AN ARCHETYPAL BASED ECG ANALYSIS SYSTEM

Manuel Duarte Ortigueira[1]
mdo@eniac.inesc.pt

Nuno Roma[2]
nuno.roma@inesc.pt

Carlos Martins[3]
carm@eniac.inesc.pt

Moisés Piedade[4]
msp@inesc.pt

INESC - Rua Alves Redol, Nr. 9 - 2º
1000 - 029 Lisboa, PORTUGAL
Tel. +351 1 3100000, Fax. +351 1 3145843

Abstract

In this paper a description of a MatLab implementation of a new ECG analysis system based on the use of archetypal analysis for the study and interpretation of ECG signals is presented. The archetypal analysis is described together with new algorithms for the study of the heart frequency variability. Some features of the system are described and some results presented.

Keywords: ECG, Archetype, Heart Rate Variability, High Resolution Electrocardiography.

1 SUMMARY

In this paper a new High Definition ECG Modelling and Processing System is presented. The complex system formed by the Autonomous Nervous System (ANS) and the Heart is modelled as if it was a modulation system, where the first generates a signal that modulates a sequence of pulses which excite the heart. This decomposition corresponds to the usual study of the Heart Rate Variability (HRV) and the High Resolution Electrocardiography. However, the system uses analysis methods that are very different from the usual ones. In the case of the HRV study, the signal coming from the ANS is estimated and studied.

In the present study of heart modelling, a new technique in ECG processing, denominated by Archetypal Analysis [1][12][13], is introduced. This analysis allows the estimation of some archetypes (or prototypes) of the heart beats or, if required, of the P, QRS and T wavelets (see Figure 1). This is achieved through the use of a new algorithm for segmenting the ECG signal, which has been developed in [1]. These algorithms allow to get "clean" waves and good estimates of the "noise". In parallel, the Independent Component Analysis is being implemented, as well as decomposition with the Wavelet Transform.

The developed software has several outputs that correspond to estimates of signals involved in the system or variables obtained from them.

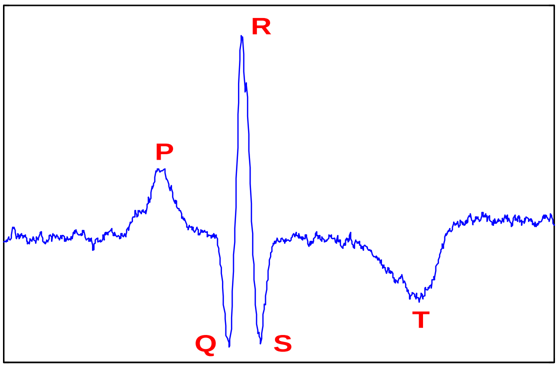

Figure 1 - Heart-beat and P-QRS-T wavelets.

2 CURRENT METHODOLOGY OF THE ECG ANALYSIS

The heart is a muscle composed by cells containing small filaments of actin and myosin. These proteins interact in the sense of forming *actomyosin* during muscle contraction, thus leading to the main purpose of the heart: pumping the blood through the circulatory system. Usually, the cardiac stimuli are originated at the *sinus* node. Pulses generated in this way are then transmitted through the *atrial* wall and septum to the atrio-ventricular node, where they are delayed approximately 0.1sec. Afterwards, they propagate rapidly through the His-Puriknje system, until they reach the non-specific muscle cells. Several pathological processes, causing disturbances at the atrio-ventricular transfer level, interruptions of the intraventricular conduction or the appearance of circular conduction of stimuli,

---

This work was supported by PRAXIS XXI under PSAU/0024/96 contract.

[1] Professor at Instituto Superior Técnico and at UNINOVA (mdo@uninova.pt).

[2] PhD Student at Instituto Superior Técnico.

[3] Professor at Escola Náutica Infante D. Henrique.

[4] Professor at Instituto Superior Técnico.

which may be responsible for dysrhithmic diseases, may compromise the transmission of these electric stimuli.

Superficial Electrocardiography captures the electric pulses through electrodes applied to the patient’s skin. The signal thus acquired enables us to obtain precious information concerning the nature of the cardiac pulses and their conductibility along the *atrium* and the *ventriculum*. It also permits the non-invasive and low-cost diagnosis of several pathological situations, which interfere in the nature and/or conductivity of *stimula* inside the heart, as well as its frequency. According to these procedures, Heart Rate Variability and Late Potentials are of special interest. Each beat originates a wavelet usually formed by 3 sub--waves: P, QRS and T.

Changes on heart rate occur as a result of the autonomous nervous system’s actions, through the Parasympathetic (P) and the Sympathetic (S) pathways, which have opposite influences. The study of these changes has contributed to detect modifications produced in several diseases in the HRV modulation by the two systems P and S, of the ECG. The P and S systems are mutually antagonised. The S stimulation leads to an increase on heart rate and the P stimulation does the opposite. These different actions result in fluctuations in the heart rate, the best known being the *sinus* respiratory arrhythmia, modulated by the P, and the ones related with the baroreflex action, modulated by the S. Through the use of an external marker on each heart beat, usually the R sub-wave of the QRS complex of the peripheral ECG, successive time intervals (RR intervals) are obtained. These intervals form a sequence, the HRV signal, which is studied using two different types of analysis: statistical and spectral [4]. It is currently known that the signal spectral components localised in the band of 0.15Hz-0.5Hz are due to the P action, and that the signal spectral components localised in the 0.04Hz to 0.15Hz band are influenced by the P and S actions together, although their relative influence is still unknown. The spectral components with frequencies lower than 0.04Hz, include, in the majority of the patients, more than 80% of the total power of the HRV signal. The S-P balance has a great relevance in the genesis of the cardiac arrhythmias, which are responsible for one of the most important causes of death. One of the main objectives of the HRV study will be the exact determination of the influence of the drugs in the P-S balance.

In the last years, the ventricular late potentials detection has been used to study the conduction disturbances in the cardiac ventricles. The ventricular late potentials are composed by high frequency and low amplitude signals that occur in the last portion of the QRS complex and/or in the beginning of the ST segment. It was postulated that these late potentials would constitute non-invasive markers of the presence of an arrhythmogenic substract, characterised by a slow and non-homogeneous propagation of the intraventricular activation wave. This is usually known as High-Resolution Electrocardiography **(**HR-ECG).

Due to the low resolution of the conventional electrocardiogram and the intense level of noise derived from the activity of other striated muscles (e.g. respiratory muscles) and other organs, the conventional ECG cannot record the stimulus of intraventricular conduction disturbances that generate micropotentials (lower than 0.1μV) and that make the intraventricular conduction wave turbulent. In order to enhance this very weak amplitude activity, it becomes necessary to use less conventional methods of electrical signal recording and analysis, which intend to reduce the noise and amplify the desired signal.

The most important methods used in the HR-ECG belong to three different categories, according with the work domain:

a) Time domain, where the most important method is the one based on mean estimation (averaging).
b) Frequency domain, using classical spectral analysis and AR/MEM.
c) Time-frequency space, based on Wigner and Gabor distributions.

Nowadays, the HR-ECG is performed through the use of signal mean value estimation techniques, based on multiple applications of average operations of the electrocardiographic signal [2][4]. This technique can be summarily described as a signal processing procedure, which consists in the averaging of succeeding segments of the signal, in order to increase the signal to noise ratio of the ECG, allowing the detection of the late ventricular potentials[5]. This method can also be viewed as a procedure that consists in computing the expected value of a stochastic process using the sum of several realisations of that process. The essential stages of this technique applied to the electrocardiographic signal are: acquisition; QRS complex detection and its correct alignment; estimation of the mean and measure of a realisation of the noise.

The estimation of the mean value operation is performed sequentially, QRS after QRS. To allow the correct application of the averaging method, each new QRS complex is detected through the selection of a fiducial (or reference) point. The objective is the precise alignment of all beats in

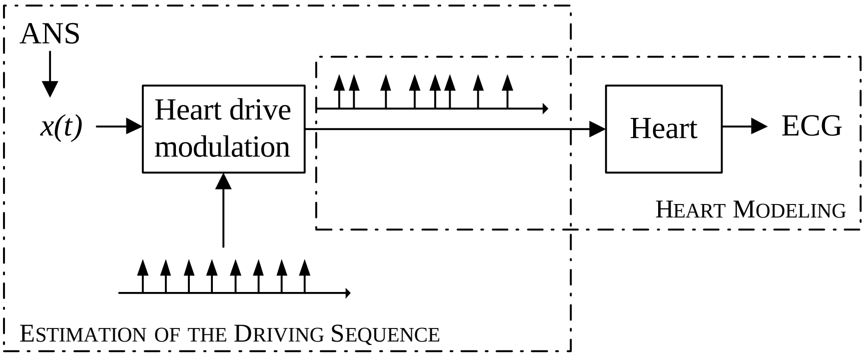

Figure 2 - Modelling of the ECG signal generator.

order to guaranty that their sum leads to a constructive interference. In practice, this procedure requires two steps:

1) Selection of a QRS complex as standard pattern.
2) Comparison of each complex with the standard one, with periodic update of this one (for example, every 8 beats).
3) Alignment, which is done by correlation. In this case, the instant time corresponding to the best alignment position serves as reference point for the measure of the RR interval.

The electrical signal generated by the heart, although not stationary, can, in some conditions, to be considered as such, allowing the use of usual methods of spectral analysis. The objective is to detect spectral components, especially those of very low frequency, which can reveal a risk to certain cardiovascular events. The signal is segmented in small portions and windowed, in order to select a QRS region of interest of analysis (usually the terminal QRS portion, where the ventricular late potentials can be found).

Recently, several papers were published describing and evaluating a new frequency domain analysis that considers each impulse as a whole and not as a set of arbitrarily identified portions. It was postulated that the pattern used to define arrhythmogenic abnormality would be composed by frequent and abrupt variations in the front wave velocity of the QRS signal. These are generated when the activation wave crosses areas with conduction abnormalities, resulting in a high degree of spectral turbulence.

The signal representation in the time-frequency space was already validated by several authors [6].

3 PROPOSED APPROACH

3.1 APPROACH OVERVIEW

Despite the effort expended to introduce new approaches in this field (successful in other areas), such as - Array Processing, Chaotic and Fractal Modelling and Wavelet Transform; the framework described in the previous section has been left unchanged in the last years [7].

The new approach, proposed in this paper, formulates the problem from a new point of view, which consists of a decomposition into two parts, as shown in Figure 2.

This decomposition corresponds, essentially, to the usual HRV and High-Resolution Electrocardiography studies. However, the analysis methods being used are very different from the

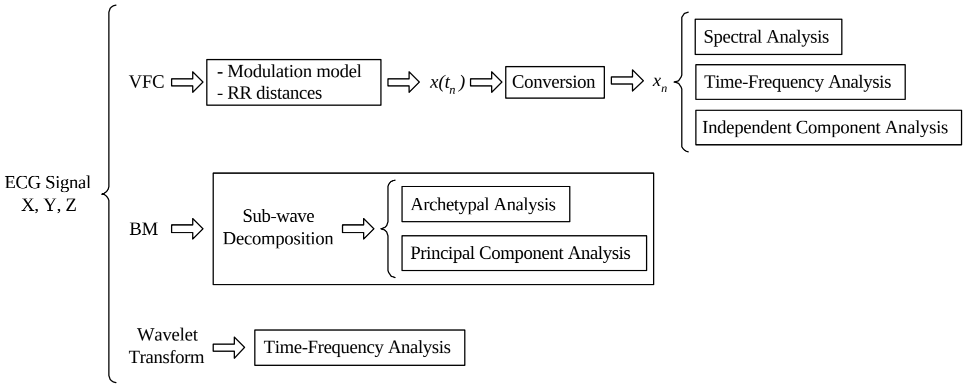

Figure 3 - Working scheme.

current ones.

3.2 WORKING SCHEME

The working scheme was based in the chart presented in Figure 3, where HRV represents the study of the variability of the heart rate and Beat Modelling (BM) the study and modelling of the heart beat.

The heart rate variability is studied from two different points of view: one is based in the use of RR distances, while the other uses several signals that result from the Archetypal Analysis. The study of the heart-beat is done by decomposing it into waves or sub-beats: archetypes and principal component. The Wavelet Transform allows a global insight of the ECG signal in the time-frequency space.

The software developed for the implementation of the above scheme is based in the following steps:

1 Reading of the ECG signal.
2 Wave isolation (beats or P, QRS and T sub-waves) using the developed algorithm.
    2.1 Use of the RR distances to obtain a signal proportional to the heart excitation signal generated by the Autonomous Nervous System (ANS).
    2.2 Study of the isolated beat sequence using Archetypal Analysis and Principal Component Analysis.
    2.3 Use of the Wavelet Transform to perform a global signal decomposition in the time-frequency space.

3.3 DEVELOPED ALGORITHMS

3.3.1 Beat isolation and positioning

From the above description it is evident that the isolation of successive beats and the establishment of a time reference are required. Two algorithms were developed to accomplish this task. The better one is based on a quadratic estimator [8]. It allows the transformation of the original signal into a new one, smoother than the original. With this signal, it is possible to separate the beats of an ECG signal (X, Y and Z leads) by defining the corresponding cut points. A cut point is the point that separates two consecutive beats. The main objective of this step is the determination of the cut point set for all the ECG beats or sub-beat waves.

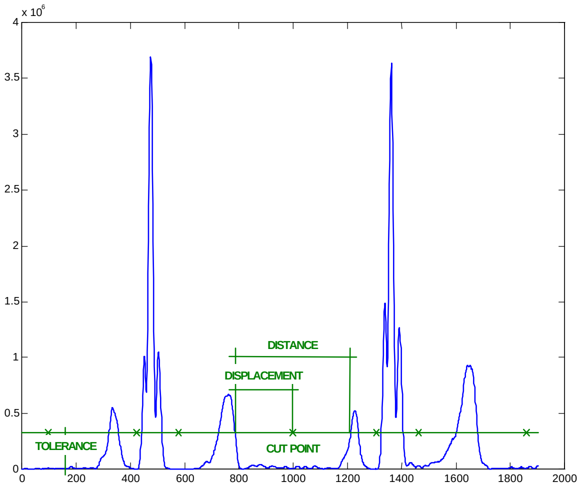

Figure 4 - Signal energy and cut points.

An initial guess to do this could be the minimum energy points. However, these are strongly affected by noise. Therefore, it has been decided to identify the points with energy above a given threshold, defined as a fraction of the energy maximum (see Figure 4). If the distance between two such consecutive points is larger than a given value, it has been considered that there is a cut point between them. This distance is also used to classify the separated waves.

To achieve a correct processing of the isolated waves it is important to normalise them. The first step of this task consists in normalising their length. It has been decided to take the longest wave as the reference. An alternative could be to take the average of all wave lengths. Hence, the cut points of the other waves have to be displaced to the left or to the right in order to obtain equal length waves. The second step consists in aligning the waves taking, again, the longest wave as the reference. This step can be done by using correlation functions. However, it has been decided to use the sum of absolute differences.

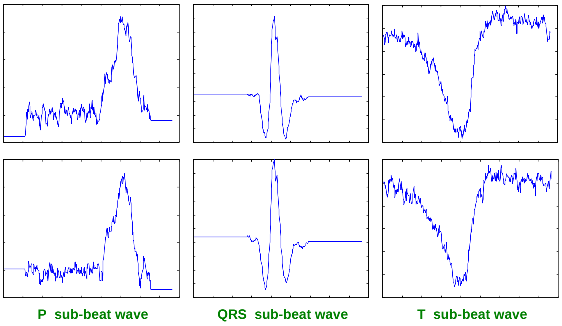

Figure 5 - Isolated and normalised waves.

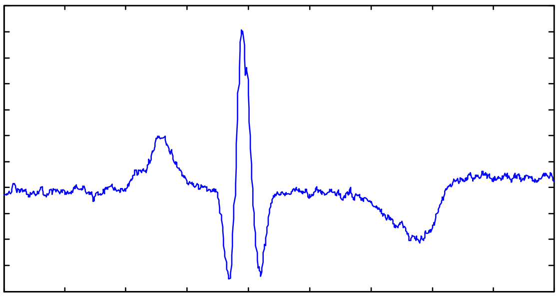

Figure 6 - A complete beat wave.

Figure 5 and 6 present some examples of wave isolation and segmentation of an ECG signal. Figure 5 presents two sets of sub-beat waves (P, QRS and T) and Figure 6 presents a complete beat. In a future work, an algorithm should be implemented to permit an automatic adjustment of the involved parameters, according to the particular characteristics of the signal being studied. The development of such algorithm is now in progress and will be presented in the very near future.

3.3.2 HRV Signal Analysis

The HRV signal is obtained from the RR distances. It has been assumed that the heart exciting signal (originated at the ANS) is proportional to the RR distances signal. This signal, *x(t)*, is known only at instants t_k, $k\,\hat{I}\,Z$, which means that the resulting discrete-time signal was obtained through a non-uniform sampling. As a consequence, the usual analysis methods can not be used without a suitable conversion [1]. To make this difference clearer, let d_n $(n=1, 2, \ldots, L)$ be a sequence of RR distances. Considering 0 as the time origin reference, we define a set of sampling instants: $t_n = t_{n-1} + d_n$, with $t_0=0$, $n=1, 2, \ldots, L$, and a signal, *x(t)*, which is proportional to the modulating signal. In the available commercial systems, sometimes the signal *x(t)* is treated as if it was obtained by uniform sampling. To demonstrate and illustrate this mistake, it has been used a signal d_n, obtained from an ECG signal and constructed using the sequence of instants t_n. After that, a sinusoid has been sampled at those instants and at a uniform time sequence nT, where the sampling interval T, is the mean value of d_n. The Fourier Transform (FT) of this pair of signals was computed by using the FFT algorithm. The obtained results are shown in Figure 7 (top and middle charts). It is clear that the use of the FFT algorithm to compute the Fourier Transform of non-uniformly sampled signals is incorrect. To avoid this problem, it has been assumed that the signal *x(t)* was sampled using a non-periodic ideal sampler, thus obtaining a signal with a Fourier Transform such as shown at the bottom of Figure 2, which evidences a good agreement with the middle chart of Figure 7. This fact means that the available approaches, used to study the variability of cardiac frequency, are intrinsically incorrect. The conversion from non-uniform to uniform sampling is done using the generalisation of the Shannon-Whitakker theorem, presented in [15]. However, before performing this conversion, it was necessary to remove the periodic components, to conform the signal with the requirements of the refereed theorem.

Once the refereed conversion has been performed, it is possible to apply the usual signal analysis techniques. At this stage, it has been decided to follow two different approaches:

a) Spectral Analysis and Time-Frequency Analysis.
b) Several analysis algorithms were developed in order to study the signal from a global (Spectral Analysis) and local (Time--Frequency Analysis and Sequential Spectral) points of view. Simultaneously, it was implemented an algebraic method for periodicity removal, which allows, for example, the elimination of the respiratory component without distorting the other components.

These sets of algorithms work in a parallel fashion. Other methods are being developed for other kind of analysis such as, chaotic and fractal detection methods.

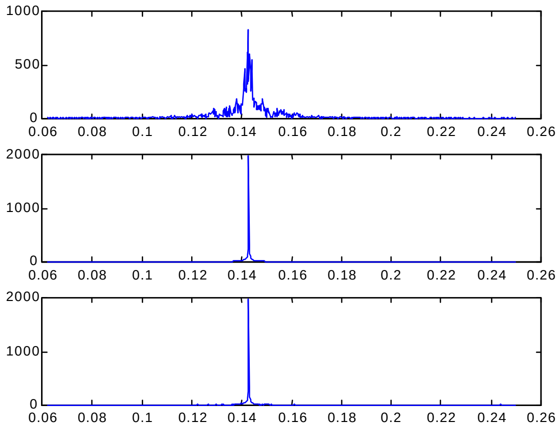

Figure 7 - FFT of a non-uniformly sampled sinusoid (top chart). FFT of a uniformly sampled sinusoid (middle chart). FFT of non-uniformly sampled sinusoid computed using the following equation: $\sum_{n=0}^{N-1} x(t_n).e^{-jwt_n}$ (bottom chart).

3.3.3 Beat modelling

The analysis of the heart-beats is accomplished by a new algorithm, based on Archetypal Analysis (AA). Essentially, it consists of a construction of several archetypes (or prototypes) of the beats [1][12][13], which are composed by weighted averages of the original beats. It is important to point out that this operation should be seen as a special "averaging" process, since it sets large weights to similar beats and small weights (eventually, zero) to the other beats. The weights (α) are computed by minimising the quadratic error, i.e., the square of the difference between the original signal and its archetypal reconstruction. This averaging procedure increases the signal to noise ratio. Furthermore, these weights give additional information about the heart excitation. It is possible to summarise the most important stages of this procedure in the following steps:

a) read the signal;
b) isolate the beats;
c) normalise the beats;
d) compute the Archetypes;
e) use the archetypes to reconstruct a "clean" ECG signal;
f) compute the signal error between the original ECG signal and the reconstructed one;
g) analyse and model each archetype;
h) study the weight sequences;
i) study the error signal.

Concerning the computational complexity, the most demanding stage of this algorithm is the weight computation step.

Once the archetypes have been computed, a synthetic "clean" version of the ECG signal is obtained. Then, an error signal is obtained by calculating the difference between the two signals, which although is a gaussian signal that seems to be also a chaotic signal. Several algorithms are being developed to test these hypotheses. The study of each archetype will supply an "image" of the heart.

Besides the AA approach, a Principal Component Analysis [14] was implemented, supplying additional data and the number of components of the signal.

3.3.4 Experimental Results

In this section, some preliminary results are shown. Figure 8 and 9 present two archetypes obtained from an ECG signal.

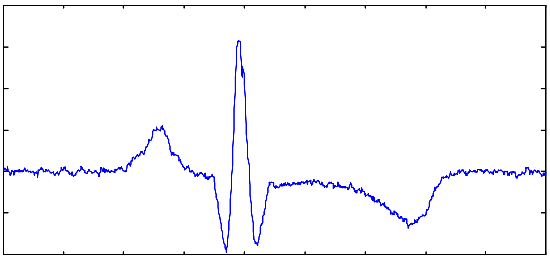

Figure 8 - Archetype I

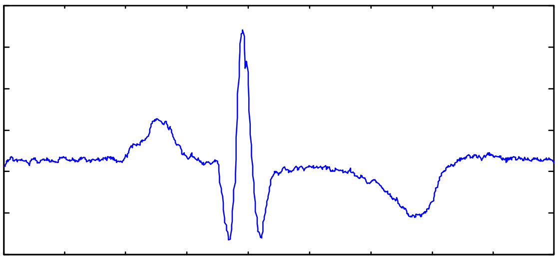

Figure 9 - Archetype II.

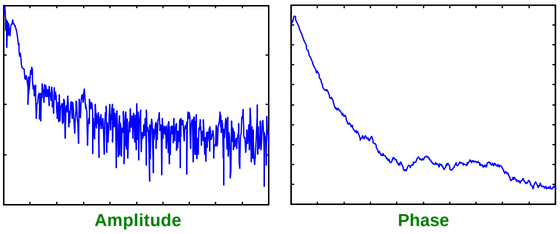

Figure 10 - Spectrum of Archetypal I.

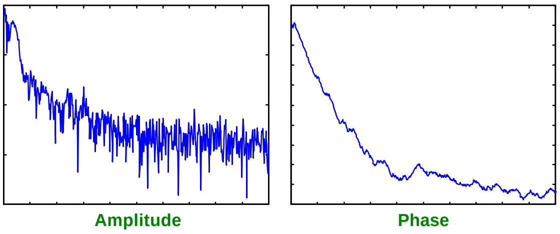

Figure 11 - Spectrum of Archetypal II

The corresponding Fourier Transform (amplitude and phase) is presented in Figure 10 and 11.

As it was mentioned before, since the archetypes are obtained by performing several averaging processes, it is possible to consider them as "clean" beat prototypes.

The difference between the original signal and the reconstructed one (see Figure 12) is the "noise" signal (see Figure 13).

In the several processing studies performed along this research, this signal has presented a set of characteristics typical of a broadband signal. However, a conjecture has been made supporting that it is a chaotic signal. This assumption is being object of further study.

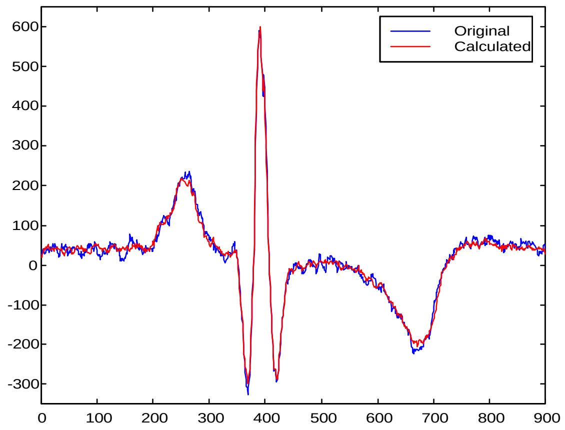

Figure 12 - Original and "clean" beats.

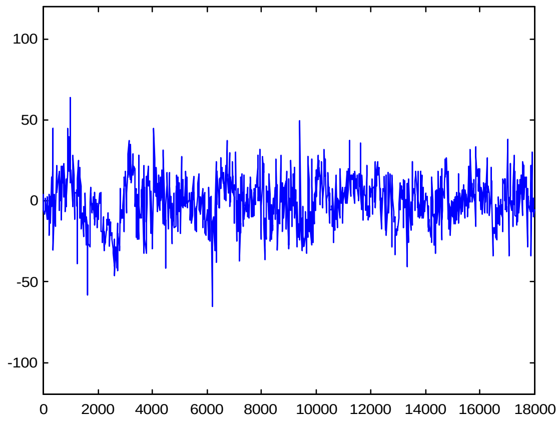

Figure 13 - Noise signal.

The plot of the histogram of this signal is presented in Figure 14. Power spectral density estimates are shown in Figure 15. Observing this chart, it is possible to realise the similarities between this signal and a gaussian signal.

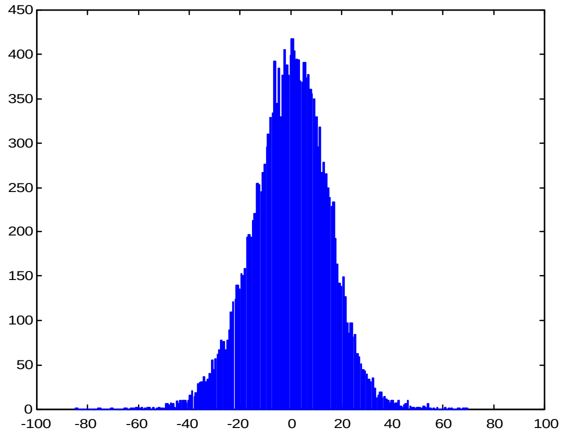

Figure 14 - Histogram of the noise signal.

To conclude the presentation of the application of AA to ECG, the weight sequence (α signals) is presented, which corresponds to the archetypals shown in Figure 8 and 9. These signals are shown in Figure 16 and 17.

From a spectral point of view these signals behave like white noise.

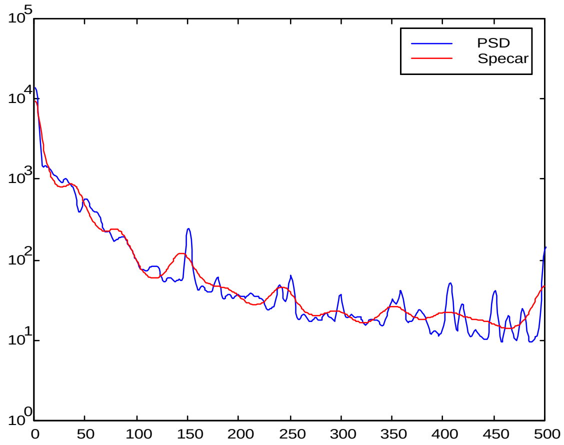

Figure 15 - Power spectral density of the noise signal computed using a Classic method and MEM.

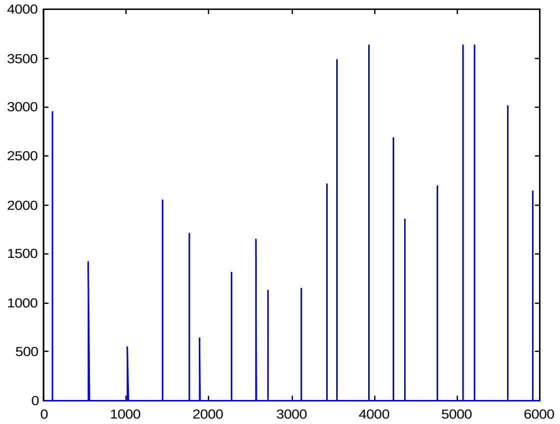

Figure 16 - α signal corresponding to archetypal I.

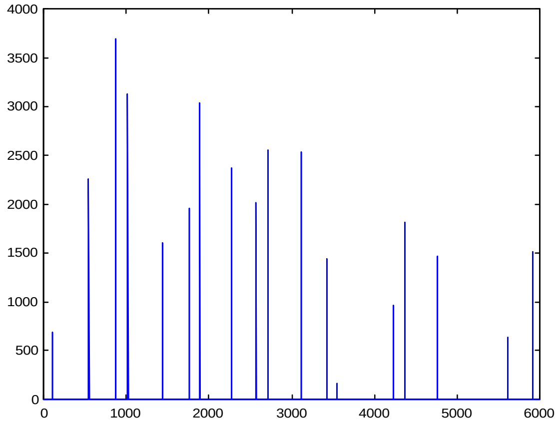

Figure 17 - α signal corresponding to archetypal II.

3.3.5 Future Work

All the referred analysis algorithms were already developed. However, additional work is required in order to optimise them and to minimise the associated computational burden. Furthermore, the different software blocks need to be integrated, allowing an easier data access and use.

In the future, it is proposed:

a) To explore all the features made accessible by the Archetypal, Independent, and Wavelet analysis.
b) To develop tests to identify some of the estimated signals in respect to: randomness (gaussianity); chaos and/or fractal characteristics and periodicity; and to compute characteristic parameters.
c) To establish the correlation between the signal and parameter estimates and cardiology anomalies.
d) To make the validation and to publicise the developed system.

4 MATLAB IMPLEMENTATION

The implementation of the described system was done using the Matlab simulation environment (version 5.2) and was denominated by "WinECG". Besides the basic Matlab toolboxes, the developed program makes use of two additional toolboxes: "Signal Processing Toolbox" and "Wavelet Toolbox".

The developed program was divided into three main blocks, performing the following processing functions: *"Heart Frequency Variability"*, *"HRV Signal Analysis"* and *"Wavelet Analysis"*. In the main window, it is asked the user to select the desired analysis (Figure 18).

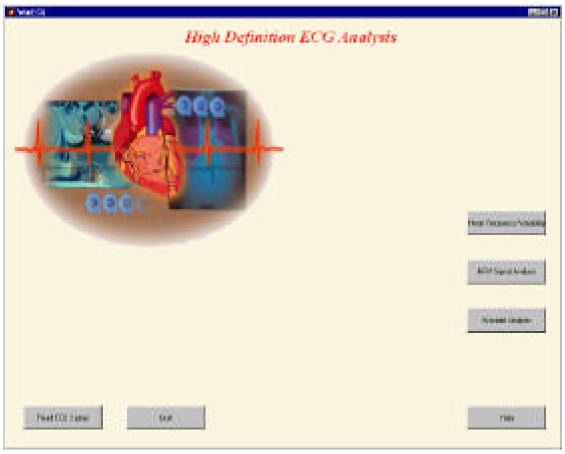

Figure 18 - Program WinECG - selection of the desired processing function.

The "Heart Frequency Variability" button enables the user to perform the study of the heart rate variability signal. With this functional block it is possible to perform several different analysis, both in the time and frequency domain (Figure 19).

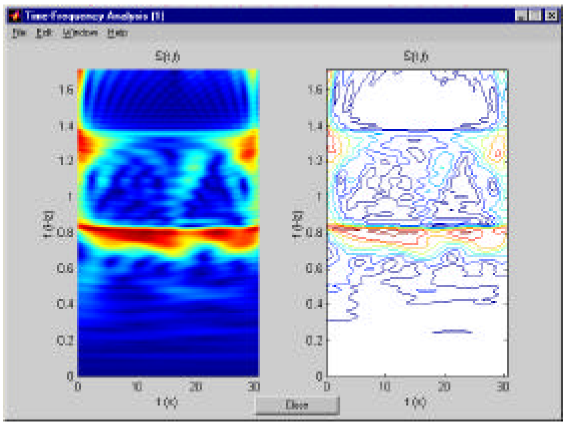

Figure 19 - Program WinECG - time-frequency analysis of the heart rate variability signal using the Shannon-Witakker method.

The "HRV Signal Analysis" button enables the user to proceed with the study of the beat modelling and heart rate variability of the current ECG signal. This functional block is responsible for performing the previously described Archetypal Analysis. It provides the user with many different types of information, such as, spectral analysis of the original and synthesised ECG signals (Figure 20), as well as the error resultant of the AA (Figure 21).

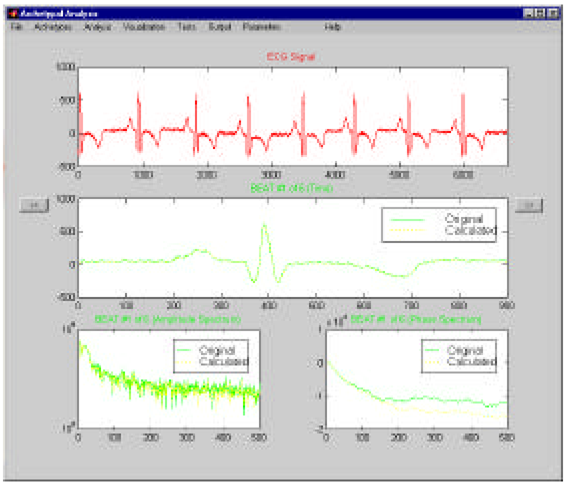

Figure 20 - Program WinECG - spectral analysis of the original and synthesized ECG signals.

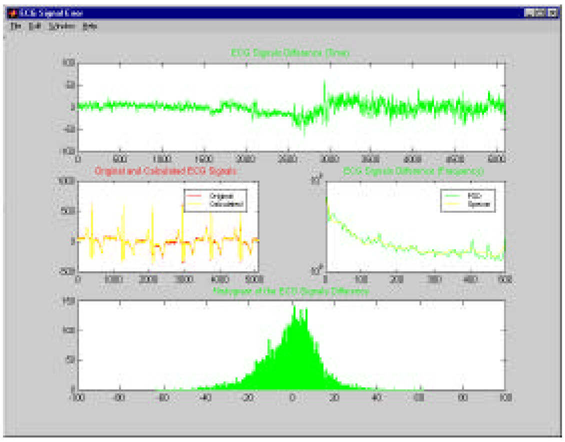

Figure 21 - Program WinECG - time-frequency study of the original and synthesized ECG signals, and of the resultant error signal.

The "Wavelet Analysis" functional block provides the user with several information resultant from the application of the Wavelet Transform. In Figure 22 it is depicted an example of the application of this functional block, using a 32 order filter and 5 decomposition levels.

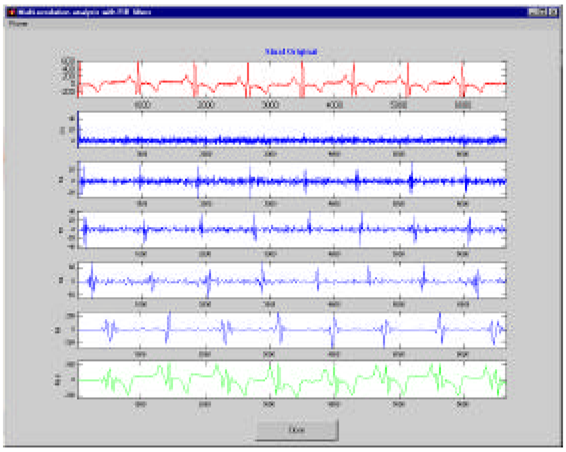

Figure 22 - Program WinECG - wavelet analysis function.

BIBLIOGRAPHY

[1] Ortigueira, M.D., “Archetypal ECG Analysis”, Proceedings of the 10th Portuguese Conference on Pattern Recognition, *Proc. RECPAD*'98, pp. 373-379, Lisboa, March 1998,.

[2] El-Sherif N, Turitto G (eds), *High Resolution Electrocardiography*, Mount Iksco, N.Y., Futura, 1992.

[3] Lander, P., Berbari, E. J., ”Principles and Signal Processing Techniques of the High--Resolution Electrocardiogram”, Progress in Cardiovascular Disease, 1992, vol. 35(3), pp. 169-188.

[4] Zimmermann, M., Adamec, R., Simonin, P. and Richez J., "Beat-to-beat Detection of Ventricular Late Potentials with High--Resolution Electrocardiography," *American Heart Journal*, Vol. 121, No. 2 part 1, pp 576--585, February 1991.

[5] Simson, M. B., “Use of Signals in the Terminal QRS Complex to Identify Patients with Ventricular Tachycardia after Myocardial Infarction”, *Circulation* 1981, Vol. 64, pp. 235-242.

[6] Buckingham, T. A., Greenwalt, R, Janosik, D. L. *et al*, “Three Dimensional Spectro--Temporal Mapping of the Signal-Averaged ECG Identifies Patients with Ventricular Tachycardia Despite the Presence of Bundle Branch Block (*Abst.*). J Am Coll Cardiol 1990, vol. 15, pp. 166A.

[7] "Bioengineering of Cardiovascular System", *Spring course*, *Corso Best*, pp. 25-30, Roma, Italy,March 1996.

[8] Fang, J. and Atlas, L. "Quadratic Detectors for Energy Estimation", *IEEE Trans. on Signal Processing*, Vol. 43, No. 11, November 1995.

[9] Soares, M.A. S. *Métodos de Análise Espectral e Sua Aplicação ao Estudo da Variabilidade da Frequência Cardíaca*, Master thesis, IST, September 1998, Lisboa, Portugal.

[10] Nóbrega, L.M.A.M., *Análise em Componentes Independentes*, Master thesis, IST, September 1998, Lisboa, Portugal.

[11] Comon, P., "Independent Component Analysis: A new concept?," *Signal Processing*, Vol. 36, pp. 287-314, 1994.

[12] Cutler, A. and Breiman, L. “Archetypal Analysis”, *Technometrics*, Vol. 36, No. 4, pp. 338-347, November 1994.

[13] Stone, E. and Cutler, A. “Introduction to Archetypal Analysis of Space-Temporal Dynamics”, *Physica D*, pp. 110-131, 1996.

[14] Vautard, R.,Yiou, P. and Ghil, M., "Singular--Spectrum Analysis: A Tool-Kit for Short, Noisy Chaotic Signals," *Physica D* 58, pp. 95--126, 1992.

[15] Ortigueira, M.D., "What about the FT of the comb signal”, *Proceedings of the IEEE* 1997 *Workshop on Sampling Theory and Applications*, *Sampta* 97, pp. 28-33, Aveiro, Portugal, June 1997.