International Journal of Electrical, Electronics and Computer Engineering **1**(2): 77-82(2012)
ISSN No. (Online) : 2277-2626
Special Edition for Best Papers of Michael Faraday IET India Summit-2012, MFIIS-12

ECG Compression Technique Using ASCII Character Encoding and Transmission Using GSM Transmitter

S.K. Mukhopadhyay*, M. Mitra**, and S. Mitra***

**Department of Applied Physics, University of Calcutta, 92 A.P.C. Road, Kolkata (WB)*
***Department of Applied Physics, Faculty of Technology, University of Calcutta, 92 A.P.C. Road, Kolkata, (WB)*
****Department of Electronics, Netaji Nagar Day College (affiliated to University of Calcutta), 170/436, N.S.C. Bose Road, Regent Estate, Kolkata, (WB)*

(Received 05 October, 2012 Accepted 01 December, 2012)

ABSTRACT: Software based new, efficient and reliable ECG data compression and transmission scheme is proposed here. It is observed that this proposed algorithm can reduce the file size significantly (CR=15.72) without altering the significant clinical information (PRD=7.89%). The compression scheme is such that the compressed file contains only ASCII characters. These characters are transmitted using Global System for Mobile Communication (GSM) based Short Message Service (SMS) system and at the receiving end original ECG signal is brought back using just the reverse logic of compression. The compression and reconstruction algorithms are implemented on C-platform and SMS generation and collection algorithms are implemented on MATLAB 7.1 environment.

***Index Terms*—** Sign Byte, Grouping, ASCII Character, GSM Transmitter, SMS.

I. INTRODUCTION

Electrocardiogram (ECG) describes different electrical phases of a cardiac cycle and represents a summation in time and space of the action potentials generated by cardiac cells. It also provides a measure of the electrical currents initiated in the extra-cellular fluid because of the potential changes across the cell membrane [1]. A typical normal ECG trace is shown in Fig. 1.

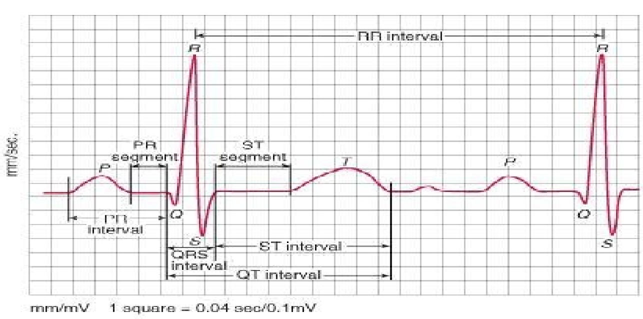

Fig.1. A typical normal ECG trace.

The ECG is described by waves, segments and intervals. Waves are tagged using the letters P, QRS, T. Segments are time spans between waves and intervals are time lengths that include waves and segments. The shape and size of the P-QRS-T wave and the time intervals between various peaks contain useful information about the nature of possible disease afflicting a heart. P, Q, R, S and T letters were chosen in the early days of ECG history and were chosen arbitrarily [2].

However, by the very nature of bio-signals, reflection of abnormalities would be random in the timescale. Hence the study of ECG pattern and heart rate variability have to be carried out for extended periods of time (i.e., for 24 h) as done in holter monitoring system. Naturally the volume of the data handled would be enormous. Therefore we must need a way to reduce the data storage place. ECG signal compression techniques can be broadly classified into three major categories: (i) direct data compression, (ii) transformation methods and (iii) parameter extraction techniques. Direct data compression techniques ([3]) generally retain samples that contain important information about the signal and discard the rest. Transformation based compression techniques generally detect the redundancies utilizing the spectral and energy distribution analysis. Among transformation schemes, wavelet transformation (WT) [4] has become very popular due to the fact that the time–frequency kernel for the WT-based method can better localize the signal components in time–frequency space. On the other hand parameter extraction methods [5] for compression are irreversible process. These methods are mainly based on linear prediction and long-term prediction methods.

One of the growing issues in rural health care system in India is to expand the service among the poor population distributed at huge geographical area with poor connectivity in terms of infrastructure. ECG is considered as one of the principal physiological signal to detect the cardiac abnormalities of human being.

In the present work, a group of 8 consecutive ECG data (sample) is taken at a time from input ECG data file. Sign byte of these 8 voltages is generated and those voltages are multiplied by a suitable large number to obtain integers. Those integers are now grouped in forward direction except few critical numbers that are reserved as some special ASCII characters and are replaced by other suitable. At last these grouped integers along with other necessary information (sign byte, amplification factor etc.) are printed in their corresponding ASCII characters.

After compression it is need to transmit the data to some distant health care center or to some expert cardiologist for immediate action. In [6], a telemedicine system was proposed where full channel model with multiple faders, co-channel or adjacent channel interference and noise profiles were considered as the main modeling parameters of GSM channel. GSM link was also used in [7] to develop a portable emergency telemedicine device that supports real-time transmission of critical bio-signals as well as still images of the patient. Use of CDMA network [8] was also proposed for telecardiology application.

Here the compressed file contains only ASCII characters. Therefore SMS system is used to transmit the compressed data. But the problem is mobile messaging system can transmit only 7-bit ASCII characters (0-127) whereas the compressed file contains 8-bit ASCII characters (0-255). Therefore every character in the compressed file is converted to two suitable 7-bit ASCII and then transmitted.

In the reconstruction part of the algorithm, all mobile messages are taken together and 7-bit ASCII characters are converted into corresponding 8-bit ASCII characters. Then, using just the reverse programming logic of compression, original ECG signal is reconstructed.

II. METHODOLOGY

Proposed algorithm is divided into following three main sections: (i) Data compression (ii) Transmission and Reception and (iii) Data reconstruction. All these compression, transmission-reception and reconstruction algorithms are explained sequentially in the rest of the sections. Block schematic of the proposed algorithm is shown in figure 2.

A. Data Compression Protocol

The algorithm proposed by us in [9] is implemented here to compress the ECG data. At a glance 8 ECG samples are taken from the digitized input ECG data file and saved in an array. A sample array is shown below.

0.060	0	0.007	-0.001	0.007	0.020	0.015	-0.040
e[0]	e[1]	e[2]	e[3]	e[4]	e[5]	e[6]	e[7]

The sign of every element of this array is checked. For every positive number a binary zero (0) and for every negative number a binary (1) is taken as sign bit for that corresponding 'Voltage' value. Decimal equivalent of this binary string will be used as the sign byte of these corresponding eight 'Voltage' values. Following this rule, from the elements of the above array (e[]), a string like ''00010001'' can be obtained whose decimal equivalent is 17. This will be treated as the sign byte of those corresponding eight 'Voltage' values. Now the sign byte is converted into its corresponding ASCII character and printed in the output file. In the next step, maximum number is found out from that array. Using this maximum number a common amplification factor for those eight "voltage" values is generated in such a way that after multiplication, integer part of each of these "voltage" values will be either less than or equals to nine (9). Here for this example, the maximum number is 0.060. So the amplification factor is 150 (9/0.060). These amplified voltage values (only integer part) will be stored in an array. After amplification, numbers of e [ ] array will become as below and saved in another array c [ ].

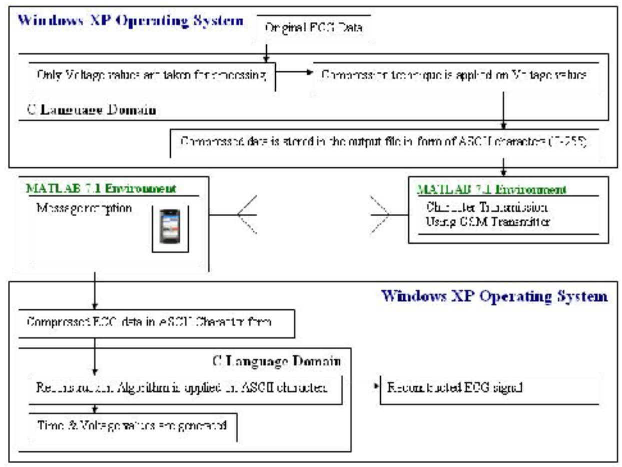

Fig. 2. Schematic of the proposed algorithm.

9	0	1	0	1	3	2	6
c[0]	c[1]	c[2]	c[3]	c[4]	c[5]	c[6]	c[7]

Now these eight amplified 'Voltage' values are grouped into just four using the following logic.

Every number in c[i] position is multiplied by 10 and added with the number in c[i + 1] position and also these grouped integers will be printed in their corresponding ASCII character. The counter, ''i'' (initially set to zero) will be incremented by 2. Now, all the eight integers are grouped into just four.

9	0	1	0	1	3	2	6
90		10		13		26	

If any grouped integer becomes 10 or 13 or 26, they will be changed to 11, 14 and 27 respectively because at the time of data reconstruction, compiler will treat these characters as End of File (EOF) and program will get terminated.

So, in this particular case, 90, 11, 14 and 27 will be printed in the output file in their corresponding ASCII character. According to the given examples in compression methodology, a string like "◄Z♂♫←û" will appear in the output file as shown below.

17	90	11	14	27	0	150
◄	Z	♂	♫	←	Space	û

Every set of four grouped integers and other associated information like 'sign-byte', 'Amplification factor' etc. will be printed in their corresponding ASCII characters maintaining the following format.

Sign Byte	Four Grouped integers	rs	Amplification Factor

Every character set is separated from each other by an 'ENTER' character. In the original ECG data file if there are only 8 time values and 8 voltage values separated by TAB delimiter, size of that file becomes 102 byte (minimum). Whereas in the compressed file there are only 7 ASCII characters (8 bit). So the compressed file size is 7 byte only. Hence 14.57 is the minimum compression ratio that can be achieved using this algorithm. The above-described algorithm will be executed repeatedly until all the ECG samples are become compressed.

B. Transmission and Reception Protocol

The compressed data file contains only ASCII characters. Therefore GSM bases SMS system is used to transmit the compressed data. An 'i-300' GSM transmitter is used for this purpose.

Mobile messaging system can transmit only 7-bit ASCII characters (0-127) but the compressed file contains 8-bit ASCII characters (0-255). Therefore every character in the compressed file is converted to two suitable 7-bit ASCII characters. Although mobile message support 7-bit ASCII characters but not all. There are some specific 7-bit characters which cannot be transmitted through GSM modem. For example, all characters between 0-31, 91 ([ ), 92 (\), 93 ( ]), 96 (`), 127 (⌂) and few more cannot be transmitted through GSM modem. Therefore these characters should be avoided. The rule followed here to convert 8-bit to 7-bit ASCII character is very simple and is described as below.

Two variables are taken and named 'MSB' and 'LSB'. For any 8-bit ASCII less than or equals to 31, MSB is set to 32 and LSB is set to '32 + that 8-bit ASCII'. Therefore, both MSB and LSB will be always greater than 31 and hence can be transmitted through SMS. For the numbers ranging from 32 to 126 a different technique is used. For these numbers, MSB is set to 33 and LSB is kept same as the original number.

X= any 8-bit ASCII (0-255)
if X <= 31
  'MSB'=32 and 'LSB' = X+32.
If X > 31 and X<= 127
  'MSB'=33 and 'LSB' = same as original

Although 91, 92, 93, 96 and 127 falls in this range but these three characters can't be transmitted through SMS. Therefore these three are treated in a different way. For these numbers MSB is set to 34 and LSB is se to 'number-6'. This is nothing but to overcome the problem.

Almost same rule is followed for rest of the 8-bit ASCII (128-255). The thing is to be noted that both MSB and LSB must reside under valid 7-bit ASCII characters and the combination of these two must be unique for every 8-bit ASCII character. At last both MSB and LSB are printed in the output file.

SMS system can send 140 characters per SMS. An algorithm is developed which divides the compressed data file into a number of small data files each containing 140 7-bit ASCII characters. Out of those 140 characters, the first character is reserved for patient ID, second and third characters are reserved for SMS numbers and rest 137 characters are used for transmitting the compressed ECG data. Now those small data files are transmitted through an 'i-300' GSM transmitter attached through computer's serial port. Standard Attention Commands or AT commands have been used to communicate with the transmitter. Patient ID and SMS number also start from 32. As the patient ID and SMS number is embedded inside the message body, multiple patients' messages can be transmitted simultaneously to a particular mobile phone and if any SMS is transferred before it's previous due to some network problem or something else, there will be no effect at the time of data reconstruction.

At the receiver end all those received messages are to be transferred from the mobile phone to the computer via USB cable, Bluetooth technology or by any means. There is also a reverse algorithm at the receiver end which concatenates all the SMS according to their SMS number of a particular patient ID, converts 7-bit ASCII characters to 8-bit ASCII characters and makes separate file for different patient ID. Each of these file is same as its original compressed data file. 7-bit to 8-bit ASCII conversion is done using just the reverse algorithm. A variable 'Y' (say) is taken to hold 8-bit ASCII characters. The conversion algorithm is given below.

if MSB = = 32
Y = LSB – 32
if MSB = = 33
Y = LSB
if MSB = = 34
Y = LSB + 6.

The SMS generation and concatenation algorithm is implemented on MATLAB 7.1 platform. A 'Samsung Wave 525' mobile is used at the receiver end to receive SMS.

C. Data Reconstruction Protocol

Reconstruction algorithm is developed using just the reverse logic of compression. From the compressed data file at a time one set of character is taken. Equivalent ASCII values of those characters are saved in an array and then reconstruction algorithm is applied on those to restore original eight 'Voltage' values. The first ASCII value among those contains the 'Sign byte' of those eight voltage values. Following four ASCII values are the grouped integers and the sixth and seventh ASCII values are 'rs' and 'Amplification factor' respectively. Now from those seven ASCII values, 'Sign byte', 'rs', and 'Amplification factor' are stored separately in different variables. Grouped integers are ungrouped using the reverse logic of grouping. Now, every ungrouped number is divided by the 'Amplification factor' and further saved in another array. In the next step, the sign-byte will be converted into its corresponding 8-bit binary equivalent. In the binary string if any bit is '1' then the corresponding ungrouped integer will be multiplied by (-1). . This array contains the reconstructed ECG voltage values. A variable 'x' (say) is declared and is initialized with zero (0). For PTD-DB ECG database the sampling frequency of the original ECG signal is 1kHz. Therefore the sampling interval is 0.001 second. Hence, in each iteration, 'x' will be incremented by the sampling interval and will be printed with the reconstructed ECG samples.

IV. RESULT

A data compression algorithm must also reproduce the data with acceptable fidelity. In biomedical data compression, we usually determine the clinical acceptability of the reconstructed signal through visual inspection. We may also measure the difference between the original and the reconstructed signal mathematically. Such a numerical measure is the percent root-mean-square difference, PRD, given by equation (1). Where y_i represents the original ECG sample and $\bar{y}_i$ represents the reconstructed ECG sample. The normalized version of PRD is PRDN, which does not depend on the signal mean value S_{mean}, given by equation (2).

$$PRD\% = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{\sum_{i=1}^{n} y_i^2}} X\,100\% \qquad \ldots(1)$$

$$PRDN\% = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - s_{mean})^2}} X\,100\% \qquad \ldots(2)$$

One another numerical measure given in equation (3) is Quality Score (QS) was proposed in [10] to quantify the overall performance of compression algorithm considering both the CR and PRD. A high score represents a good compression performance. The compression ratio (CR), which is defined in equation (4), is also calculated.

$$QS = \frac{CR}{PRD} \qquad \ldots(3)$$

$$CR = \frac{B_o}{B_c} \qquad \ldots(4)$$

Where B_o is total number of bits in the original is file and B_c is total number of bits in the compressed file. The compression algorithm achieves PRD of about 7.89%, CR of about 15.72, PRDN of about 20.6% and QS of about 1.99. Figure 3 to 5 shows different original (Blue-1) and reconstructed (Green-2) ECG signal processed by this algorithm.

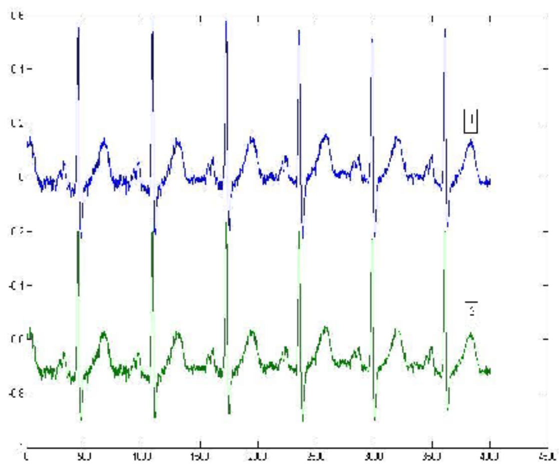

Fig. 4. Original (Blue-1) & Reconstructed (Green-2) ECG signal, File:0274, Lead I (Normal).

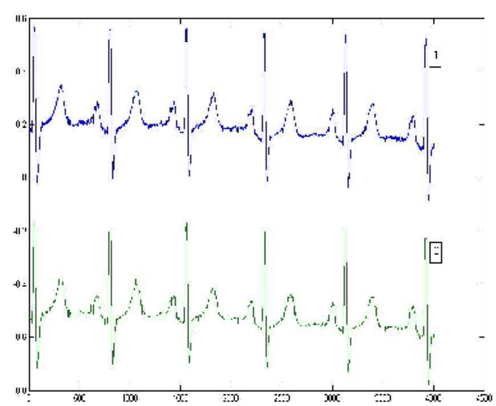

Fig. 5. Original (Blue-1) & Reconstructed (Green-2) ECG signal, File: S0301 Lead I, (Normal).

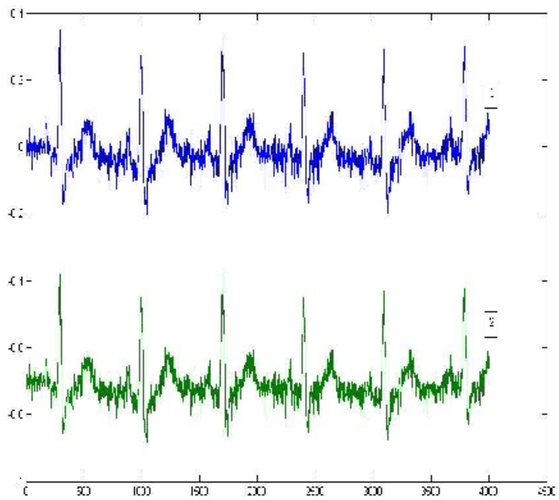

Fig. 6. Original (Blue-1) & Reconstructed (Green-2) ECG signal, File: S0036LRE, Lead I (Myocardial Infarction, Anterior).

Original and reconstructed ECG signals are checked by renowned doctors and they have given their valuable comments. According to their visual inspection there is no difference between the original and reconstructed ECG signal. To evaluate the performance of the proposed algorithm, PTB diagnosis ECG database (http://physionet.org/cgi-bin/atm/ATM) is used. This library contains a huge collection of all 12 lead simultaneous ECG recordings and the database contains a high percentage of pathological ECG. Both the compression & reconstruction algorithms require less than 1.2 seconds for processing 29,345 samples of one signal which makes suitable to compress, transmit and reconstruct real time ECG signals. Compression and reconstruction was done by ‘‘TURBO C++ IDE Version 3.0’’ and message generation and reconstruction was done by MATLAB 7.1 with a computer having RAM 2 GB DDR1, Pentium 4 CPU 2.66 GHZ and Windows XP operating system.

V. CONCLUSION AND DISCUSSION

A new approach for compression and transmission of ECG signal is proposed here. As we get ASCII characters as output, some standard ASCII compression algorithm can further compress the data. Compressed file can be preserved for later diagnosis purpose or it can also be transmitted to doctors' mobile phone to consult about the condition of the patient's heart in case of emergency. Things to be done at the doctors' end are to transfer all those SMS to the computer or laptop via USB cable, Bluetooth technology or by any means and to run the reconstruction algorithm. It is observed that on average 10 SMS is required to transmit two complete ECG cycles. Anyone who has only the preliminary knowledge about computer and internet can use the module at the transmitter end. Research on ECG feature extraction [11-13] and classification [14] is being carried out. This combined module of ECG signal compression, transmission, reconstruction, feature extraction and classification could be a better choice in telemedicine application. Moreover the system could be a solution to the poor infrastructure in rural health management in India. In certain season, the connectivity in remote villages particularly in hilly regions becomes very poor. Therefore the vast rural population in India can be benefited from using this system. The algorithm was tested by two types (Normal and Myocardial infarction) of 12 different leads i.e., of 24 types of ECG waves taken from PTB diagnosis ECG database in order to evaluate the performance of the proposed algorithm.

ACKNOWLEDGEMENT

The authors acknowledge their deepest gratitude to Dr. Suranjan Bhattacharya and Dr. R.C. Saha, cardiologist, Kolkata, India, for their valuable clinical advice.

REFERENCES

[1] L. Sornmo, and P. Laguna, “Bioelectrical signal processing in cardiac and neurological applications”, 2005 (Amsterdam: Elsevier Academic Press).

[2] J. R. Hampton, “The ECG Made Easy”, Churchill Livingstone, Sixth Edition.

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

[4] C.T. Ku, H.S. Wang, K.C. Hung and Y.S. Hung, “A novel ECG data compression method based on non-recursive discrete periodized wavelet transform”, IEEE Trans. Biomed. Eng 2006, 53(12), pp. 2577–83.

[5] G. Nave and A. Cohen, “ECG compression using long-term prediction”, IEEE Trans. Biomed. Eng1993, Vol. 40, No. 9, 877–885, Sep.

[6] R. S. H. Istepanian and A. A. Petrosian, “Optimal Zonal Wavelet-Based ECG Data Compression for a Mobile Telecardiology System”, IEEE Transactions on Information Technology in Biomedicine, Vol. 4, No. 3, September 2000.

[7] S. Pavlopoulos, E. Kyriacou, A. Berler, S. Dembeyiotis and D. Koutsouris “A novel emergency telemedicine system based on wireless communication technology - AMBULANCE,” *IEEE Trans. Inform. Tech. Biomed*. – Special Issue on Emerging Health Telematics Applications in Europe, vol. 2, no. 4, pp. 261-267, 1998.

[8] B. S. Kim and S. K. Yoo, “Performance evaluation of wavelet-based ECG compression algorithms for telecardiology application over CDMA network”, Medical Informatics and the Internet in Medicine, September 2007, 32(3), pp. 177-189.

[9] S. K. Mukhopadhyay, S. Mitra, M. Mitra, “An ECG signal compression technique using ASCII character encoding”, Measurement 45 (2012) 1651–1660.

[10] C.M. Fira and L. Goras, “An ECG signals compression method and its validation using NNs”, IEEE Trans. Biomed Eng 55 (4) (2008), pp. 1319 1326.

[11] S. K. Mukhopadhyay, M.Mitra, S.Mitra, “Time Plane ECG Feature Extraction Using Hilbert Transform, Variable Threshold and Slope Reversal Approach”, Proc. of (CD) *IEEE International Conference on Communication and Industrial Applications - ICCIA* 2011, 26-28th December 2011, Kolkata, India,pp.1-4.

[12] S. K. Mukhopadhyay, M.Mitra, S.Mitra, “ECG Feature Extraction Using Differentiation, Hilbert Transform, Variable Threshold and Slope Reversal Approach”, *Journal of Medical Engineering and Technology*, October 2012, Vol. 36, No. 7, pp- 372-386.

[13] S. K. Mukhopadhyay, M.Mitra, S.Mitra, “QRS Complex Identification Using Hilbert Transform, Variable Threshold and Slope Reversal Approach”, International Journal of Biomedical Engineering and Technology, Vol. 9, No. 4, 2012, pp-301-315.

[14] S. Mitra, M. Mitra, B.B. Chaudhuri, “A Rough-Set-Based Inference Engine for ECG Classification” IEEE Transactions on Instrumentation and Measurement, Vol. 55, No. 6, December 2006, pp- 2198- 2206.