typical Amplitude
P wave	0.25 mV
R wave	1.60 mV
Q wave	25% of R wave
T wave	0.1 to 0.5 mV

Table 2: Normal ECG Wave Durations

P-R interval	0.12 to 0.20 second
Q-T interval	0.35 to 0.44 second
S-T segment	0.05 to 0.15 second
P wave interval	0.11 second
QRS interval	0.09 second

2. EXPERIMENTS

2.1 Dataset

To validate the proposed extraction algorithm, MIT-BIH Arrhythmia Database is used. It is a standard and publicly available dataset which is developed for testing and evaluating research works on arrhythmia detectors and cardiac dynamics [11]. The database contains 48 30-minute extracts of two-channel ambulatory ECG recordings gathered from 47 participants. The recordings have been taken between 1975 and 1979 at BIH Arrhythmia Laboratory. As the current work heavily utilizes essential feature extraction from ECG signals, MIT-BIH database has been selected on its fitness for the role.

2.2 ECG Feature Extraction

The overall proposed feature extraction algorithm is best summarized in Figure 2. As recorded ECG signals are contaminated with noise, the original signal is first decomposed to smoothen it by removing the contaminating noise and picking appropriate approximation coefficients. The wavelet decomposition process down samples the signal by transforming the samples to a much lower frequency than the original signal. Consequently the details are reduced and QRS complex is preserved. Detection is made easier with reduced details. The signal is decomposed up to the fourth level beyond which the R peak detection is performed. Detection of R peaks in the down sampled approximated coefficients is achieved using a threshold which is set to be half the highest QRS complex of the down sampled signal. The plausibility of picking a point in the R wave other than its peak led to the inclusion of the next step after the R detection (fourth block in Figure 2). It can be closely observed in Figure 3 and Figure 4 that the R-peak is not a single impulse peak, but rather multiple points in the same peak satisfying the criteria of being higher than the selected threshold.

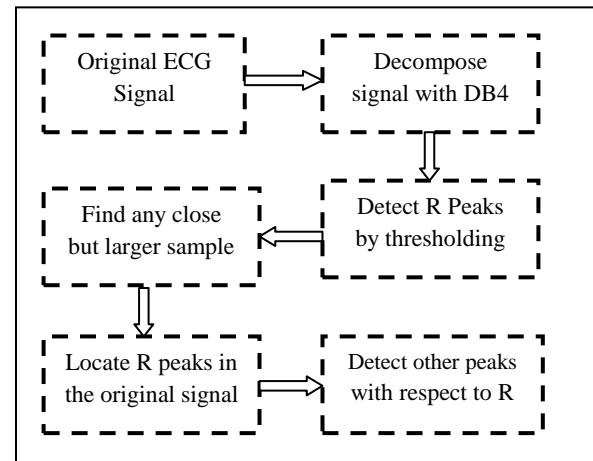


Figure 2: Proposed ECG Feature Extraction Algorithm

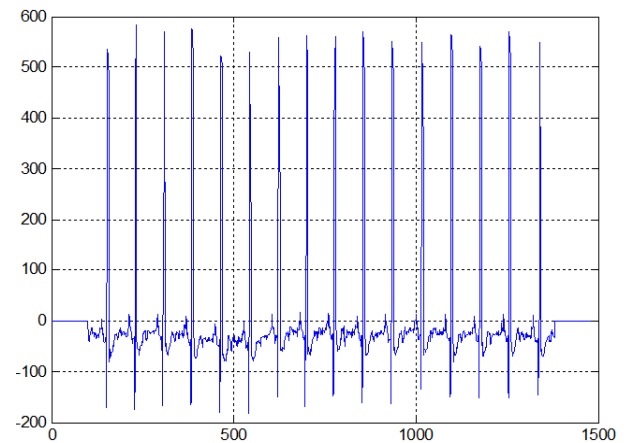


Figure 3: Original un-decomposed signal

Technically speaking, R-peak detection process is performed on the down sampled signal which requires locating any extracted R signal in the original ECG signal. After successfully detecting R peaks, the remaining P, T, Q and S waves are determined with respect to the R peaks. For the feature extraction and analysis, ECG signals of both Normal Sinus Rhythm and ST-Elevated signals are considered in the first round of experiments. Figure 4 demonstrates reconstructed signal of a selected MIT-BIH record. A close scrutiny to the plots reveals that the frequency bands are separated and the reconstruction of the decomposition coefficients 1 to 4 are smoother and cleaner as compared to the original signal shown in Figure 3. Nevertheless, as decomposition levels increase, they will have less number of samples than the original signal due to the down sampling. One can see that first signal (top left Figure 4) resembles the original signal but has exactly one fourth the number of samples because the signal is decomposed in 4 levels. 2nd level has exactly half the number of samples of 1st level, and 3rd level has exactly half the number of samples of the 2nd level and so on.

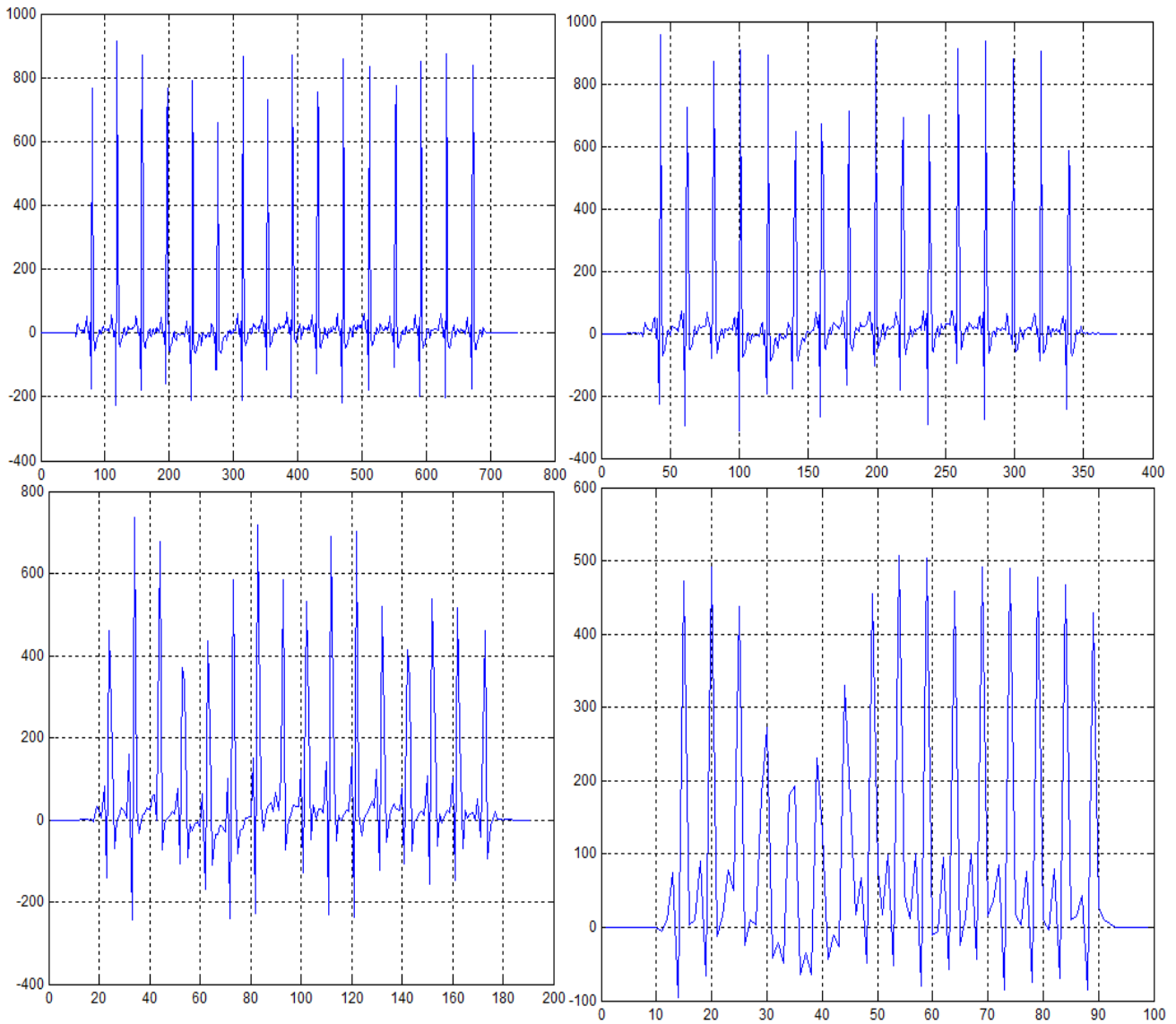


Figure 4: Reconstructed ECG signal from the 4th Level DB4 Decomposition

2.3 ECG Waves' Detection

The study's approach of the ECG waves' detection is centred on the R-Peak extraction. Therefore, maximizing the detection accuracy of this cornerstone wave is a prerequisite for the entire extraction accuracy. This is to say that all other waves of the ECG signal including T, P, Q and S waves can be easily located with reference to R-peak and with the assumption that wave intervals and segment durations are known as per tables 1 and 2. The third block of the extraction algorithm summary given in Figure 2 encapsulates the R peak detection. In this section, this block is further expanded to illustrate the actual R peak detection mechanism (cf. Figure 5).

As stated, all values exceeding the threshold are organised in a vector called R peaks. As decomposed signals are noise free, the second level approximation coefficients (top right Figure 4) is selected for further processing. This means that the initial R peaks are extracted from the down sampled signal. However, the ultimate goal is to detect the R Peaks in the original signal. So the strategy here is to first detect the R peaks in the down sampled signal and then cross verify those points in the actual signal. Since fourth level wavelet

decomposition is used, the detected R waves in the down sampled signal can be mapped to the original signal by multiplying 4 to their location indices. The R wave extraction procedure is demonstrated in the flow chart given in Figure 5.

3. RESULTS AND DISCUSSION

The implementation of the R peak extraction algorithm as described in Figure 5 enabled the detection of ECG features. The analysis made during the algorithm evaluation and testing is confined to Normal Sinus Rhythm and ST-Elevated signals of the MIT-BIH Arrhythmia Database. Figure 6² elaborates the extracted features in terms of the P, R and T waves on the original ECG signal. Most of the clinically useful information in the ECG is found in the intervals and amplitudes defined by its features. Therefore the actual amplitude can be determined by referring to the digitized data of the processed database. If you observe the raw amplitudes of the signal in Figure 6, it

² Figure 6 comes after 7 for the suitability of the latter to fit double column page while Figure 6 needs to span across two columns.

goes to around 600mV, a figure which is far larger than the expected physical quantity in millivolts. The relationship in equation 1 is used to obtain the actual amplitude in the desired physical unit from the raw data. It is noteworthy that the ADC³ gain is specified for each database record in the information header files of MIT-BIH Arrhythmia Database.

$$\text{ActualAmplitude(mV)} = \frac{\text{Raw Amplitude}}{\text{ADCGain}} \quad (1)$$

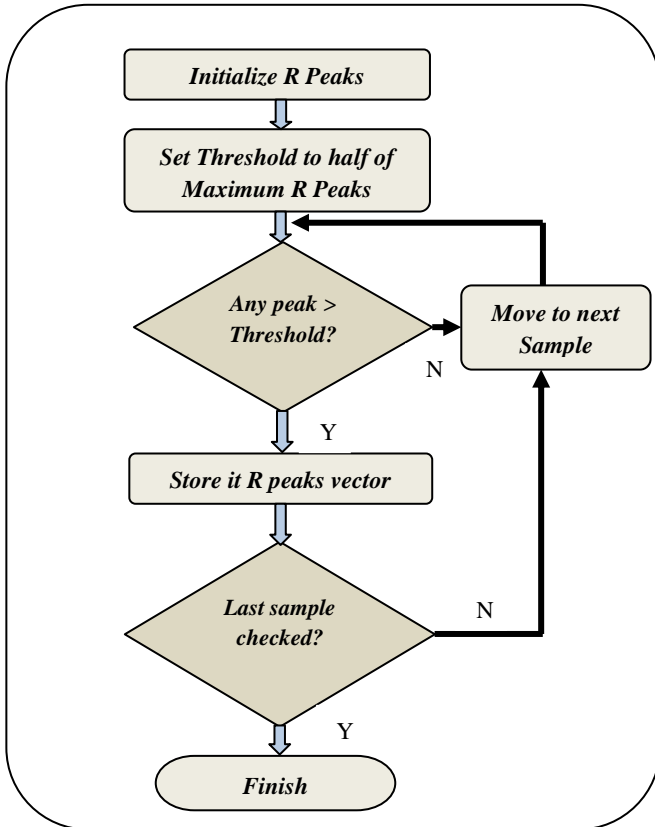


Figure 5: R Peak Detection Flow Chart

Another group of fundamental extracted features are the durations between successive waves with the most important of all being the R-R interval. It is from these duration intervals where the heart rate can be easily determined. The basic idea used in determining this interval is based on the sampling frequency and the number of samples between any two successive R peaks. The average number of samples between consecutive R peaks was calculated. This number is then multiplied by the sampling time to obtain the R-R interval.

Equation 2 is employed for the computation assuming that the average number of samples between two consecutive R peaks is N_s . The results of randomly selected 5 MIT-BIH records and their declared decisions are summarized in Table 3.

$$R - R \text{ Interval(seconds)} = \frac{1}{F_s} * N_s \quad (2)$$

The decision of declaring a given record to be normal or abnormal is based the deviation of the values of the extracted parameters from the known normal ECG features. For

instance the record **301** was declared abnormal as two of its features namely the R and P wave amplitudes greatly deviate from the normal values. Similar logic applies to record 310. The remaining database records showed considerable normality as compared to normal values in Table 1 and Table 2. One can compare parameter values in Table 3 with tables 1 and 2 to observe the similarity and discrepancies. To evaluate the performance of the algorithm, an excerpt containing the actual recording time for MIT-BIH record was taken from the Physionet website to compare with the obtained average R-R interval. Through carefully scrutinized observation, one can glance at Fig. 7, which shows an excerpt of the recording time for record 18184, and note that the average of the actual times is approximately 0.71 seconds. The experimentally extracted R-R interval for the same record was found to be 0.7 seconds as indicated in the second row Table 3. This results in a small deviation error of 0.01 seconds. To this end, a complete ECG features extraction is achieved where the detection of the primary waves facilitated the computation of time intervals.

t0	b0	RR (sec)	b1	t1
[09:34:00.000]	[0]	0.609	N	[09:34:00.609]
[09:34:00.609]	N	0.758	N	[09:34:01.367]
[09:34:01.367]	N	0.805	N	[09:34:02.172]
[09:34:02.172]	N	0.805	N	[09:34:02.977]
[09:34:02.977]	N	0.773	N	[09:34:03.750]
[09:34:03.750]	N	0.727	N	[09:34:04.477]
[09:34:04.477]	N	0.711	N	[09:34:05.188]
[09:34:05.188]	N	0.664	N	[09:34:05.852]
[09:34:05.852]	N	0.711	N	[09:34:06.563]
[09:34:06.563]	N	0.680	N	[09:34:07.242]
[09:34:07.242]	N	0.688	N	[09:34:07.930]
[09:34:07.930]	N	0.680	N	[09:34:08.609]
[09:34:08.609]	N	0.680	N	[09:34:09.289]
[09:34:09.289]	N	0.688	N	[09:34:09.977]

Fig. 7: Excerpt of original recording times for Record MIT-BIH 18184⁴

4. CONCLUSION

In this paper, we described our implementation and evaluation of an ECG feature extraction algorithm using Daubechies Wavelet. The ECG signal is first smoothed by decomposing and omitting much of the details while keeping the 4th level approximation coefficients. The R peaks are then extracted from the down sampled clean ECG signal. This is followed by locating the detected R peaks in the original signal which has enabled positioning other peaks (P, T, Q and S) with reference to the identified R peaks. The proposed wavelet based feature extraction algorithm achieved a high detection performance on the MIT-BIH Arrhythmia Database in terms of interval and amplitude extraction. The detection algorithm encounters a very low deviation error found through the comparison of extracted RR intervals with real and original recording times of corresponding database records.

³ Analog to digital conversion

⁴ www.physionet.org

Table 3: Extracted Features from 5 MIT-BIH Database records

MIT Record	R Wave Amplitude	P Wave Amplitude	T wave Amplitude	RR Interval	Decided Result
18184	1.5	0.23	0.11	0.70	Normal
301	0.95	-0.1	0.02	0.66	Abnormal
16786	2.5	0.275	0.6	0.82	Normal
310	3.96	0.22	1.35	0.66	Abnormal
16420	1.75	0.24	0.09	0.66	Normal

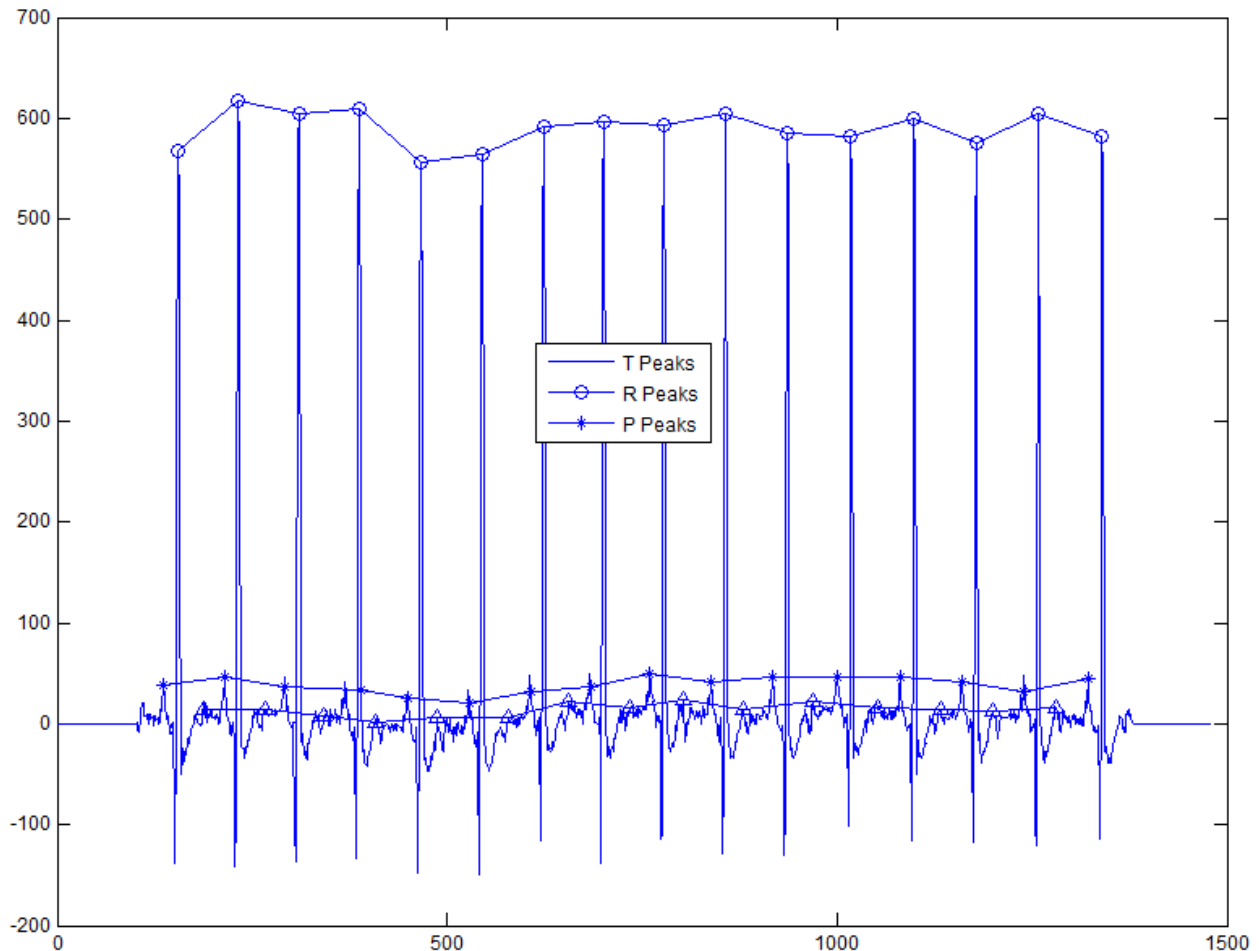


Figure 6: Detection of the ECG Primary Waves (P, R and T)

5. REFERENCES

- [1] K. Huang and L. Zhang, "Cardiology knowledge free ECG feature extraction using generalized tensor rank one discriminant analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, pp. 1-15, 2014.
- [2] J. L. Garvey, "ECG techniques and technologies," *Emergency medicine clinics of North America*, vol. 24, pp. 209-225, 2006.
- [3] S. Mahmoodabadi, A. Ahmadian, and M. Abolhasani, "ECG feature extraction using Daubechies wavelets," in *Proceedings of the fifth IASTED International conference on Visualization, Imaging and Image Processing*, 2005, pp. 343-348.
- [4] Q. Zhao and L. Zhang, "ECG feature extraction and classification using wavelet transform and support vector machines," in *Neural Networks and Brain, 2005. ICNN&B'05. International Conference on*, 2005, pp. 1089-1092.
- [5] E. B. Mazomenos, D. Biswas, A. Acharyya, T. Chen, K. Maharatna, J. Rosengarten, J. Morgan, and N. Curzen, "A low-complexity ECG feature extraction algorithm for mobile healthcare applications," *IEEE JOURNAL OF*

BIOMEDICAL AND HEALTH INFORMATICS, vol. 17, p. 459, 2013.

- [6] B. Salsekar and A. Wadhvani, "FILTERING OF ECG SIGNAL USING BUTTERWORTH FILTER AND ITS FEATURE EXTRACTION," *International Journal of Engineering Science & Technology*, vol. 4, 2012.
- [7] F. M. Vaneghi, M. Oladazimi, F. Shiman, A. Kordi, M. Safari, and F. Ibrahim, "A comparative approach to ECG feature extraction methods," in *Proceedings of 3rd international conference on intelligent systems modeling and simulation*, 2012, pp. 08-10.
- [8] A. Espiritu-Santo-Rincon and C. Carbajal-Fernandez, "ECG feature extraction via waveform segmentation," in *Electrical Engineering Computing Science and Automatic Control (CCE), 2010 7th International Conference on*, 2010, pp. 250-255.
- [9] S. Karpagachelvi, M. Arthanari, and M. Sivakumar, "ECG feature extraction techniques-a survey approach," *arXiv preprint arXiv:1005.0957*, 2010.
- [10] J. Saraswathy, M. Hariharan, V. Vijejan, S. Yaacob, and W. Khairunizam, "Performance comparison of Daubechies wavelet family in infant cry classification," in *Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on*, 2012, pp. 451-455.
- [11] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 20, pp. 45-50, 2001.