Pre-Processing and Compression of ECG Records for Biotelemetry Purposes

Neville Aquinas [1]
[1]Department of E & C Engineering
St Joseph Engineering College,
Mangaluru, India

Ramya P C [2]
[2]Department of E & C Engineering
St Joseph Engineering College,
Mangaluru, India

Abstract— **Cardiovascular heart diseases are one of the leading causes of deaths in India and throughout the world. A majority of population suffers from these cardiovascular diseases due to inadequate access to healthcare facilities and non availability of healthcare personnel. Advanced technologies like Telemedicine, biotelemetry have been developed to eliminate the barriers of distance and topology. These technologies when used for detection of cardiovascular heart disease using ECG have limited operational bandwidth and face archival problems. This work is based on efficiently using the bandwidth and solving the archival problem by preprocessing the signal and ECG compression. ECG data compression algorithm will reduce the amount of data to be transmitted, stored and analyzed, without losing the clinical information. Direct compression techniques are evaluated to find an optimal compression strategy for ECG data compression.**

Keywords—ECG, ECG Preprocessing, ECG Data Compression, Telemedicine, Biotelemetry

I. INTRODUCTION

Cardiovascular heart diseases are one of the leading causes of deaths in India and throughout the world. To help people suffering from these diseases, research is being carried out in medical technology to find ways to improve detection of such diseases. By early detection of the symptoms of these diseases, mortality and morbidity of the patients can be decreased significantly. A majority of population in India and many parts of the world suffers from these cardiovascular diseases due to inadequate access to healthcare facilities and non availability of healthcare personnel. To overcome these problems technologies like Telemedicine, biotelemetry have been developed to eliminate the barriers of distance and topology. Biomedical telemetry or Biotelemetry is a special field of biomedical instrumentation that often enables transmission of biological information from an inaccessible location to a remote monitoring site. Thus making use of these technologies people from remote places can have access to healthcare personnel and healthcare facilities.

Electrocardiogram (ECG) is one of the most practiced methods to detect any abnormalities in heart function. Thus technologies like Telemedicine, biotelemetry are used to transmit ECG signal for detection of cardiovascular heart disease. While using these technologies for transmission of ECG, few problems come into picture like limited operational bandwidth and archival of received ECG signal for further analysis.

This work is based on efficiently using the bandwidth while transmitting ECG signal and solving the archival problem of received ECG signal for further analysis. The problem of limited operational bandwidth in biotelemetry and telemedicine and archival can be solved by preprocessing the signal before transmission, thus removing unwanted signal from the ECG signal and ECG data compression. ECG data compression algorithm is needed since it will reduce the amount of data to be transmitted, stored and analyzed, but without losing the clinical information content [4]. ECG data compression allows real-time transmission over telephone networks, economic off-line transmission to remote interpretation sites [8]. Various Direct compression techniques are evaluated to find an optimal compression strategy for ECG data compression.

The electrocardiogram (ECG or EKG) is a diagnostic tool that measures and records the electrical activity of the heart. The ECG is a graphic record of the direction and magnitude of the electrical activity that is generated by polarization and depolarization of the atria and ventricles. One cardiac cycle in an ECG signal consists of the P-QRS-T waves. Most of the clinically useful information in the ECG is found in the intervals and amplitudes defined by its features (characteristic wave peaks and time durations).

There are 3 main deflections in an ECG: the P-wave, the QRS complex, and the T-wave. These are shown in Figure 1 below
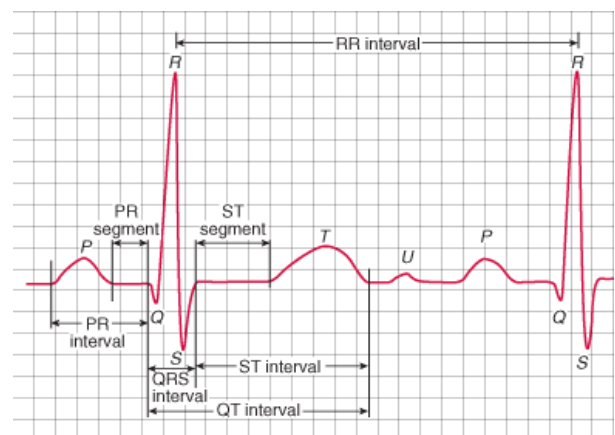

Figure 1: A typical ECG tracing [Courtesy: en.wikipedia.org]

Electrical sensing devices or electrodes are placed strategically on top of the body to detect the electrical activity of the heart and diagnose patients with different heart

anomalies. The trace depends on the position of the lead. An electrocardiogram is obtained by measuring electrical potential between various points of the body using a biomedical instrumentation amplifier. The electrodes usually consist of a conducting gel, embedded in the middle of a self-adhesive pad onto which cables clip. Sometimes the gel also forms the adhesive [9].

II. METHODOLOGY

A. Loading Ecg Signal

The ECG signals are obtained from the PTB Diagnostic ECG Database [5]. Each record includes 15 simultaneously measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) together with the 3 Frank lead ECGs (v_x, v_y, v_z). Each signal is digitized at 1000 samples per second. Each of these patient records consists of 2 files: Header File and Data File. Header file contains patient data. Data file contains ECG data and is stored as .mat [5] which can be read by loading it into Matlab. The 1st signal (i) from the 12 lead is used in this work. Figure 2 shows the ECG signal loaded from PTB database.
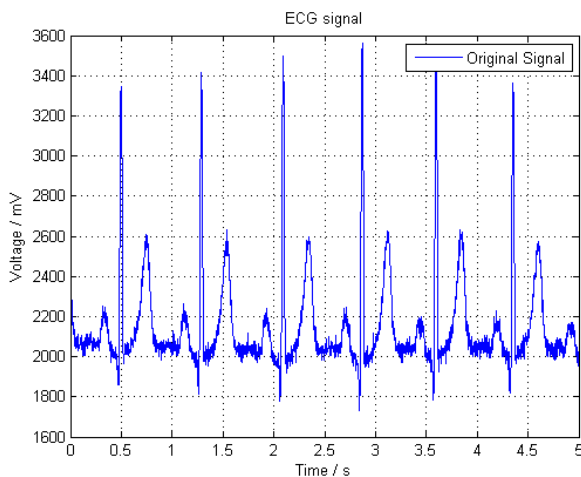

Figure 2: Snapshot of ECG signal loaded from PTB database

B. Ecg Signal Preprocessing

ECG recordings are often corrupted by noise artifacts. The two dominant noise artifacts present in ECG recordings [10] are: 1).High-frequency noise caused by electromyogram induced noise, power line interferences, or mechanical forces acting on the electrodes; 2).Baseline wander (BW) that may be due to respiration or the motion of the patients or the instruments. These noise artifacts severely limit the efficiency of the process and thus need to be removed for better computational results. Hence these recordings are subjected to pre-processing prior to the main processing stage.

Wavelet transform is used to eliminate these noise artifacts. A discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. The reason behind using discrete wavelet transform is ECG signal can be decomposed without any loss of information and energy [6]. Wavelet decomposition on raw ECG signal

results in approximate coefficients (CA), and detailed coefficients (CD). L_i represents length of first approximate coefficient and all detailed coefficients. The sampling frequency of the ECG signal from PTB database is known i.e. 1000Hz [5]. By substituting in Nyquist Criterion, the maximum frequency of the ECG signal is calculated to be 500Hz. The levels of decomposition are restricted to 10, since the 10th level decomposition gives an approximate coefficient of 0-0.48; this frequency range contains the baseline wandering frequencies (0-0.48) [5]. The important features of the ECG signal are within 60Hz and the baseline wandering frequencies are within 0-0.48Hz [3]. Hence co-efficients containing these frequencies are forced to zero and then the coefficients are reconstructed back to obtain filtered signal. Figure 3 shows the preprocessed signal.
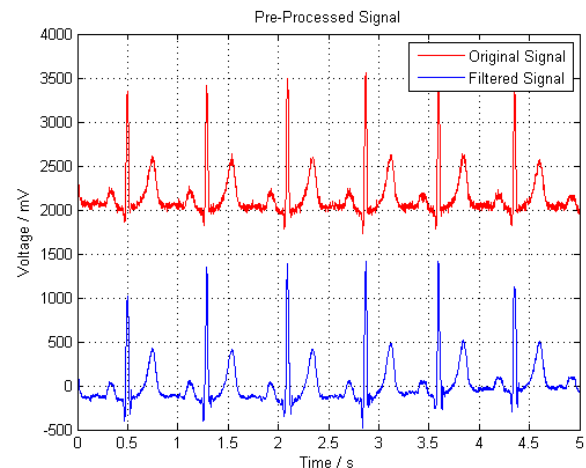

Figure 3: Snapshot of preprocessed ECG signal

C. Qrs Complex Detection

Detection of QRS complex is required in order to obtain the R-peak and its location from the pre-processed signal. QRS complex is detected using modified Pan Tompkins's algorithm [2]. Typical frequency components of a QRS complex range from about 5 Hz to about 15 Hz [2][3]. Therefore, QRS detection algorithm uses a filter stage prior to the actual detection in order to attenuate other signal components and artifacts, such as P-wave, T-wave [3].

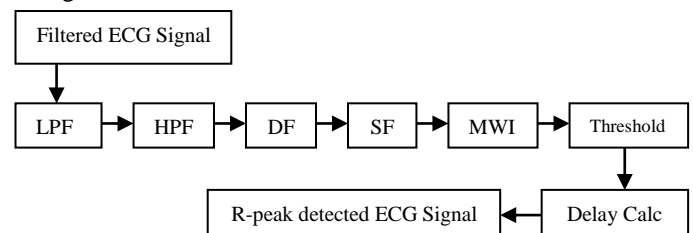Block diagram of QRS Complex Detection is shown in Figure 4,

Figure 4: Block Diagram of QRS Complex Detection

The filtered signal is passed through a second order Lowpass filter (LPF) with cut-off frequency of about 11Hz. The Lowpass filtered signal is then passed through High pass filter (HPF) with cut-off frequency of about 5Hz. Filtering the ECG signal with Lowpass filter and Highpass filter gives a

combined effect of bandpass filter. The bandpassed signal is then passed through derivative filter (DF) to calculate the slope of the signal. The derivated signal is squared (SF) point by point. Squaring makes all data points positive. The squared signal is then passed through moving average filter to obtain waveform feature information in addition to the slope of the R wave [2]. The signal is then thresholded for detecting the R-peak. The purpose of the thresholding is to obtain maximal slope or the peak of the R wave. This is done by selecting the samples that satisfy the thresholding condition.

Max.Peak = [(Max.Peak-2) < (Max.Peak-1) < (Max.Peak) > (Max.Peak+1)> (Max.Peak+2)] && Max.Peak > Threshold Level (i.e. 0.25*Max.Amplitude)
(1)

The R-peak detected will have a delay due to the filtering process carried out during earlier stages. This group delay is calculated manually by overlaying the original signal and the filtered signal and the delay is compensated by shifting the located R-peak by the delay calculated. Finally the detected R-peak is marked on the signal. Figure 5 shows the R-peak detected in the ECG Signal.
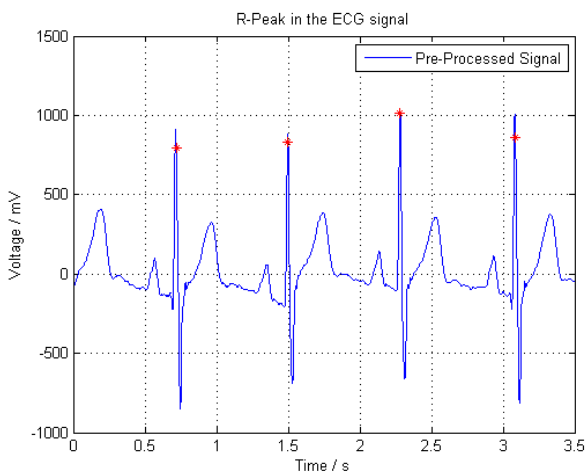

Figure 5: Snapshot of R-peak located signal

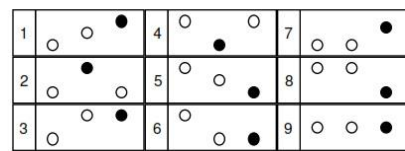### D. Ecg Data Compression

The practical importance of ECG data compression has become evident in many aspects of Computerized electrocardiography [8] including: a) increased storage capacity of ECGs as databases for subsequent comparison or evaluation , b)Feasibility of transmitting real-time ECGs over the public phone network, c)Implementation of cost effective real-time rhythm algorithms, d)Economical rapid transmission of off-line ECG"s over public lines to a remote interpretation center, and e) Improved functionality of ambulatory ECG monitors and recorders. The ECG data is compressed with a view to storing it and later on decompressing it without any loss of diagnostic information. Biomedical signals can be broadly classified into: direct data compression, transformation methods, and parameter extraction method. In this work direct data compression techniques like turning point (TP) [11], amplitude- zone-time epoch coding (AZTEC) [12], and Fan algorithms have been analyzed [13].

1. Turning Point Algorithm:

The algorithm [1] processes three data points at a time. It stores the first sample point and assigns it as the reference point X_0. The next two consecutive points become X_1 and X_2. The algorithm retains either X_1 or X_2, depending on which point preserves the turning point (i.e., slope change) of the original signal. Figure 6(a) shows all the possible configurations of three consecutive sample points. In each frame, the solid point preserves the slope of the original three points. The algorithm saves this point and makes it the reference point X_0 for the next iteration. It then samples the next two points, assigns them to X_1 and X_2, and repeats the process. A simple mathematical criterion is used to determine the saved point. First consider a sign(x) operation

$$sign(x) = \begin{cases} 0 & x = 0 \\ +1 & x > 0 \\ -1 & x < 0 \end{cases}$$

(2)

(a)

Pattern	$s_1 = sign(X_1 - X_0)$	$s_2 = sign(X_2 - X_1)$	NOT(s_1) OR ($s_1 + s_2$)	Saved sample
1	+1	+1	1	X_2
2	+1	−1	0	X_1
3	+1	0	1	X_2
4	−1	+1	0	X_1
5	−1	−1	1	X_2
6	−1	0	1	X_2
7	0	+1	1	X_2
8	0	−1	1	X_2
9	0	0	1	X_2

(b)

Figure 6: Turning point (TP) algorithm. (a) All possible 3-point configurations. Each frame includes the sequence of three points X_0, X_1, and X_2. The solid points are saved. (b) Mathematical criterion used to determine saved point.

Using the mathematical formula, the following are then obtained s1 = sign(X_1 − X_0) and s2 = sign(X_2 − X_1), where (X_1 − X_0) and (X_2− X_1) are the slopes of the two pairs of consecutive points. If a slope is zero, this operator produces a zero result. For positive or negative slopes, it yields +1 or −1 respectively. A turning point occurs only when a slope changes from positive to negative or vice versa.

Figure 7 shows the snapshot of ECG signal before compression using Turning Point algorithm and after reconstruction.
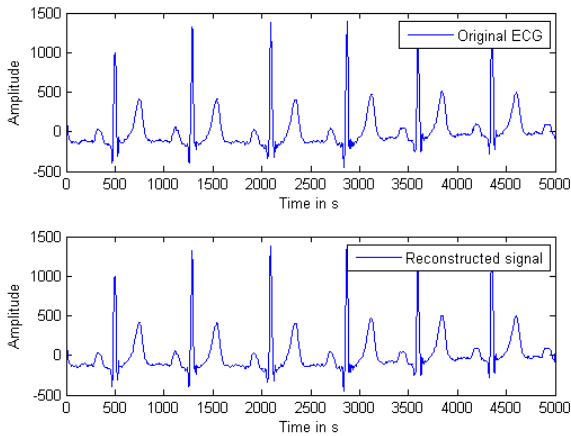
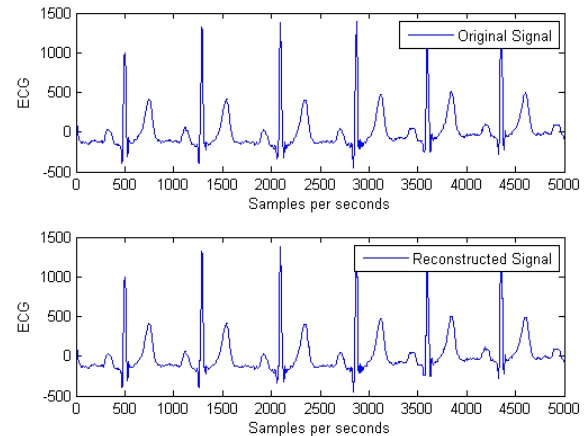Figure 7: Snapshot of ECG signal before compression using Turning point algorithm and after reconstruction.

Figure 8: Snapshot of ECG signal before compression using FAN algorithm and after reconstruction.

2. Aztec Algorithm:

AZTEC (Amplitude Zone Time Epoch Coding) [1] converts the ECG waveform into plateaus (flat line segments) and sloping lines. AZTEC retains only the samples for which there is sufficient amplitude change. AZTEC Algorithm is implemented in 2 phases:

Horizontal Mode:

1) Acquire the ECG signal
2) Assign the first sample to X_{max} and X_{min} which represents highest and lowest elevations of the current line.
3) Check for the following condition and store the plateau if
   a. If $X1 > X_{max}$ then $X_{max} = X1$ and
   b. If $X1 < X_{min}$ then $X_{min} = X1$ and so on till X_n samples, repeat this until the following 2 conditions are satisfied
      i. the difference between V_{MAX} and V_{MIN} is greater than a predetermined threshold or
      ii. If line length is > 50 are satisfied
4) The stored values are the length L=S-1, where S is no. of samples and L is length and the average amplitude of the plateau $(V_{MAX}+V_{MIN})/2$.
5) Algorithm starts assigning the next samples to X_{max} and X_{min}.

Slope Mode:

1) If no. of samples ≤ 3, then the line parameters are not saved. Instead the algorithm begins to produce slopes.
2) The direction of the slope is determined by checking the following conditions.
   a) If $(X_2 - X_1) * (X_1 - X_0)$ is +ve then the slope is +ve.
   b) If $(X_2 - X_1) * (X_1 - X_0)$ is -ve then the slope is -ve.
3) The slope is terminated if the no. of samples is ≥ 3 and if direction of slope is changed.

Figure 8 shows the snapshot of ECG signal before compression using FAN algorithm and after reconstruction.

3. FAN Algorithm:

Originally used for ECG telemetry, the Fan algorithm [1] draws lines between pairs of starting and ending points so that all intermediate samples are within some specified error tolerance, Figure 9 Illustrates the principles of the Fan algorithm.
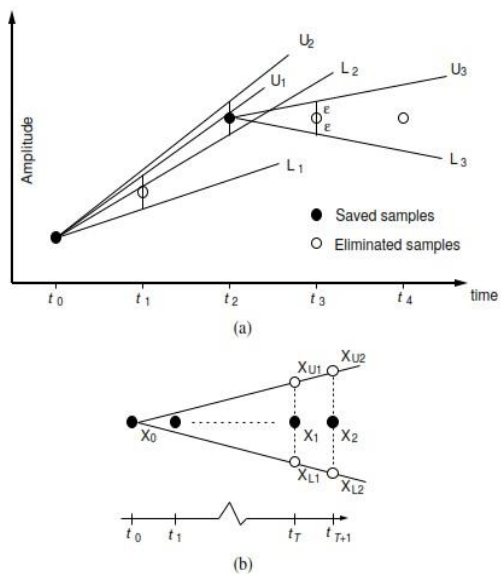

Figure 9. Illustration of the Fan algorithm. (a) Upper and lower slopes (U and L) are drawn within error threshold around sample points taken at t_1, t_2, … (b) Extrapolation of X_{U2} and X_{L2} from X_{U1}, X_{L1}, and X_0.

The algorithm starts by accepting the first sample X_0 as the non redundant permanent point. It functions as the origin and is also called the originating point. It then takes the second sample X_1 and draw two slopes $\{U_1, L_1\}$. U_1 passes through the point $(X_0, X_1+\varepsilon)$, and L_1 passes through the point $(X0, X1-\varepsilon)$. If the third sample X_2 falls within the area bounded by the two slopes, it generates two new slopes $\{U_2, L_2+\varepsilon\}$ that pass through points (X_0, X_2) and $(X_0, X_2 - \varepsilon)$. It compares the two pairs of slopes and retains the most converging (restrictive) slopes (i.e., $\{U_1, L_2\}$).

Next it assigns the value of X_2 to X_1 and read the next sample into X_2. As a result, X_2 always holds the most

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICESMART-2015 Conference Proceedings

recent sample and X_1 holds the sample immediately preceding X_2. It repeats the process by comparing X_2 to the values of the most convergent slopes. If it falls outside this area, it saves the length of the line T and its final amplitude X_1 which then becomes the new originating point X_0, and the process begins anew. The sketch of the slopes drawn from the originating sample to future samples forms a set of radial lines similar to a fan, giving this algorithm its name.

From Figure 9(b), we can see that

$$x_{U2} = \frac{x_{U1} - x_0}{T} + x_{U1} \tag{3}$$

$$x_{L2} = \frac{x_{L1} - x_0}{T} + x_{L1} \tag{4}$$

where, $T = tT - t0$. It reconstructs the compressed data by expanding the lines into discrete points. The Fan algorithm guarantees that the error between the line joining any two permanent sample points and any actual (redundant) sample along the line is less than or equal to the magnitude of the preset error tolerance. The algorithm's reduction ratio depends on the error tolerance.

Figure 10 shows the snapshot of ECG signal before compression using AZTEC algorithm and after reconstruction.
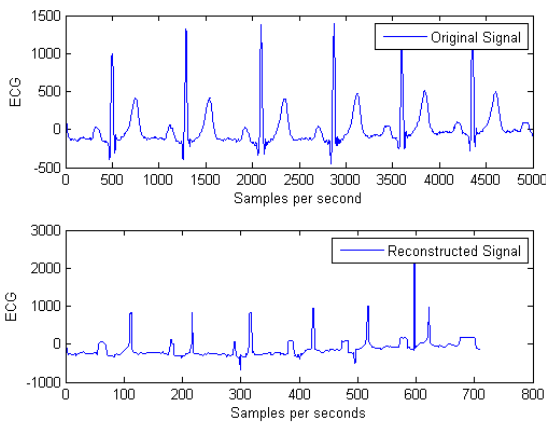

Figure 10: Snapshot of ECG signal before compression using AZTEC algorithm and after reconstruction.

The effectiveness of an ECG compression technique is described in terms of:

i) Compression Ratio (CR): CR is the ratio of the original data to compressed data without taking into account factors such as bandwidth, sampling frequency, precision of the original data, word length of compression parameters, reconstruction error threshold, database size, lead selection, and noise level. It is given by:

$$CR = \frac{OriginalFileSize}{CompressedFileSize} \tag{5}$$

Higher the CR, smaller the size of the compressed file.

ii) Percentage Mean Square Difference (PRD): is a measure of error loss. This measure evaluates the distortion between the original and the reconstructed signal. PRD calculation is as follows:

$$PRD = 100 * \sqrt{\frac{\sum_{i=1}^{n}(OriginalSignal(i) - Re\,constructedSignal(i))^2}{\sum_{i=1}^{n}(OriginalSignal(i))^2}} \tag{6}$$

The Table 1 below shows the comparison between compression techniques of turning point (TP), amplitude-zone-time epoch coding (AZTEC), and Fan algorithms.

Table 1. Comparison of Compression Techniques

ECG Signal	Turning Point Algorithm	
	Compression Ratio	PRD
s0486	2	0
s0503	2	0
s0506	2	0
s0487	2	0
s0531	2	0
s0499	2	0
s0500	2	0
s0502	2	0
s0551	2	0
s0532	2	0
ECG Signal	AZTEC Algorithm	
	Compression Ratio	PRD
s0486	1.1652	85.8200
s0503	1.2168	82.2000
s0506	1.2416	80.5400
s0487	1.5142	66.0400
s0531	1.6898	59.2000
s0499	1.3298	75.2400
s0500	1.3889	72.4400
s0502	1.6551	61
s0551	1.4405	69.4400
s0532	1.6938	59.0400
ECG Signal	FAN Algorithm	
	Compression Ratio	PRD
s0486	1.9365	0.1400
s0503	1.9794	0.1400
s0506	2.0186	0.1200
s0487	2.0137	0.1200
s0531	1.9670	0.1400
s0499	2.0416	0.1200
s0500	2.0048	0.1600
s0502	2.0194	0.1200
s0551	2.0392	0.1400
s0532	2.0129	0.1400

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICESMART-2015 Conference Proceedings

III. DISCUSSION

Preprocessing the ECG signal using discreet wavelet transform removes most of the noise artifacts present in the ECG signal. Comparison of direct ECG data compression algorithms show that the Turning Point algorithm produces better signal fidelity and better reduction ratio which can be used for real-time transmission.

IV. CONCLUSION

The problem of limited operational bandwidth in biotelemetry and telemedicine can be solved by preprocessing the signal before transmission; this can be clearly seen by comparing the ECG signal before preprocessing and ECG signal after Preprocessing. ECG data compression algorithm can solve the archival problem by reducing the amount of data to be stored and analyzed, but without losing the clinical information content. In the future work, analyzing transformation methods and feature extraction methods for efficient storage of ECG will be carried out. Moreover additional statistical data need to be utilized for evaluating the performance of an algorithm in ECG compression. Thus comparison of best data compression technique can be analyzed.

REFERENCES

[1] Willis J. Tompkins, Biomedical Digital Signal Processing, Prentice-Hall, 1993.

[2] Jiapu Pan, Senior Member, IEEE and Willis J Tompkins, Senior Member, IEEE, "A Real – Time QRS Detection Algorithm", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. BME-23, NO. 3, March 1985

[3] Bert-UweKöhler, CarstenHennig, Reinhold Orglmeister, "The Principles of Software QRS Detection", IEEE ENGINEERING IN MEDICINE AND BIOLOGY, January/February 2002.

[4] Hierarchical Neural Network Based Compression of ECG signals - Bekir Karlik, P.M.A. Sloot et al. (Eds.): ICCS 2003, LNCS 2657, pp. 371-377,2003.

[5] "PTB Database", [Online]
Available: http://physionet.org/physiobank/database/ptb/, [Accessed: Mar. 2, 2014]

[6] Md. Ashfanoor Kabir, Celia Shahnaz, "Denoise of ECG signals based on noise reduction algorithms in EMD and Wavelet domains", Biomedical Signal Processing and Control, Vol. 7, pg. 481-489, 2012.

[7] Min Dai, Shi-Liu Lian, "Removal of Baseline Wander from Dynamic Electrocardiogram Signals", Image and Signal Processing, 2009. CISP '09. 2nd International Congress on, vol., no., pp.1, 4, 17-19 Oct. 2009

[8] ECG Data Compression Using DWT , Anubhuti Khare, Manish Saxena, Vijay B. Nerkar , International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-1, October 2011

[9] "Electrocardiography", [Online]
Available: https://en.wikipedia.org/wiki/Electrocardiography, [Accessed: Mar. 2, 2014]

[10] Blanco-Velasco, M.; Weng, B.; Barner, K.E., "A New ECG enhancement algorithm for stress ECG tests," Computers in Cardiology, 2006 , vol., no., pp.917,920, 17-20 Sept. 2006

[11] W. C. Mueller, ―Arrhythmia detection program for an ambulatory ECG monitor,‖ Biomed. Sci. Instrument., vol. 14, pp. 81-85, 1978.

[12] J. R. Cox, F. M. Nolle, H. A. Fozzard, and G. C. Oliver, ―AZTEC, a preprocessing program for real-time ECG rhythm analysis,‖ IEEE Trans. Biomed. Eng., vol. BME-15, pp. 128-129, Apr. 1968.

[13]. J . Sklansky and V. Gonalez, ―Fast polygonal approximation of digitized curves,‖ Pattern Recog., vol. 12, pp. 327-331, 1980.