Header Format for Selective Mapping Technique on Wavelet based ECG Compression

Yotaka Chompusri*, **Kobchai Dejhan**, Guest members, and **Surapan Yimman**, Member

ABSTRACT

The error from applying compression method based on Selective Mapping technique on both regular and irregular period ECG signal is very small. But, when more beat requires the Dynamic Time Warping technique than Period Scaling technique, the more additional information is produced and the size of this header data is large. This data contains the mapping information which is importance information in reconstruction process. In case of Dynamic Time Warping technique, the header size is greater than 8 times of residual size therefore it enormously reduces the Compression Ratio. To improve the compression performance by not affecting the error, the designed header format is proposed to lower the size of additional data producing from mapping process. As the result, this arrangement of new header format can reduce the header size lower than 7.5 times of original header size. Therefore, this header format will increase the compression performance.

Keywords: ECG, Compression, Dynamic Time Warping, Discrete Wavelet Transform, Header Format

1. INTRODUCTION

Electrocardiogram (ECG) is the electric conduction of heart rhythm. It is the useful clinical information being used to diagnose the heart disease condition. The size of ECG is large therefore large memory is needed to store or transmit them. Many research have been proposed to reduce the memory consuming of ECG processing such as time domain technique and transform domain technique [1-3]. Among many compression methods, there are some methods working well on high varying period ECG signal [4, 5]. On this work, the goal is to implement the compression method which its performance is good on both regular and irregular period ECG signals. This method is based on residual difference from reference beat subtraction and Wavelet Transform (WT). To receive small value of residual difference and to gain small error between original signal and reconstructed signal, the proper matching technique which its process does not change any information of original information is applied. From those reasons, two matching techniques are used which first technique is Period Scaling and second technique is Dynamic Time Warping (DTW). Furthermore, each technique is proper to difference class of period regularity and produces difference size of additional data. Period Scaling technique is suitable to regular period signal and small size of additional data is produced. In the meantime, the DTW is appropriate to irregular period signal and generates large size of additional data. Since this additional data is summed up to total compressed data, this will lose the ability to compress signal resulting in low value of Compression Ratio. In this research, the arrangement of those additional data is proposed. Because the additional data from matching process has significant effect on performance of compression, lowering the additional datas size will improve the compression performance. The suitable header format to rearrange the additional data from mapping technique can reduce the size of those additional data resulting in smaller header information. Consequently, the Compression Ratio is better while the error in reconstructed signal is not affected by header format.

2. THE ALGORITHM

Fig. 1, shows the algorithm of this compression method. First, the signal is separated into each signal beat by QRS complex detection and then the reference beat is computed. Then all signal beats are matching to the reference beat by selective mapping technique and the additional data from mapping technique is formatted into header data. Then, the residual difference between signal beat and reference beat is calculated. After that, the residual is decomposed by Discrete Wavelet Transform (DWT) and thresholding process eliminates redundant information. Lastly, the Huffman encoding method is applied to the residual after thresholding and the header data.

* Corresponding author.
Manuscript received on XXX XX, XXXX, revised on XXX XX, XXXX.
Surapun Yimman is with the Industrial Physic & Medical Instrumentation Department, King Mongkuts University of Technology North Bangkok, 10800, Thailand.
E-mail addresses: sym@kmitnb.ac.th

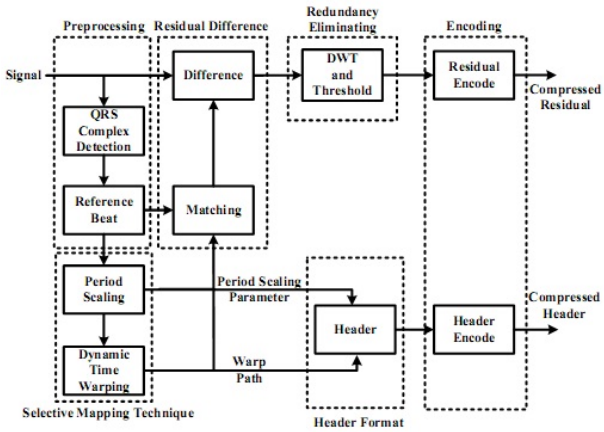

Fig.1: *Block diagram of proposed compression algorithm.*

2.1 Preprocessing

In preprocessing block of Fig. 1, the QRS complex is detected by Hamilton and Tomkins [6] then signal beat is defined between R-R intervals as shown in Fig. 2 . After that, the reference period is searched from every beat and defined from the most frequent period. And then, each value of reference beat is computed from averaging all signal beats which period is equal to reference period. When all signal beats and reference beat are found, the matching technique is applied.

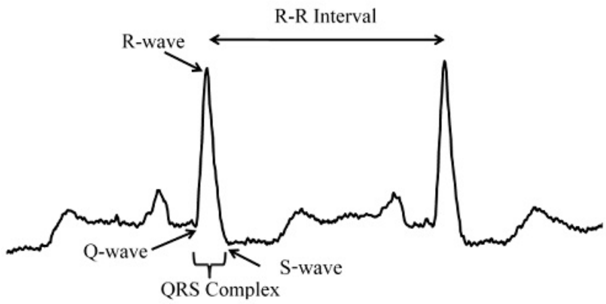

Fig.2: *ECG Signal.*

2.2 Selective Mapping Technique

This process will select the most appropriate mapping technique for mapping each beat. The Percent Root Mean Square Difference (PRD) which is the parameter for measuring the error of reconstructed beat is used to evaluate which mapping technique will be applied to map signal beat to reference beat.

Firstly, Period Scaling technique is a linear mapping technique. This technique scales the period of reference beat to be equal to the period of signal beat. This technique decides which reference point will be repeated or skipped depending on how much difference between two signal periods. And, the period difference is defined as period scaling parameter. From this process, signal beat is mapped to reference beat without concerning the similarity between two signals which could cause the error in reconstructed beat. So that, this technique is only suitable for signal beat which its shape is similar to the shape of reference beat. The most benefit of this technique is that only one additional data is produced, which is the period scaling parameter. As the result, this period scaling parameter is not much affected to Compression Ratio.

After that, the beats PRD is calculated. If the error is lower than the beat PRD threshold level then the next signal beat will be processed. On the other hand, if the error is higher than beat PRD threshold level, the Dynamic Time Warping technique is applied to that signal beat instead of Period Scaling technique. Dynamic Time Warping technique is the nonlinear mapping technique. This technique can measure the similarity between two signals which are not equal in time series [7]. The DTW technique can map the signal point to the most correlate point of reference beat by not alter the original signal, consequently the residual difference value is always small [8] both in case of regular and irregular period signal. And another advantage is that peak point of each beat is always matched, thereby peak is surely brought back in reconstruct signal. The disadvantage of DTW technique is the additional data which is the warp path producing from DTW process. Warp path contains the matching index between signal beat and reference beat. This additional data causes of lowering the Compression Ratio

2.3 Residual Difference

Residual difference value is the difference between signal beat and reference beat. When Period Scaling is selected, the subtraction point from reference beat is computed from Period Scaling algorithm. If the mapping technique is Dynamic Time Warping, the warp path indicates which point of reference beat will be subtracted from signal beat.

2.4 Discrete Wavelet Transform and Threshold

Residual Then, the Discrete Wavelet Transform is applied to decompose the residual difference signal into wavelet coefficient. The dabechies mother wavelet is used in three levels of decomposition. After that, the thresholding is applied to remove the redundant information from residual signal. The threhold value is based on Energy Packing Efficiency (EPE). The EPE is the percentage of preserved energy after threshold with respect to the total energy before threshold in each level of sub-band [9] as in (1).

$$EPE_{Ci} = [\overline{E}_{Ci}/E_{Ci}] \times 100\% \tag{1}$$

Where $\overline{E}_{Ci}$ is the total energy of i level coefficient after thresholding and E_{Ci} is the total energy of i level coefficient before thresholding.

Thresholding eliminates nonessential information and remains information which contributes the desire energy. The coefficient which magnitude value is less than threshold value is set to zero value. And this causes majority of signal to be zero. Since more than half of energy is contributed in approximate coefficient, the high *EPE* value of approximation coefficient, EPE_{AC}, is selected to keep the essential information. In contrary, the lower *EPE* value of detail coefficient, EPE_{DC}, is chosen to mitigate the non necessary information. signal.

2.5 Encoding

The residual signal after thresholding and the header data after formatting are encoded by Huffman method. This encode method is the minimum-redundancy coding which is preferable to the data which are repeatedly. This process efficiently reduces the size of ECG signal.

2.6 Reconstruction process

Fig. 3, shows the reconstruction method. At the beginning, the compressed residual and the compressed header are decoded and then Inverse Discrete Wavelet Transform is applied to residual signal. In the mean time, the algorithm finds which mapping technique is applied to each beat and then the matching points are calculated back from header data. After that, the reconstructed signal is received from adding the matching point of reference beat back into the output signal from Inverse Discrete Wavelet Transform process.

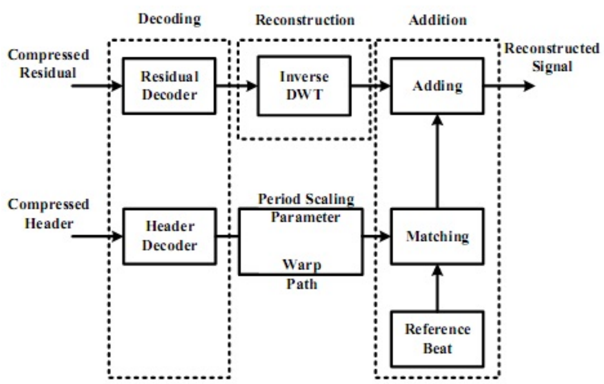

Fig.3: *Reconstruction method.*

3. HEADER FORMAT

After compression, the header data and residual signal of one beat is arranged as shown in Fig. 4.

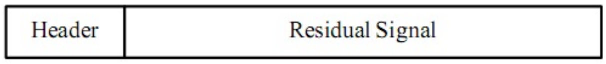

Fig.4: *Compressed signal of one beat.*

Fig. 5, shows header format which composes of header I and header II. The header I consists of 11 bits. First bit indicates the selected mapping technique, Period Scaling or Dynamic Time Warping. The second bit indicates whether the signal beat is shorter or longer than reference beat. And, the rest bits which are bit 3 to bit 11 store the period difference between signal beat and reference beat. The header II is the additional information which is the warp path generating from Dynamic Time Warping technique.

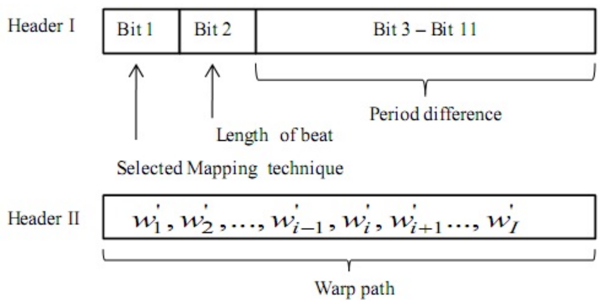

Fig.5: *Header Format.*

In Period Scaling case, the mapping information is the period scaling parameter which is the period difference between signal beat and reference beat. This mapping information requires only 11 bits per signal beat. Thereby, only header I is generated from Period Scaling technique.

In Dynamic Time Warping case, the mapping information of this technique is warp path, therefore this information requires both Header I and Header II. The additional header II is used to store the warp path which its size is large. In this work, the header format is proposed to lower the size of warp path.

Given signal beat, X, which length is I as in (2) and reference beat, Y, which length is J as in (3).

$$X = x_1, x_2, ..., x_i, ..., x_I \tag{2}$$

$$Y = y_1, y_2, ..., y_j, ..., y_J \tag{3}$$

Warp path, W, contains the matching index between signal beat and reference beat as defined in (4).

$$W = w_1, w_2, ..., w_i, ..., w_I \tag{4}$$

where w_i = *(i, j)* is the matching pair, i is the index point of signal beat and j is matching index point of reference beat.

The warp path length is equal to the length of signal beat and each warp pair consists of two index points. Therefore, the size of warp path is about two times the size of original signal. So this additional data will enormously reduce the Compression Ratio.

To improve the compression performance, the warp path is rearranged in this research. Only the matching index point of reference beat is kept in warp pair therefore warp path size is reduced to a half. Then the index pair is rearranged as (5).

$$w_i = j \tag{5}$$

where i is the index point of signal beat and j is the matching index point of reference beat. From the property of warp path, the warp path is monotonically increasing therefore the matching index point of upper warp pair is always higher than or equal to the matching index point of lower warp pair as in (6).

$$w_i = j, w_i + 1 = j'; j \le j' \tag{6}$$

To reduce the number of bit storing warp path, the modified warp path, W, is generated as in (7) and (8).

$$w'_I = w_I \tag{7}$$

$$w'_I = w'_i - w'_{i-I}; 2 \le i \le I \tag{8}$$

And the block diagram of calculation process is shown in Fig. 6. The modified warp path will keep the difference value between position of matching index point and the position of lower adjacent matching index point instead of the value of matching index point.

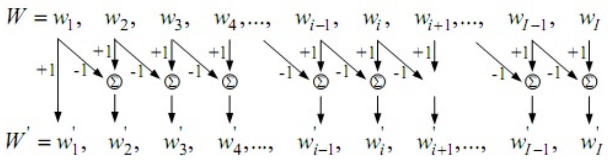

Fig.6: *Modified Warp Path*

This modification makes the warp paths size to be smaller. Consequently, the index value of modified warp path is small and more repeated. Therefore, least bits are required to store them and this kind of data is suitable to the encoding method such as Huffman

4. EVALUATION

The ECG signal in this experiment is from MIT-BIH database. Only 100,000 points of each signal which sampling rate is 360 Hz and 11 bits/sample is tested. The aim of this research is to reduce the size of header data producing from mapping technique, especially the warp path from Dynamic Time Warping technique.

On this experiment, 43 signals including both types of signal classified by regularity of period are tested.

Two parameters to assess the performance of compression method are Percent Root Mean Square Difference and Compression Ratio.

The Percent Root Mean Square Difference (PRD) is the indicator to measure the error between the original signal and the reconstructed signal as in (9).

Where x_i is original signal and is the reconstructed signal.

Compression Ratio (CR) is the ratio between original signals data size and the total compressed signals data size as in (10).

$$CR = \frac{N_x}{N_{rb} + N_{ch} + N_{cr}} \tag{9}$$

Where N_x is the number of original signals bit, N_{rb} is the number of reference beats bits, N_{ch} Ncompressed header is the number of compressed headers bits, N_{cr} is the number of compressed residual signals bits.

In this work, the new parameter is defined as Header Size Ratio (HSR) which is the ratio between the compressed original headers size to the compressed proposed headers size as in (11) to indicate the performance on decreasing the header data

$$HSR = \frac{N_{oh}}{N_{ph}} \tag{10}$$

Where N_{oh} is the number of compressed original header datas bits, N_{ph} is the number of compressed proposed format header datas bits

For thresholding process, the EPEAC for approximation coefficient is set in high value, 99% and EPEDC for three level of detail coefficient are set at the same value which is 90%. The reason for high threshold value setting both of approximation coefficient and detailed coefficient is that this experiments aim is to justify the performance of the algorithm for header reduction therefore most of essential information is kept.

From the experiment, the numbers of bit storing each original ECG signal are approximately 1,100,000 bits and the numbers of bit storing reference beat of each signal record are vary between 1,375 4,895 bits which depend on feature of each signal. The numbers of bit storing compressed residual signal of each signal record are not different which are varied from 140,000 - 190,000 bits.

The size comparing between the original header format and proposed header format is shown in Table I. In this case, the PRD threshold level is set at 0 % in selective mapping process thereby all beats are mapped by Dynamic Time Warping technique. Consequently, the header data of both original format and proposed format are maximum size which could be the most affected to reduce the CR value. When only DTW technique is selected, the PRD result is very small which is always less than 0.75 although the signals are high varying such as signal record no. 119 and 203. But, the number of original headers bit is very large which its size is varying from 7.7 to 11.2 times the size of compressed residuals bit. When the header information was formatted, this rearrange process can reduce bits storing header data to be lower than 8.2 times of bits storing original header data and HSR value is about 8.2 11.1 as shown in last column of Table I. This formatting process makes the size of header data nearly to the size of compressed residual.

Table 2 and Table 3 are the results when selective mapping technique is applied. This process will lower the number of beats mapping via DTW technique by setting Beat PRD threshold level which is set to 0.5 % in both non high varying period signal and high varying period signal. If the DTW technique is less applied, the size of warp path generated from this technique will be reduced.

In Table 2, the total number of beat and number of beat which mapping by DTW technique when PRD threshold level was set on Selective Mapping technique of each signal are shown. The setting of Beat PRD Threshold level on Selective Mapping Technique lowers the number of beat switching to DTW technique. The amount of beats switching to DTW techniques is dependent on how varying of each beat. If signal is likely to be regularity, less beats are switched to DTW. Vice versa, more beats will use DTW if signal is irregular.

In Table 3, the PRD resulting was increased when decreasing the number of beats mapping by DTW. However, a number of original headers bits were decreased that was because some beats were switched to DTW technique. Thus, the header data size depends on number of beats using DTW technique that is more beats using DTW technique will return more header information. In this case, most of the herder data size is still larger than the residual signal size. Only when few beats are mapped by DTW technique which less than 8% of total beats such as signal record no. 102, 201 and 222, the original header data is smaller than the residual signal. If original header size is more than residual size, this proposed header format can reduce the size of header more than 8.2 times of original header size. The HSR value in this case of Table II is not different from Table I. On the other hand, the HSR value is clearly reduced when the original header size is less than residual size. But the HSR in this case is still high which is varying from 5 8 times.

5. CONCLUSION

By this compression technique, the result of PRD is good which is relatively low comparing to other

Table 1: *THE MAXIMUM HEADER DATA OF PROPOSED METHOD*

Signal Number	PRD (%)	Residual (bit)	Original Header (bit)	Proposed Header (bit)	HSR
100	0.14	171,104	1,520,652	169,640	8.96
101	0.48	154,802	1,467,255	142,653	10.28
102	0.18	166,654	1,476,038	175,821	8.39
103	0.16	153,314	1,583,623	184,564	8.58
104	0.48	151,452	1,524,218	174,204	8.74
105	0.17	171,689	1,511,714	180,956	8.35
106	0.5	144,888	1,554,222	168,802	9.2
107	0.42	166,718	1,512,976	170,446	8.87
108	0.34	169,724	1,555,776	164,308	9.46
109	0.24	164,090	1,525,036	181,272	8.41
111	0.26	181,100	1,520,898	167,537	9.07
112	0.2	188,408	1,467,525	167,187	8.77
113	0.37	154,832	1,484,526	158,102	9.38
114	0.19	165,203	1,606,593	165,557	9.7
115	0.31	157,936	1,420,109	144,087	9.85
116	0.34	159,332	1,497,852	172,439	8.68
117	0.21	187,595	1,575,750	169,810	9.27
118	0.29	175,777	1,549,568	171,658	9.02
119	0.75	146,524	1,566,021	160,159	9.77
121	0.35	185,044	1,502,608	156,455	9.6
122	0.28	189,003	1,498,100	172,816	8.66
123	0.24	165,790	1,614,341	173,018	9.33
124	0.32	177,772	1,578,336	151,799	10.39
200	0.74	148,825	1,489,036	165,560	8.99
201	0.12	189,039	1,459,213	175,728	8.3
202	0.15	188,037	1,578,819	173,079	9.12
203	0.68	153,411	1,475,353	159,355	9.25
207	0.55	168,324	1,448,495	129,631	11.17
209	0.23	163,518	1,481,828	178,734	8.29
210	0.34	159,995	1,432,789	165,336	8.66
212	0.29	164,314	1,466,061	175,527	8.35
217	0.32	154,856	1,545,250	171,009	9.03
219	0.25	152,027	1,530,854	171,717	8.91
220	0.19	142,419	1,576,161	186,342	8.45
221	0.3	172,425	1,480,195	161,430	9.16
222	0.21	182,461	1,449,351	154,222	9.39
223	0.21	144,626	1,552,153	183,059	8.47
228	0.6	164,481	1,540,812	161,878	9.51
230	0.31	140,523	1,491,496	175,789	8.48
231	0.16	148,580	1,645,791	181,235	9.08
233	0.73	151,635	1,451,801	173,192	8.38
234	0.26	175,805	1,433,062	170,493	8.4

methods. However, the compression performance is not high enough that is because of the additional data generating from Dynamic Time Warping technique. This warp path is very importance in reconstruction process so that this information could not be neglected. Since this additional data has enormous effect on Compression Ratio, the Selective Mapping Technique is applied. This selective process will lower the number of beats mapping by DTW technique, therefore the addition data from DTW will be smaller. Moreover, to lower the size of additional data information, the header format to rearrange the information of mapping technique and warp path is proposed. This proposed header format can efficiently reduce the size of header data. Therefore, this method can increases the Compression Ratio by no effect on error between original signal and reconstructed signal.

References

[1] S. M. S. Jalaleddine, C. G. Hutchens, R. D. Strattan, and W. A. Coberly, *ECG data compression techniques-a unified approach,* Biomedical Engineering, IEEE Transactions on, vol. 37, pp. 329-343, 1990.

[2] M. Shaou-Gang, Y. Heng-Lin, and L. Chih-Lung, *Wavelet-based ECG compression using dynamic vector quantization with tree codevectors in single codebook,* Biomedical Engineering, IEEE Transactions on, vol. 49, pp. 671-680, 2002.

[3] L. Zhitao, K. Dong Youn, and W. A. Pearlman, *Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm,* Biomedical Engineering, IEEE Transactions on, vol. 47, pp. 849-856, 2000.

[4] C. Hsiao-Hsuan, C. Ying-Jui, S. Yu-Chien, and K. Te-Son, *A high performance compression algorithm for ECG with irregular periods,* in Biomedical Circuits and Systems, 2004 IEEE International Workshop on, 2004, pp. S2/4-9-12.

[5] C. Hsiao-Hsuan, C. Ying-Jui, S. Yu-Chien, and K. Te-son, *An effective and efficient compression algorithm for ECG signals with irregular periods,* Biomedical Engineering, IEEE Transactions on, vol. 53, pp. 1198-1205, 2006.

[6] P. S. Hamilton and W. J. Tompkins, *Compression of the ambulatory ECG by average beat subtraction and residual differencing,* Biomedical Engineering, IEEE Transactions on, vol. 38, pp. 253-259, 1991.

[7] V. Tuzcu and S. Nas, *Dynamic time warping as a novel tool in pattern recognition of ECG changes in heart rhythm disturbances,* in Systems, Man and Cybernetics, 2005 IEEE International Conference on, 2005, pp. 182-186

[8] B. Huang and W. Kinsner, *ECG frame classification using dynamic time warping,* in Electrical

Table 2: *THE THRESHOLD LEVEL SETTING AND NUMBER OF BEAT APPLYING DTW*

Signal Number	Beat PRD Threshold Level (%)	Total Beat	Beat Mapping by DTW
100	0.5	341	205
101	0.5	317	305
102	0.5	336	28
103	0.5	328	316
104	0.5	358	358
105	0.5	383	161
106	0.5	288	288
107	0.5	324	324
108	0.5	326	293
109	0.5	387	380
111	0.5	320	250
112	0.5	393	265
113	0.5	265	265
114	0.5	253	51
115	0.5	291	290
116	0.5	362	362
117	0.5	231	213
118	0.5	334	334
119	0.5	299	299
121	0.5	280	209
122	0.5	390	377
123	0.5	229	229
124	0.5	228	226
200	0.5	398	398
201	0.5	408	10
202	0.5	243	35
203	0.5	450	450
207	0.5	319	319
209	0.5	441	431
210	0.5	414	365
212	0.5	424	420
217	0.5	333	333
219	0.5	347	347
220	0.5	327	322
221	0.5	365	360
222	0.5	337	25
223	0.5	368	361
228	0.5	337	337
230	0.5	367	367
231	0.5	267	264
233	0.5	475	475
234	0.5	423	375

Table 3: *THE HEADER SIZE RATIOS RESULT OF PROPOSED METHOD*

Signal Number	PRD (%)	Residual (bit)	Original Header (bit)	Proposed Header (bit)	HSR
100	0.25	144,418	920,571	103,181	8.92
101	0.49	155,775	1,413,197	137,872	10.25
102	0.39	176,520	120,619	15,731	7.66
103	0.16	153,883	1,524,850	177,866	8.57
104	0.48	151,452	1,524,218	174,204	8.74
105	0.31	184,207	651,063	78,079	8.33
106	0.5	144,888	1,554,222	168,802	9.2
107	0.42	166,718	1,512,976	170,446	8.87
108	0.36	173,502	1,422,372	150,821	9.43
109	0.25	164,599	1,497,389	177,938	8.41
111	0.32	186,801	1,192,760	131,971	9.03
112	0.31	179,059	1,004,654	117,624	8.54
113	0.37	154,832	1,484,526	158,102	9.38
114	0.38	178,034	315,328	33,360	9.45
115	0.31	157,922	1,415,605	143,640	9.85
116	0.34	159,332	1,497,852	172,439	8.68
117	0.23	188,137	1,457,978	156,886	9.29
118	0.29	175,777	1,549,568	171,658	9.02
119	0.75	146,524	1,566,021	160,159	9.77
121	0.42	193,473	1,132,024	120,188	9.41
122	0.29	190,117	1,450,425	167,440	8.66
123	0.24	165,790	1,614,341	173,018	9.33
124	0.32	178,052	1,565,149	150,587	10.39
200	0.74	148,825	1,489,036	165,560	8.99
201	0.37	177,109	35,238	6,955	5.06
202	0.37	185,932	231,304	26,133	8.85
203	0.68	153,411	1,475,353	159,355	9.25
207	0.55	168,324	1,448,495	129,631	11.17
209	0.23	164,389	1,447,374	174,723	8.28
210	0.37	160,034	1,267,624	146,820	8.63
212	0.3	164,262	1,452,509	173,978	8.34
217	0.32	154,856	1,545,250	171,009	9.03
219	0.25	152,027	1,530,854	171,717	8.91
220	0.19	143,031	1,552,832	183,618	8.45
221	0.3	171,945	1,461,927	159,557	9.16
222	0.39	180,958	100,685	12,845	7.83
223	0.22	145,719	1,523,533	179,870	8.47
228	0.6	164,481	1,540,812	161,878	9.51
230	0.31	140,523	1,491,496	175,789	8.48
231	0.16	149,237	1,628,740	179,355	9.08
233	0.73	151,635	1,451,801	173,192	8.38
234	0.28	177,853	1,273,594	151,838	8.38

and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on, 2002, pp. 1105-1110 vol.2.

[9] B. A. Rajoub, *An efficient coding algorithm for the compression of ECG signals using the wavelet transform,* Biomedical Engineering, IEEE Transactions on, vol. 49, pp. 355-362, 2002.