# Electrocardiogram Waveform Feature Extraction Using the Matched Filter

Felipe E. Olvera, Jr., *Student Member, IEEE*

***Abstract*— The matched filter was used to detect different signal features on an human heart electrocardiogram signal. The waveform features of interest were the QRS Complex, the R-R intervals, and the ST segments of four different electrocardiogram signals. The detection of the QRS Complex and the R-R interval were compared for accuracy and used in determining the length of the heart beat interval which is necessary to determine the heart rate variability. The detection of the ST segment, which is a precursor of possible cardiac problems, was more difficult to extract using the matched filter due to noise and amplitude variability.**

***Index Terms*— Electrocardiogram, feature extraction, matched filter, signal preprocessing, heart rate variability.**

## I. Introduction

HUMANS have always been interested in the workings of the heart, whether physical or non-physical. The first bioelectric recorder named the capillary electrometer, was devised by Marey and Lippman in 1876 [1] and it was used to record the first human electrocardiogram in 1887 by Dr. Augustus Waller. The capillary electrometer used a solution of mercury and sulfuric acid. The meniscus formed by the mercury-sulfuric acid interface varied in height when an electric current was applied. The varying height of the meniscus was recorded onto photographic material and thus biological electrical activity could be recorded.

In 1901, the first paper on the string galvanometer was published by its inventor, Dr. Willem Einthoven, detailing the improvements in recording high fidelity human electrocardiograms [2]. This device superseded the capillary electrometer which was not able to provide the level of detail that the string galvanometer could. Dr. Einthoven won the 1924 Nobel Prize in Medicine for inventing the electrocardiograph.

Unfortunately, the string galvanometer was bulky, weighed more than 600 pounds, required five people to operate and required water cooling of the electromagnet assembly, similar to modern day high-power klystron technology. What the patient had to endure was even worse: hands and feet in glass cylinders containing a saline solution and the electrode was a zinc plate. Because of the size of the string galvanometer, it could not be used in a hospital room but had to be operated remotely through long wire connections. This problem was solved in 1926 when the first portable, 80 pound, electrocardiogram machine was introduced.

This work was completed as part of a course project for *Statistical Signal Processing II* at Portland State University during the Winter term of 2006.

F. Olvera is a graduate student at the Portland State University Maseeh College of Engineering and Computer Science.

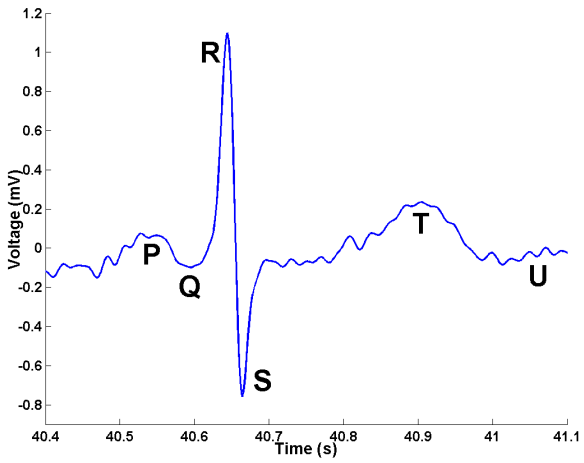


Fig. 1. The Electrocardiogram Waveform with Notated Features.

Nonetheless, early pioneers such as Dr. Horatio B. Williams were able to make significant discoveries using the string galvanometer. He created the first heart station at Columbia University College of Physicians and Surgeons in 1911, using this device to record electrocardiograms [2]. Dr. Williams was able to investigate the individual sections of the human electrocardiogram waveform and show how the QRS complex for example, is affected by conduction defects in the ventricles.

The electrocardiogram (ECG) waveform consists of several sections, labeled P, Q, R, S, T, and U, as shown in Fig. 1. Each section has its own characteristics and importance to the function of the human heart. The major sections are: the P-wave, the QRS complex, the T-wave, and the U-wave. The important morphological features are the QRS complex, the R-R interval, the P-R interval, the Q-T interval, the ST segment and the R-wave amplitude [3].

Each of these sections has significance on the health of the human heart. For example, in the QRS section, its normal time interval is from 0.04 to 0.09 seconds and any time interval greater than 0.1 seconds indicates a form of delay in the ventricular depolarization or beating of the heart.

Each electrocardiogram section can be extracted by using different methods but the only method used in this analysis was the matched filter specific to the ECG section of interest. Matched filters maximize the signal-to-noise ratio for a noisy signal so that the signal of interest can be extracted [4]. The sections extracted using the matched filter were the QRS complex, the ST segment and the R-wave peak.

The QRS complex and the R-wave peak are useful in examining the heart rate variability because the heart rate can be determined by the interval between the QRS complexes or the R-wave peaks. Healthy biological systems are not periodic but tend towards *disorder* and the human heart rate displays such variability [5]. Since the human heart beat varies slightly from beat to beat, analyzing this beat-to-beat variability leads to the detection of possible heart problems.

The R-wave peak is part of the QRS complex so both sections were used as the signal of interest in their respective matched filters to detect the individual heart beats in four different ECG signals. Since the R-wave signal is much smaller than the QRS complex, 20 milliseconds vs. 40 to 90 milliseconds, a comparison was made to determine whether using the R-wave peak or the QRS complex resulted in better R-R interval detection.

The more important section extracted using the matched filter was the ST segment. If the ST segment is elevated or has slope changes, this could be a precursor of an heart attack. Since the ST segment is usually low in voltage level in relation to the rest of the ECG signal and sometimes buried in noise, it is much harder to detect than the QRS complex or the R-wave peak and thus was a good test of the use of the matched filter in ECG signal feature extraction.

## II. Methodology

### A. Experimental Data

The electrocardiogram signals were obtained from the MIT-BIH database via the Physionet [6] web site [7]. The MIT-BIH database contains many data sets of electrocardiogram signals, mostly abnormal or unhealthy electrocardiograms, but it also contains normal electrocardiograms that can be used as a reference base. One electrocardiogram signal was selected from four different databases: 1) the MIT-BIH Arrhythmia Database, 2) the MIT-BIH Normal Sinus Rhythm Database, 3) the Sudden Cardiac Death Holter Database and, 4) the Long Term ST Database.

Each of the selected electrocardiogram signals is thirty minutes in length but only five-minute sections were used for signal processing as suggested by Szilágyi [8] and Malik *et al.* [9], to determine proper spectral components in short term ECG recordings that are important when analyzing the heart rate variability. The ECG signals were chosen on the basis of age. The Normal ECG used as a reference was recorded from a young woman so an effort was made to check each record in each of the databases to try and obtain an ECG recording made from someone within the same age group.

The ECG signal used from the Normal Sinus Rhythm Database was Record 16272. This signal was recorded from a 26-year old female who had a normal sinus rhythm with no history of heart problems or medication. This signal was used as the reference template for the features to be extracted from the other ECG signals. In the five-minute segment used, there were 371 heart beats.

The ECG signal used from the Arrhythmia Database was Record 113. This signal was recorded from a 24-year old female patient with a pacemaker to help her problem of sinus arrhythmia. She was not on any medication and her ECG signal was mostly normal with areas of R-R interval variability possibly caused by her atrial pacemaker. In the five-minute segment used, there were 351 heart beats.

The ECG signal used from the Sudden Cardiac Death Holter Database was Record 30. This signal was recorded from a 43-year old male with normal sinus rhythm. The five minute interval used from this record has a section of cardiac inactivity. No heart or medication history was available on this individual. This was the only ECG signal that did not conform to the age group selected. This was due to the lack of an available ECG signal in this age group. As expected, most of the ECG signals from this database were recorded from much older individuals.

Unfortunately, the number of heart beats in the five-minute segment was not available either via the cardiologist's annotations on the original database record or extracted by the matched filter due to the difficulty in determining what was a heart beat and what was not a heart beat.

The ECG signal used from the Long Term ST Database was Record s20081. This signal was recorded from a 39-year old female who suffered from ST segment flattening and downsloping along with T-wave inversions. These type of problems suggest ischemia or future heart attacks. She was not on any medication and had no history of heart problems. In the five-minute segment used, there were 426 heart beats.

### B. Procedure

Each ECG signal was preprocessed before the application of signal processing techniques. Each ECG signal was originally sampled at a different sampling rate. Record 16272 was originally sampled at 128 Hz, Records 30 and s20081 were originally sampled at 250 Hz and Record 113 was originally sampled at 360 Hz. In order to apply the signal processing techniques equally to each record and preserve the most information, each record was resampled to 500 Hz [10]. After resampling, the mean was removed from each ECG signal and then each ECG signal was passed through a Chebyshev Type 1, $6^{th}$-order low-pass filter to remove any signal components beyond 50 Hz.

From the Normal ECG record, an ECG signal template was extracted as shown in Fig. 1. Since each R-R interval varies slightly, the template interval was chosen on the purely subjective basis on how *ideal* it looked. From this template, each feature of the ECG waveform was used as the signal of interest for the specific feature's matched filter. The feature's causal matched filter was applied to each of the four ECG signals and the QRS complex, the R-wave peak, and the ST segment were extracted using the same detection method.

## III. Results

Each of the preprocessed ECG signals are shown in Fig. 2. Only 60 seconds of the 300 seconds of each ECG signal is shown so that individual heart beats are easily discernable.

Since resampling introduces aliasing errors, Record 16272 and Record 113 were checked for possible problems. The comparison of these ECG signals before and after resampling

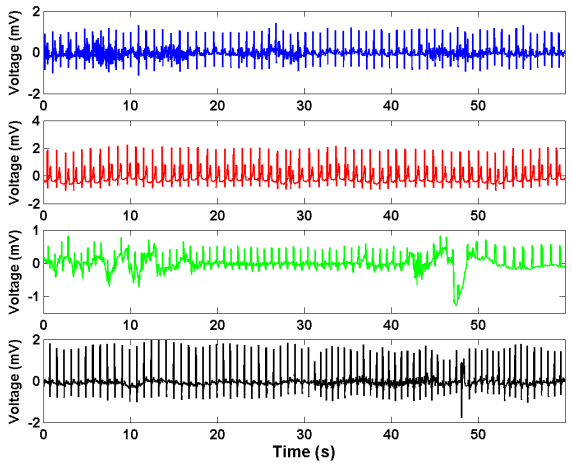


Fig. 2. The Normal, Arrhythmia, Cardiac Arrest, and Long Term ST ECGs.

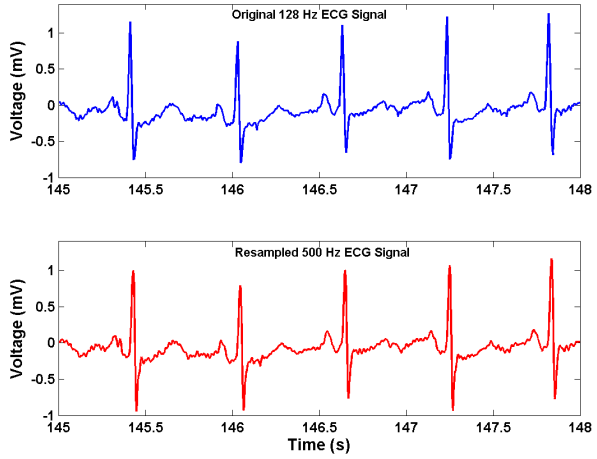


Fig. 3. The Normal ECG before and after resampling.

and low-pass filtering is shown in Fig. 3 and Fig. 4. There were only small amplitude changes in both ECG signals after resampling and more importantly, the ST segment noise did not increase significantly though the resampled ECG transition point of the S-wave did show more noise than in the original ECG signal, but not enough to require a different resampling frequency.

For the detection of the R-R interval, the matched filter outputs were better than expected for the Normal and Long Term ST ECGs but for the Arrhythmia and Sudden Cardiac Death (Cardiac Arrest) ECGs, the results were not as good. The results of the specific feature matched filter outputs are shown for each ECG in Fig. 5, Fig. 6, and Fig. 11.

The R-wave peaks and QRS complexes were detected using their respective matched filters and the results are shown in Fig. 7, Fig. 8, Fig. 9, and Fig. 10. With one exception, all of the R-wave peaks and QRS complexes were detected in the Normal ECG. On the Long Term ST ECG, all but one of the QRS complexes were detected and all but three of the R-wave peaks were detected.

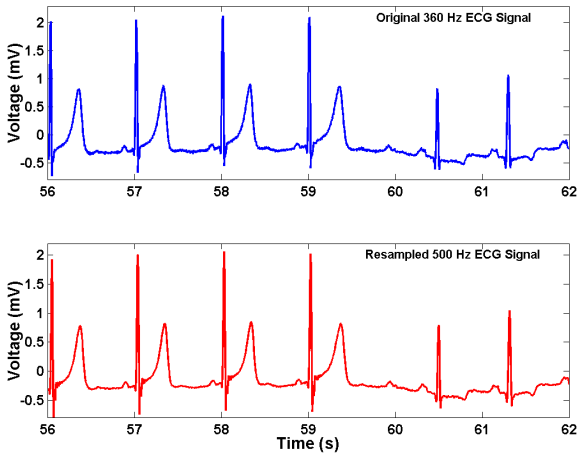


Fig. 4. The Arrhythmia ECG before and after resampling.

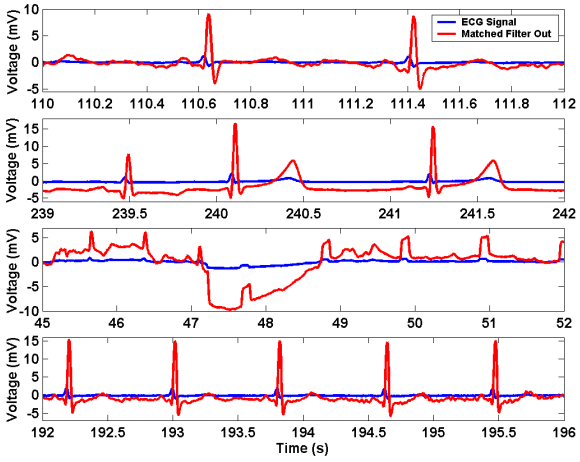


Fig. 5. The R-wave Peak Matched Filter Output for the Normal, Arrhythmia, Cardiac Arrest, and Long Term ST ECGs.

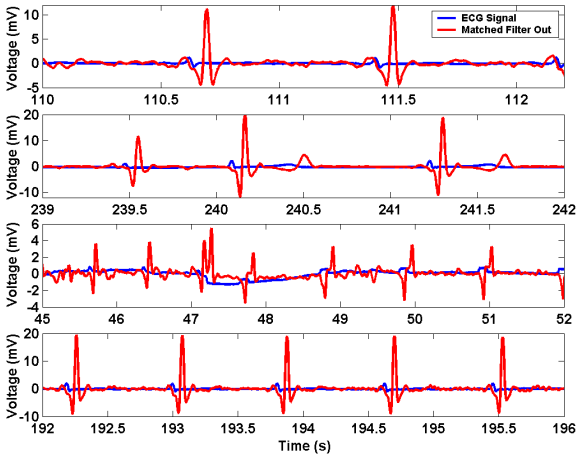


Fig. 6. The QRS Complex Matched Filter Output for the Normal, Arrhythmia, Cardiac Arrest, and Long Term ST ECGs.

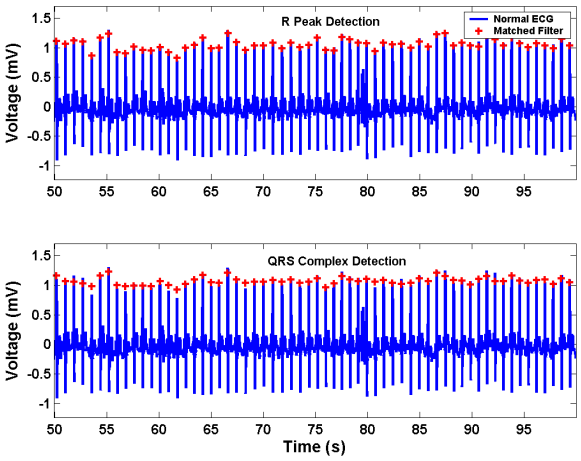


Fig. 7. The R-wave Peak and QRS Complex Matched Filter outputs for the Normal ECG after feature detection.

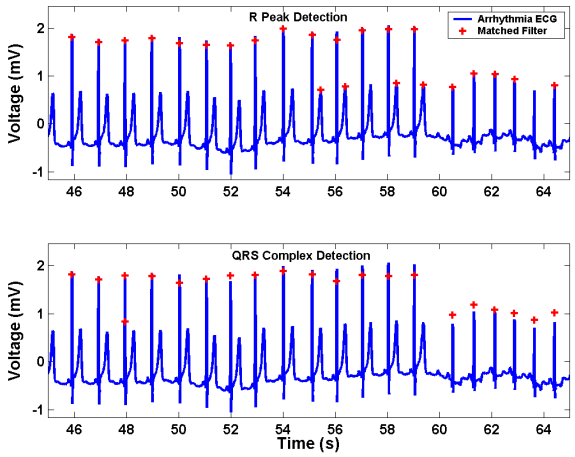


Fig. 8. The R-wave Peak and QRS Complex Matched Filter outputs for the Arrhythmia ECG after feature detection.

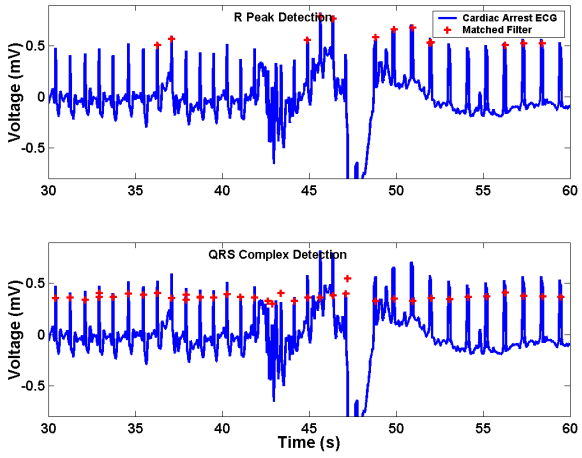


Fig. 9. The R Peak and QRS Complex Matched Filter outputs for the Cardiac Arrest ECG after feature detection.

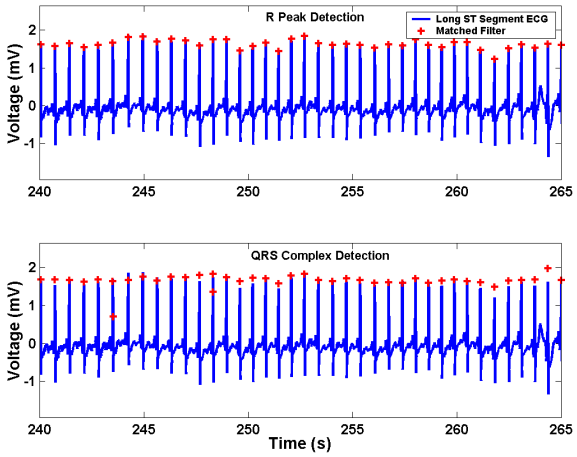


Fig. 10. The R Peak and QRS Complex Matched Filter outputs for the Long Term ST ECG after feature detection.

Unfortunately, this was not the case with the Arrhythmia and Cardiac Arrest ECGs when using R-wave peak detection. The Arrhythmia ECG has areas where the ST segment was highly elevated and the matched filter output considered these peaks as R-wave peaks. The problem was due to the detection threshold used, not the matched filter implementation. The peak detection threshold was two-thirds of the average value of the output of the matched filter for the specific feature of interest.

The peak detection resolution was not sensitive enough to distinguish between the R-wave peaks and the ST segment peaks due to the use of a feature-length, sliding detection window . When the ST segment peaks were not present, the R-wave peaks were detected. Good R-wave peak detection is necessary in order to determine the heart rate variability. In order to obtain good R-wave peak detection, the ECG signal should not have too much baseline wander, amplitude changes or noise which can cause poor detection [11] and both of these ECG signals had more noise and extraneous peaks than the other two ECG signals.

The use of QRS complex detection had much better results. Only one heart beat was not detected and incorrectly detected beats (false positives) were much lower than with R-wave peak detection. The peak detection threshold for the QRS complex was higher in level (9 mV vs. 6 mV), so it was not as sensitive to the elevated ST segment peaks.

The Cardiac Arrest ECG had a different detection problem. There were areas within the ECG signal where the ECG signal was lost, noisy or chaotic due to cardiac arrest. The preceding R-wave peaks before the arrest section were low in voltage level in relation to the rest of the signal so the matched filter threshold algorithm had a difficult time distinguishing the R-wave peaks. When the patient was receiving cardiopulmonary resuscitation, the ECG displayed this as a pulse, similar to a square wave and surprisingly, the R-wave peak matched filter did not detect these peaks as R-wave peaks due to their square wave shape.

The QRS complex matched filter detection with the Cardiac

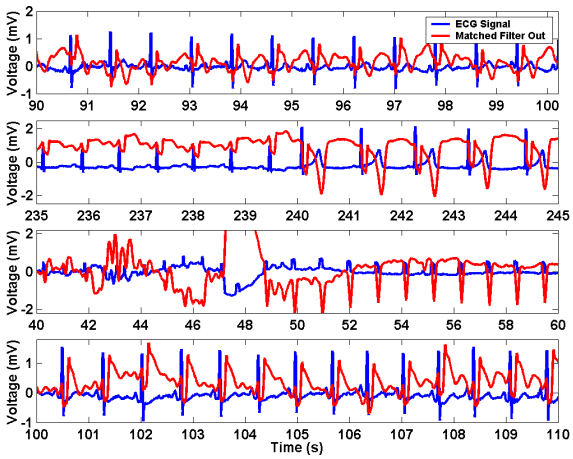


Fig. 11. The ST Segment Matched Filter Output for the Normal, Arrhythmia, Cardiac Arrest, and Long Term ST ECGs.

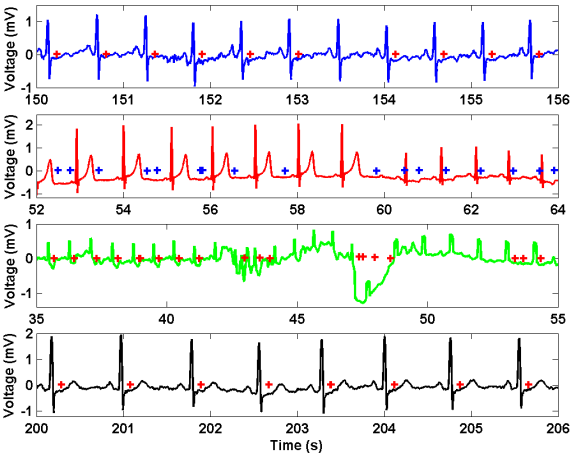


Fig. 12. The ST Segment Matched Filter output for the Normal, Arrhythmia, Cardiac Arrest, and Long Term ST ECGs after feature detection.

Arrest ECG was much better in detection, especially in the noisy areas where the R-wave peak matched filter failed. Unfortunately, the QRS complex matched filter detected the chest compressions as QRS complexes even though they were not, but the detection of false positives was lower than with R-wave peak detection.

The ST segment was much harder to detect using the ST segment matched filter and the results while good with the Normal and Long Term ST ECGs, were poor for the Arrhythmia and Cardiac Arrest ECGs. The ST segment matched filter results before and after detection are shown in Fig. 11 and Fig. 12.

The ST segment matched filter for the Normal ECG had problems because of the noise in the ST segment. Many ST segments were not detected and the incident of false positives was high. Nonetheless, out of 371 heart beats, the matched filter detected 337 ST segments.

The matched filter output for the Arrhythmia ECG was worse. In the second plot segment shown on Fig. 12, while

TABLE I
R-WAVE PEAK, QRS COMPLEX, AND ST SEGMENT DETECTION RESULTS

| R-Wave Peak | Found | Miss | FP | Sensitivity | 1-Specificity |
|---|---|---|---|---|---|
| Normal | 370 | 1 | 0 | 0.99 | 0 |
| Arrhythmia | 329 | 22 | 44 | 0.94 | 0.13 |
| Cardiac Arrest | 146 | - | 97 | - | - |
| Long Term ST | 423 | 3 | 15 | 0.99 | 0.04 |
| **QRS Complex** | **Found** | **Miss** | **FP** | **Sensitivity** | **1-Specificity** |
| Normal | 370 | 1 | 3 | 0.99 | 0.01 |
| Arrhythmia | 350 | 1 | 8 | 0.99 | 0.02 |
| Cardiac Arrest | 306 | - | 132 | - | - |
| Long Term ST | 425 | 1 | 17 | 0.99 | 0.04 |
| **ST Segment** | **Found** | **Miss** | **FP** | **Sensitivity** | **1-Specificity** |
| Normal | 337 | 34 | 54 | 0.91 | 0.15 |
| Arrhythmia | 209 | 142 | 256 | 0.60 | 0.73 |
| Cardiac Arrest | 205 | - | 133 | - | - |
| Long Term ST | 400 | 26 | 9 | 0.94 | 0.07 |

the elevated ST segments triggered no ST segment detection unfortunately, only false positives were detected. When the ST segments disappeared after 60 seconds, the matched filter began to detect the ST segments but incorrect detections continued. The ST segment detection for this ECG signal was poor.

The ST segment detection for the Cardiac Arrest ECG was better but only for areas where the ECG signal was consistent and with very little noise. Since many areas of the Cardiac Arrest ECG contained noise due to cardiac problems, detection of non-peak features such as the ST segment was difficult so accurate metrics could not be determined. All of the detection results are shown in Table I. *(Note: The abbreviation FP stands for False Positive.)*

The last detection section involved the heart rate variability. The use of the matched filter to detect the R-R intervals made this possible. Once the R-wave peaks were detected, the R-R intervals were then calculated. In Fig. 13, all four of the ECG heart rate variabilities (HRV) are plotted. The Normal and Long Term ST ECGs have similar HRVs, but the Arrhythmia and Cardiac Arrest ECGs are different with periods of non-variability which is not normal.

According to Laguna *et al.* [12], in order to obtain the power spectral density (PSD) of the HRV series, the HRV series must be resampled in order to provide an evenly sampled signal even though possible artifacts can be introduced by the low-pass filtering. In order to get a unique representation of the HRV signal, it must be band-limited to the inverse of the R-R interval mean. In this case, the R-R interval mean was less than 2 Hz so the resampling rate was set to 2 Hz.

The PSDs of the HRVs showed that the power was concentrated at less than 0.1 Hz as shown in Fig. 14. The peaks at 0.1 Hz correspond to the Low Frequency (LF) component of the heart rate variability. Due to the small number of data points, fewer than 500, in the HRV signal, each ECG's HRV was zero-padded to the same 1024 sample points length before the PSD was determined.

## IV. DISCUSSION

The detection of different features in the ECG waveform was much harder than anticipated but it was not due to the

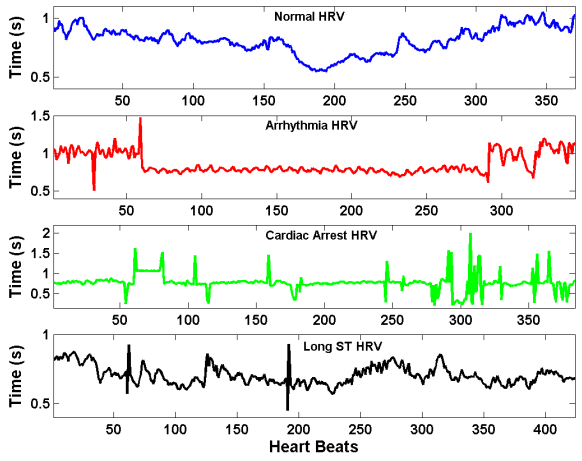


Fig. 13. The HRV Tachograms for the Normal, Arrhythmia, Cardiac Arrest, and Long Term ST ECGs.

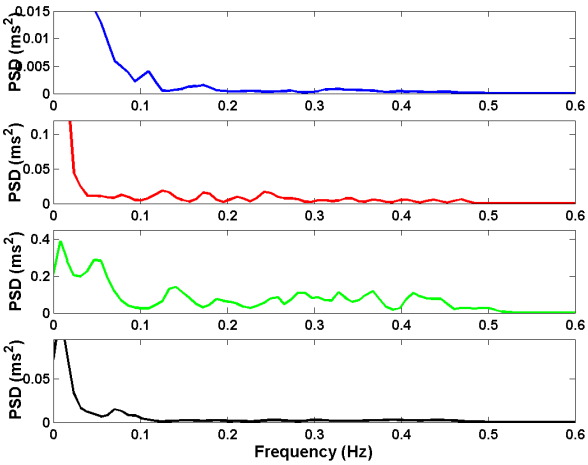


Fig. 14. The PSDs of the Normal, Arrhythmia, Cardiac Arrest, and Long Term ST ECG HRVs.

implementation of the matched filter. The more difficult part was creating the detection method to extract the feature of interest in each ECG signal. The Normal and Long Term ST ECG signals were somewhat similar so their respective detection thresholds were not as affected by extraneous peaks as the Arrhythmia and Cardiac Arrest ECGs signals. Unfortunately, the matched filter was sensitive to the presence of extraneous or multiple peaks and had difficulty extracting the signal of interest when its duration was short. When the matched filter signal template had short, sharp peaks to detect, such as the R-wave peak, the detection was not as good as with a larger, less *sharp* template such as the QRS complex.

Detecting the level and slope of the ECG ST segment proved much harder than anticipated because while the ST segment could usually be found, the slope information could not be extracted with the matched filter alone. The ST segment varies from beat to beat and it is highly susceptible to noise which makes slope detection difficult without some type of multiple threshold detection that takes into account the beginning and end points of the ST segment slope. This makes the problem difficult and requires multiple detection passes.

## V. Conclusion

This project was difficult but the results were better than expected. The variability of the human heart beat was unexpected and detecting the features of interest was made harder by this fact. Surprisingly, the matched filter itself was not difficult to implement or use. What was difficult was the detection threshold on the signal features of interest.

A better method for QRS complex detection along with R-wave peak and ST segment detection would involve using an adaptive filter to whiten the noise in the QRS complex [13]. By improving on the methods used; using a different form of the matched filter and better threshold detection, the matched filter ECG feature extraction could be made more successful.

## References

[1] L. A. Geddes, "Retrospectroscope: Contributions of the Vacuum Tube to Early Electrophysiological Research," *IEEE Engineering in Medicine and Biology*, pp. 118–126, January/February 2001.

[2] L. A. Geddes and A. Wald, "Retrospectroscope: Horatio B. Williams and the First Electrocardiographs Made in the United States," *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, no. 5, pp. 117–121, September/October 2000.

[3] H. Gholam-Hosseini and H. Nazeran, "Detection and Extraction of the ECG Signal Parameters," *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, no. 1, pp. 127–130, 1998.

[4] D. G. Manolakis, V. K. Ingle, and S. M. Kogan, *Statistical and Adaptive Signal Processing*. Artech House, 2005.

[5] P. Cugini, F. Bernardini, C. Cammarota, L. Cedrone, D. Cipriani, S. Coda, M. Curione, C. Danese, G. D. Francesco, R. D. Rosa, S. Fontana, A. Pellegrino, and E. Proietti, "Disorder and Circadian Periodicity In Within-Day Variability of Sinusal R-R Intervals In Myocardial Infarcted Patients," *Journal of Clinical and Basic Cardiology*, vol. 3, no. 1, pp. 53–58, 2000.

[6] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13). [Online]. Available: http://circ.ahajournals.org/cgi/content/full/101/23/e215

[7] [Online]. Available: http://www.physionet.org/physiobank/database/mitdb/

[8] S. M. Szilágyi, "Event Recognition, Separation and Classification From ECG Recordings," *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, no. 1, pp. 236–239, 1998.

[9] M. Malik, J. T. Bigger, R. E. K. A. John Camm, A. Malliani, A. J. Moss, and P. J. Schwartz, "Heart Rate Variability, Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, pp. 354–381, 1996.

[10] T. Srikanth, S. Napper, and H. Gu, "Bottom-Up Approach to Uniform Feature Extraction in Time and Frequency Domains for Single Lead ECG Signal," *Journal of the International Society for Biolectromagnetism*, vol. 4, no. 1, 2002.

[11] R. Logier, J. Dejonckheere, and A. Dassonneville, "An Efficient Algorithm For R-R Intervals Series Filtering," *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, pp. 3937–3940, September 2004.

[12] P. Laguna, G. B. Moody, and R. G. Mark, "Power Spectral Density of Unevenly Sampled Data by Least-Square Analysis: Performance and Application to Heart Rate Signals," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 6, pp. 698–715, June 1998.

[13] P. S. Hamilton and W. J. Tompkins, "Adaptive Matched Filtering for QRS Detection," *IEEE Engineering in Medicine and Biology Society 10th Annual International Conference*, pp. 0147–0148, 1988.