Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 53(67), Fascicola 2, 2008

The Classification of Electrocardiographic Signal (ECG) Perturbed by Noise Using the Wavelet Theory

Cristina Laura Stolojescu[1]

Abstract: ECG signal is a non-stationary signal meaning that it changes its statistical proprieties over time. Therefore, the most powerful tool for analyzing this type of signals is the wavelet theory. The aim of this paper is to classify the ECG signals belonging to a given database in four classes: one class for ECGs without noise and three classes corresponding to ECGs perturbed by three types of noise. The Discrete Wavelet Transform and Daubechies wavelets were used to filter and analyze the four types of signals. The ECG data is taken from the standard MIT-BIH Arrhythmia database, while the signals of noise belong to MIT-BIH Noise Stress Test Database. Various tests were elaborated in this sense and the method used was the descriptive statistics.

Keywords: ECG signal, Daubechies wavelets, multi-resolution analysis, statistical analysis.

I. INTRODUCTION

During time Biomedical Engineering passed through an evolution process and became a very important and vital domain, while biomedical engineers succeeded to perform in diverse activities such as assisting in the diagnosis and treatment of patients. The application of computers to the signals for machine interpretation was one of the earliest uses of computers in medicine and enabled the recording of events that are predictive of future life-threatening happenings, [1].

Very important information according with the activity of human internal organs is provided by signals recorded on the human body surface.

The electrocardiogram (ECG) is the recording of the electrical activity of the heart, [1-3] measured over time, in order to analyze the signals characteristics and proprieties and to connect these characteristics and proprieties with normal or pathological functions. Unfortunately, such information cannot be available directly from the recorded signals because ECG signals are always contaminated by various types of noise like the higher-frequency noise due to muscle contraction (EMG noise), the lower-frequency noise due to motion artifacts (baseline wandering), the low-frequency noise in the skin-electrode interface or the effect of respiration, [1-3]. Therefore, some additional processing techniques are required to extract signal features and components that are of diagnostic importance, [1].

Kadbi *et al* [9] and Prasad *et al* [10] used the wavelet transform and artificial neural networks to accurately classify ECG arrhythmias considering signals from the same database we used: MIT-BIH Arrhythmia database.

In the present paper we propose an algorithm based on Daubechies wavelets and statistical analysis to realize the classification of ECGs in four classes. We consider two cases: in the first one the four classes are: ECG without additional noise and ECG perturbed by two of the three types of noise: baseline wonder (the isoelectric line), 'bw', muscular noise (EMG noise), 'ma', and electrode motion artifact, 'em', while in the second case the four classes are: simple ECG and the other three classes correspond to ECG perturbed by only one type of noise.

The paper is organized as follows. The Discrete Wavelet Transform is described in Section II. The proposed algorithm is presented in Section III, while the final section is dedicated to results and conclusions.

II. DISCRETE WAVELET TRANSFORM

Biomedical signals can be analyzed in time domain, frequency domain, or time-frequency domain.

While the Fourier transform is useful for analyzing the spectral content of a stationary signal and for transforming difficult operations into very simple ones in the Fourier dual domain, the Wavelet Transform (WT) is a powerful time-frequency signal analysis tool which it is used in a wide variety of applications including biomedical signal processing, [2-4].

The Discrete Wavelet Transform (DWT) has two parameters: the wavelet mother ψ and the number of iterations, [4]. Discrete wavelets can be scaled and translated in discrete steps and a wavelet representation is the following:

1 Facultatea de Electronică şi Telecomunicaţii, Departamentul Comunicaţii
Bd. V. Pârvan Nr. 2, 300223 Timişoara, e-mail: c.stolojescu@etc.utt.ro

$$\psi_{j,n} = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t - n2^j}{2^j}\right) \qquad (1)$$

where j is the scale factor and n is the translation factor, [4-6].

The discrete wavelet transform is a method of obtaining the time-scale representation of a digital signal using digital filtering techniques. One of the most used signal processing functions is the filtering, therefore the discrete time-domain signal to be analyzed in the wavelet domain is passed through a series of successive low-pass filters ($\bar{g}$) to analyze the low frequencies and high-pass filters ($\bar{h}$) to analyze the high frequencies, followed by under-samplers. The filtering operations determine the signal's resolution, meaning the quantity of detail information in the signal, while the scale is determined by up-sampling and under-sampling operations, [5].

The decomposition algorithm, also known as Mallat-tree decomposition is shown in fig.1. At each level of decomposition, the high-pass filter $\bar{h}$ produces the detail information $d[n]$, while the low-pass filter $\bar{g}$ produces the approximations, $a[n]$.

The reconstruction operation is the reverse process of decomposition, meaning that at each level the details and the approximation coefficients are up-sampled by two and then are passed through the low-pass and high-pass filters, [2-5].

In our case the ECGs are decomposed in eleven levels, so the signals have the form:

$$a_0 f \approx d_{-1} f + d_{-2} f + \ldots + d_{-11} f \qquad (2)$$

where d and a represent the detail respective the approximation information.

To provide perfect reconstruction, from the variety of wavelets families, of particular interest are the Daubechies' orthonormal wavelets of compact support with the maximum number of vanishing moments, [4, 5]. From these, the most common are D4, D6 and D8, where the index number refers to the number of coefficients of the corresponding discrete-time filters $\bar{h}$. The number of vanishing moments is equal to half of the number of those coefficients, [6].

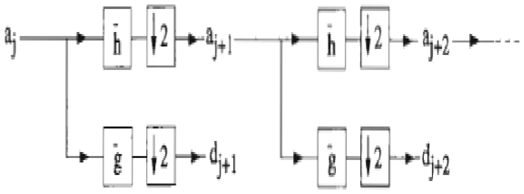

Fig.1: Wavelet decomposition tree, [4].

From the Daubechies wavelets family we choose D6 (with 3 vanishing moments) because it is similar in shape to a real ECG signal (fig.2).

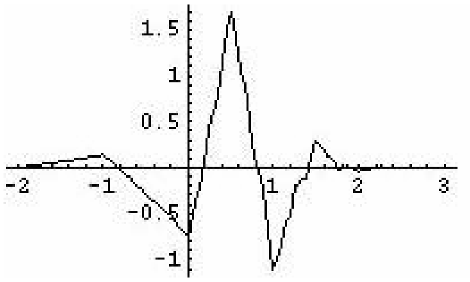

Fig.2: Daubechies6 wavelet.

III. STATISTICAL ANALYSES

The 'boxplot' (box-and-whisker diagram) is a very useful function from Matlab statistical toolbox used to handle many data values, to graphically display the mean and variance of the distribution, to compare different categories of data and to draw conclusions. The box has lines at the lower quartile, median and upper quartile values, while the whiskers are the lines extending from each end of the box to show the extent of the rest of the data.

In fig.3 are presented the results obtained applying the 'boxplot' function at one of the eleven levels of decomposition. The graphical representations correspond to the four types of signals used to test the proposed classifier. These signals are exemplified in Section IV, in the fig. 5-9. The order of the analyzed signals in fig. 3 and also in table II is: 1- ECG+bw+em, 2- ECG+bw+ma, 3- ECG+ em+ma, 4- ECG.

To apply the 'boxplot' function it is necessary to do some measures on the data set. The variance is a measure of numerical data's spread in the set, [8].

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \qquad (4)$$

In this formula X_i represents the data's values, $\bar{X}$ their statistical mean value, X is the set of data, and n is the quantity of data. The spacing between the different parts of the box indicates the degree of dispersion, also called statistical variability or data's spread.

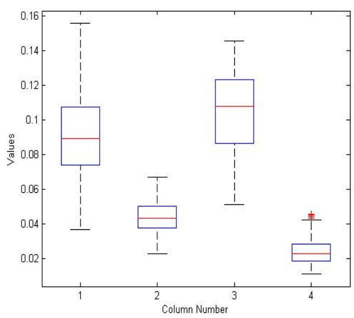

Fig.3: 'Boxplot' function applied to the ECGs used in the project.

The values of the variance for the four considered ECGs have been saved in four tables (*tvd* 1, 2, 3 and 4- table variance detail) of one hundred lines, because

we choose one hundred signals of each type and eleven columns, because the signals were decomposed in eleven levels of details, fig.4.
In order to compare the four ECGs at each level, eleven new tables were conceived. They contain the values saved before in the *tvd* tables and are defined in the following:

level_p=[tvd1(:,p);tvd2(:,p);tvd3(:,p);tvd4(:,p)]; (3)

where p is the index of the level (1, 2…11).

Applying the 'boxplot' function on these new eleven tables we obtained eleven graphical representations like the one shown in fig.3.
The Standard Deviation (SD) of a data set is another important parameter that measures how the data contained in a set is spread out and it represents the average distance from the mean of the data set to a point, (a particular data value) [8].
The standard deviation is calculated with the formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}} \qquad (5)$$

The SD is useful in the Principal Component Analysis (PCA) which is a classical statistical method used to extract relevant information from data sets, [1, 2, 8].
The PCA gives variables plane representations ('projections') that are obtained by retaining only two principal components among most significant.
In [2] various tests were made in order to obtain the signals' classification. In this paper is presented only the test that brought the best results.

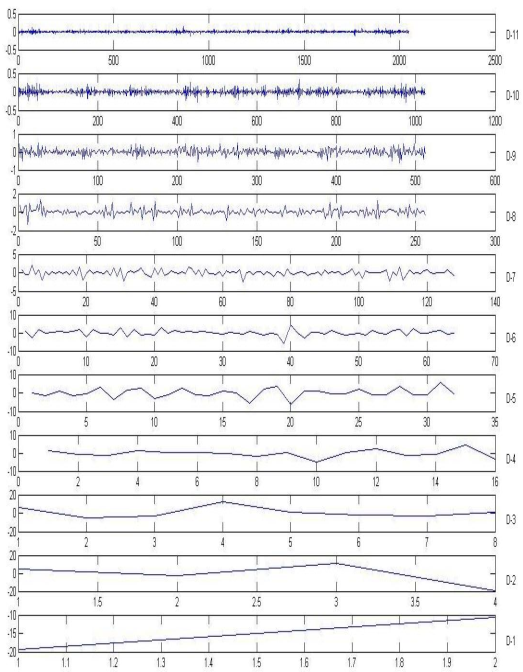

Fig.4: An example of the eleven levels decomposition.

IV. RESULTS AND CONCLUSIONS

We tested our classifier using an ECG record from MIT-BIH Arrhythmia Database and three recordings of typical noise in ambulatory ECG signals from MIT-BIH Noise Stress Test (NST) Database. The sampling frequency for all these signals is of 360 Hz, with a 11 bits/sample resolution, [2, 7].
An excellent environment for the simulation of signal processing methods based on wavelets is the Matlab toolbox WaveLab, which contains a collection of functions used in order to implement a variety of algorithms related to wavelet transform.
In [2] the classification algorithm was tested using ECGs corrupted only by one of the three types of noise.
This time two different types of noise have been applied at the same time on the ECG signal and the acquired signals are those presented in fig.5-8.
The three noise records used are: the baseline wander 'bw', the muscle artifact 'ma', and the electrode motion artifact 'em', while the ECG considered is the record '108'. For this record the subject is a women, aged 87 years old and the characteristics of the signal are described in [7]. The signals records have been converted into a numerical form before processing.

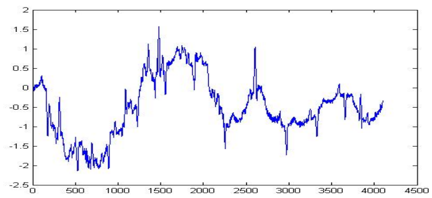

Fig.5: Signal ECG+bw+em

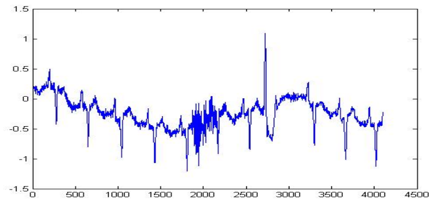

Fig.6: Signal ECG+bw+ma

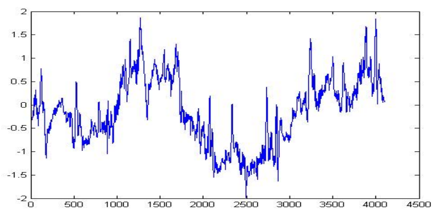

Fig.7: Signal ECG+em+ma

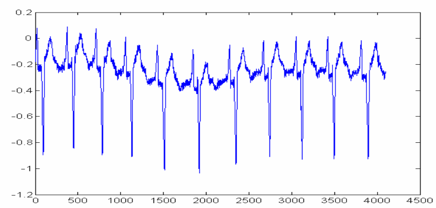

Fig.8: Signal ECG

The classification algorithm is based on the statistical analysis and the algorithm is described in [2]. We analyzed the figures we obtained applying the 'boxplot' function and we choose the signal's details from those levels that allowed us to differentiate the four signals. In this paper were considered two cases, but for both we used the same input signals, those presented in fig.5-8, with the order shown in fig.3.

1. In the first case the four classes are: class 1- ECG+bw+em, class 2- ECG+bw+ma, class 3- ECG+em+ma, class 4- ECG, and we choose the details from the third, the ninth and the seventh level of decomposition. The results are presented in table I.
We observe that choosing these four classes the percentage of the well classed signals is very good, thinking that we have 100 input signals and we obtain around 90 in the end.

2. In the second case were proposed the following four classes: class 1- ECG+bw, class 2- ECG+em, class 3- ECG+ma, class 4- ECG and we choose the details from the fifth, the ninth respective the third level of decomposition. The results are shown in table II.
Looking at the results we can observe that they are pretty good. Signal 3 (ECG+em+ma) belongs to the class 2 (ECG+em) and also to the class 3 (ECG+ma).
Signal 4 (ECG) has also a good classification, even if 10% are lost. The ECG record from MIT-BIH Arrhythmia Database is considered 'clean' meaning that it is used without applying an additional noise but in reality it is corrupted by different types of noise during the acquisition process, in this case muscular noise induced by the patient's movement.
Talking about the classification of signal 1 and signal 2 the classes are the good ones, but the classifier doesn't recognize the baseline wonder. For the signal 1, composed by ECG, 'bw' and 'em', we have the 10 % of incorrect classification in the class 3, meaning ECG perturbed by 'ma'.

TABLE I

CLASSIFICATION OF THE FOUR TYPES OF SIGNALS

ECG+bw+em	94%
ECG+bw+ma	95%
ECG+em+ma	80%
ECG	98%

TABLE II.

CLASSIFICATION OF THE FOUR TYPES OF SIGNALS

	1	2	3	4
Class 1	0	0	0	0
Class 2	90	0	20	0
Class 3	10	100	80	10
Class 4	0	0	0	90

We explain the presence of the 'ma' noise the same as before.
The signal 1 is composed by 'em' noise and 'bw' noise. The 'em' noise is generally considered the most troublesome type of noise while it can mimic the appearance of abnormal beats and cannot be removed easily using simple filters. So, the principal difficulty for the classification of the signal 1 comes from the fact that the proposed classifier doesn't recognize the 'bw' noise. The signal 2, is composed by a clean ECG, by 'bw' noise and by 'ma' noise. Muscle contraction noise causes artificial milivolt level potential to be generated and as baseline is usual in the microvolt range, it is usually insignificant. Therefore we have 100% ECG+ma. So, the difficulty for the classification of the signal 2 comes also from the fact that the proposed classifier doesn't recognize the 'bw' noise.
To conclude, we developed a structure of detection and classification, using four hundred signals, hundred for each type, and we obtained the signal's classification, meaning the percentage of the well classified or bad classified signals. Tested on real ECG signals affected by noise, the classification algorithm showed good results, assuring a good classification in the first case when the four types of signals and the four classes are similar. In the second case, when the signals must belong to two different classes, the classifier ignores the isoelectric line. So, there are two alternatives:
- to perform the denoising of the ECG eliminating the 'bw' noise, or
- a new research in order to find out if other ECG records from the same database determine the same behavior of the classifier.

REFERENCES

[1] J. D. Bronzino, The biomedical engineering handbook, Second Edition, vol. 1, Boca Raton: CRC Press LLC, 2000.
[2] Cristina Stolojescu, "Estimation of Noise in ECG Signals Using Wavelets"- Proceedings of the 11th International conference on Optimization of Electrical and Electronic Equipment OPTIM 2008,Vol.4, 22-24 May 2008, Brasov, Romania, pp.113-118.
[3] L.Senhadji, Approche multiresolution pour l'analyse de signaux non stationnaires, PhD Thesis, 1993,
[4] Stephane Mallat, A Wavelet Tour of Signal Processing, Second Edition, 1999,
[5] http://etd.lib.fsu.edu/theses/available/etd-11242003-185039/ unrestricted/09_ds_chapter2.pdf,
[6] Soumya Sakha Tripathy, System for Diagnosing Valvular Heart Disease using Heart Sounds, Master of Technology thesis, Department of Computer Science &Engineering, Indian Institute of Technology, 2005,
[7] http://www.physionet.org/physiobank/database/ mitdb/,
[8] Lindsay I Smith, A tutorial on Principal Components Analysis, February 26, 2002.
[9] Kadbi, M.H.; Hashemi, J.; Mohseni, H.R.; Maghsoudi, A. "Classification of ECG Arrhythmias Based on Statistical and Time-Frequency Features," Advances in Medical, Signal and Information Processing, 2006, pp.1-4,
[10] Prasad, G.K. Sahambi, J.S, "Classification of ECG arrhythmias using multi-resolution analysis and neural networks," TENCON 2003, vol.1, pp.227-231,