Non-linear fit of a 6 skewed Gaussian template for filtering, compression and clinical identification of the ECG using a DSP wireless setup

Luís Pina Soares [1]*, Raul Carneiro Martins* [2]

[1] Instituto de Telecomunicações, Lisbon, Portugal, lmbpsoares@gmail.com
[2] Instituto de Telecomunicações, Lisbon, Portugal, rcmartins@ist.utl.pt

Abstract: In this paper we present an improved method for the acquisition, filtering, compression and clinical parameter identification of the ECG. The method involves the use of an improved instrumentation setup with high resolution and sample rate, as well as DSP monitored automatic gain and offset control (AGOC) which maximizes the range of the signal. It also involves the fitting of the signal with a nonlinear model consisting of six generalized Gaussians (with skew). The method further includes the use of sine-fitting techniques to selectively remove electromagnetic interference such as the power-line (50 Hz or 60 Hz component). We will further describe the experimental setup developed, which is a DSP based portable 3-lead device, and show some experimental and characterization results.

Keywords: ECG acquisition, clinical model template nonlinear fit, sine-fit.

1. INTRODUCTION

Most current filtering, compression and ECG wave detection schemes are relatively blind to the actual morphology of the wave. They rely on time-frequency schemes like filter banks [1], wavelets [2], clustering [3], singular value decomposition [4], stochastic methods with blind source separation [5, 6] and neural network approaches [7]. These methods use mostly the knowledge of the frequency band of interest disregarding the morphology of the ECG being also, to some extent, sensible to base-line wander, power-line interference and electromyograms (EMG). Adaptive filtering, like the methods proposed in [6, 8], require another reference signal or a generic model of the signal as an input which seldomly is available.

In this paper we propose the use of a nonlinear fit of one period of the ECG with a template of 6 generalized 4-parameter Gaussian functions. The nonlinear fit algorithm is based on the Marquardt-Levenberg compromise. This method allows for an extraction of some clinical information regarding the ECG in an optimized set of parameters describing the relative location, width, height and skew of each of the six features of the ECG (P, Q, R, S, T and U).

To maximize the efficiency of the fitting algorithm and to increase the signal-to-noise ratio (SNR) of the ECG we developed a portable, DSP based, wireless acquisition module which incorporates a DSP monitored automatic offset and gain control, and a 4 parameter sine-fitting algorithm for extraction of coherent electromagnetic interference. With this approach we are able to maximize the resolution of the ADC, being able to use up to 75 % of its input range, increase the signal gain and suppress both the base-line wander and power-line interference. We further use the most significant 15 of an 18-bit ADC (AD7678) at a high sampling rate of ~1 kSps to improve the SNR. The high sampling rate is important because the solution to the fit in the hyper-plane of the error gets narrower the more points the ECG has and upsampling is counterproductive since it increases the relative weight of noise.

2. METHODS

We will now present a detailed description of the electronic instrumentation setup with its AGOC system and the template fitting method.

2.1. Hardware

The electronic instrumentation consists in a portable, battery operated, DSP-based, wireless acquisition system whose block diagram is depicted in fig. 2.1 and where are clearly identified the amplification, sampling and processing and communication stages.
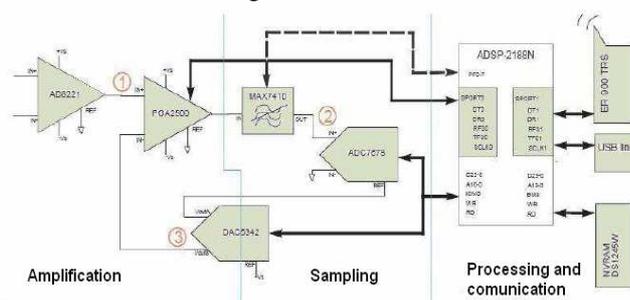

Fig. 2.1.   Acquisition system developed for the ECG signal recording.

Nowadays, the ECG systems commercially available present in the market use low levels of signal amplification, essentially to accommodate the heavy base-line wander inherent to cardiac signal. The fluctuation of the signal mean introduces problems in the signal measurement. Since these fluctuations might be large in amplitude and of random nature, the gain levels must be low when no type of method is used to compensate or suppress the phenomenon. This

reality is visible in fig. 2.2, which depicts the high resolution acquisition of a real
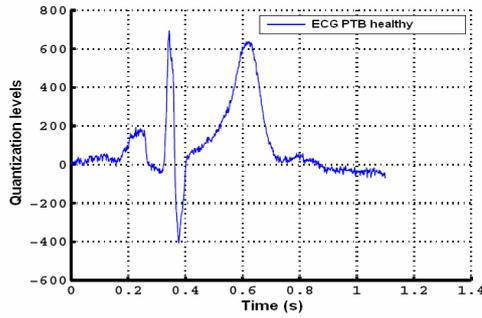

Fig. 2.2. Healthy ECG signal extracted from the MIT database, belonging to the PTB.

healthy cardiac signal. This signal was extracted from the PTB database (MIT-BIH) [9, 10], and acquired at 1kHz sampling frequency with a 16bit ADC. Careful analysis of the wave reveals that the ADC dynamic range put to use by the acquisition topology represents only 1% of the total dynamic range, which is clearly low.

The electronic platform implemented to acquire the signal and to perform some of the signal processing was devised so that it would be possible to substantially reduce the effects of the base-line wander. The control strategy used to achieve such a goal is depicted if figure 2.3. To control the
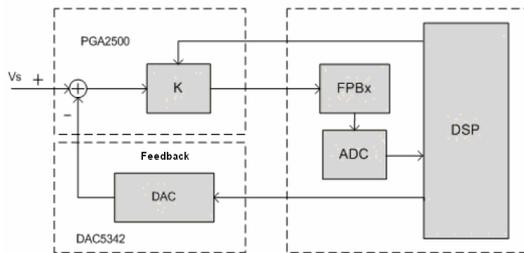

Fig. 2.3. Control topology devised for the base line wander compensation, used in the acquisition system.

mean fluctuation we introduced a feedback loop using an DAC (DAC5342). The feedback loop allows compensating the slow fluctuations of the signal, permitting the introduction of higher levels of amplification using a digitally programmed differential gain amplifier (PGA2500). This control structure, designated as AGOC, is
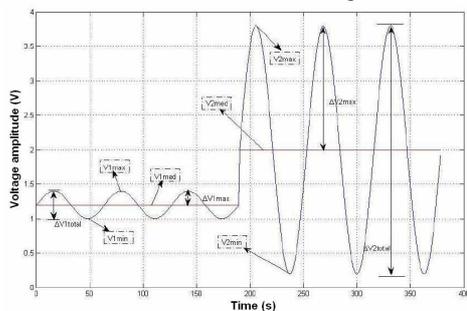

Fig. 2.4. Example of the AGOC system operation. Two waves are presented, the initial and the transformed one exhibiting better effective resolution.

possible thanks to the Digital Signal Processor (ADSP-

2188N), which executes a 2 step cyclic algorithm (depicted in table 2.1) that manages the gain and offset compensation values.

Effectively this control topology provides us with the means to define (as in figure 2.4 if one considers the dynamic range equal to the graph boundaries) the percentage of the total dynamic range that the signal will use and the signal placement within the total dynamic range. This adjustment capability brings obvious improvements to the system's effective resolution.

Table 2.1. Steps of the algorithm used to compensate the base-line wander, executed in the DSP.

Init. state: Calibration (standard conditions) process to make the first precise estimation of mean value using low values of gain.

Def. state: Definition of the desired dynamic range occupation and of the signal mean placement in the available dynamic range.

While (1)

1º - Estimate biological signal mean, considering present gain and compensation offset.

2º - Compute the future values of gain and compensation offset, respecting the parameters imposed in the definition state for the dynamic range.

2.2. Power-line interference removal using Sine-fitting

Very synthetically, the acquired data goes through a power-line identification and suppression algorithm which consists in an interpolated DFT of the data for accurate identification of the power-line frequency, 4 parameter non-linear sine-fitting for precise parameterization and time suppression using a reconstructed model of the interference.

In the interpolated DFT [11, 12], knowledge of the time window used allow us to fit the amplitude of the two bins nearest to the frequency to be identified with the analytical expression of the Fourier transform of the window. In our case the window is rectangular and the fit is fairly simple. Once the fit is achieved, the determination of the frequency becomes simply the value at which occurs the maximum of the function used for the fit. This frequency is used as a starting point for the iterative process of fitting the sine wave since the regression is not linear in the coefficients.

The "blind" path taken by the parameter vector in the solution space is governed by the Levenberg-Marquardt algorithm [14] which appears to provide a good compromise between convergence speed, robustness and final bias.

From the knowledge of the parameters the interference can be modelled in time and subtracted from the acquired signal.

Algebraically the regression model can be expressed as,

$$y = J^{(i)} \cdot \theta^{(i)} + \varepsilon^{(i)} \quad (1)$$
$$\underset{(n,1)}{} \quad \underset{(n,4)}{} \quad \underset{(4,1)}{} \quad \underset{(n,1)}{}$$

in which y is the vector with the acquired samples, θ is the parameter vector evaluated at iteration $^{(i)}$, ε the error and J is the Jacobian of the analytical model,

$$f\left(t,\theta\right) = \theta_1 \cos(\theta_3 \cdot t) + \theta_2 \sin\left(\theta_3 \cdot t\right) + \theta_4, \quad (2)$$

consisting of,

$$J = \frac{\partial f\left(t,\theta_1,\ldots,\theta_4\right)}{\partial\left(\theta_1,\ldots,\theta_4\right)} = \begin{bmatrix} \dfrac{\partial f\left(t_1\right)}{\partial\theta_1} & \cdots & \dfrac{\partial f\left(t_1\right)}{\partial\theta_4} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f\left(t_n\right)}{\partial\theta_1} & \cdots & \dfrac{\partial f\left(t_n\right)}{\partial\theta_4} \end{bmatrix}, \quad (3)$$

also evaluated at iteration $^{(i)}$.

In each iteration the increment vector for the parameters, δ_θ, is determined by,

$$\delta_\theta^{(i)} = \left(J^{T(i-1)} \cdot J^{(i-1)} + \xi I \right)^{-1} J^{T(i-1)} \underbrace{\left(y - f^{(i-1)}\left(t,\theta\right) \right)}_{\varepsilon^{(i-1)}}, (4)$$

where I is the eye matrix and ξ is the adaptive coefficient whose behaviour is determined by the algorithm in table 2.2. The parameter vector is updated through,

$$\theta^{(i+1)} = \theta^{(i)} + \delta_\theta^{(i)}. \quad (5)$$

Table 2.2. Steps of the Levenberg-Marquardt algorithm [14].

$\rho_\nabla = 0,25; \ \rho_G = 0,75; \ \xi = 1; \ k_1 = 2; \ k_2 = 3$
iteration (i)
 determine $J^{(i)}, \ \varepsilon^{(i)}$ e ρ_l
 $\rho_l < \rho_\nabla : \ \xi = k_1 \cdot \xi$
 $\rho_l > \rho_G : \ \xi = \xi / k_2$
 $\rho_l > 0 \ : \theta^{(i+1)} = \theta^{(i)} + \nu\left(\varepsilon^{(i)}\right)$

The suppression is achieved by subtracting an analytical replica of the power-line interference, obtaining a new sample vector,

$$y = y - J^{(f)} \cdot \theta^{(f)} = \varepsilon^{(f)} , \quad (6)$$

where $^{(f)}$ corresponds to the last iteration.

This technique of suppressing power-line interference is equivalent to the use of an infinitely narrow notch filter centered on the frequency of the mains.

2.3. Generalized Gaussian Template Non-linear fit

This method uses a multiresolution wavelet transform procedure [13] for identification of the R feature, localization in a continuous ECG, with local maximum search for better time resolution. This enables the segmentation of a continuous ECG signal vector in its periodic heart beats, in a way that will permit the use of the non-linear representation model.

The non-linear model consists of the sum of six generalized Gaussians, as expressed by,

$$T\left(\theta,t\right) = \sum_{n=1}^{6} G_g\left(\underbrace{q_{44}c_{42}d_{44}f_{43}}_{\theta}, t \right) . \quad (7)$$

Each of the ECG features is described by a generalized Gaussian exponential defined by,

$$G_g(a,\,d,e,f) = f \cdot e^{\left(\dfrac{\pi^2(t-a)^2}{d^2\left(\pi + 2\,\mathrm{atan}(-e(t-a))\right)^2} \right)}, \quad (8)$$

where f is the scaling factor, a is the time location, d is the aperture and e the skew of the Gaussian. The template defines those parameters individually for each of the 6 features. In this way the template can freely adapt to other morphologies of the ECG period that might depict pathologies. In fig. 2.5 can be seen the representation of a generic (simulated) heart beat using the non-linear model (template) for the coefficient values present in the table 2.3.

The non-linear model shown in fig. 2.6 is a template. The goal is to describe any heart beat period, healthy or unhealthy, using the given template. To describe the real data using the non-linear model, iterative methods are used. Specifically, the method used to compute the parameter values of the model, which minimizes the error between the adaptive template and the real signal we pretend to represent through the model, is the Levenberg-Marquardt Method [14], similar to what was done in section 2.2 for the power-line suppression.
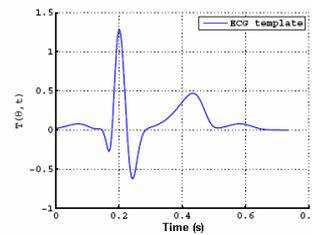

Table 2.3. Coefficients used to construct the simulated wave of an heart-beat.

Feature\Coef	a	c	d	f
P	0,200	8	0,0655	0,1195
Q	0,360	6	0,0210	-0,1235
R	0,400	0,0036	0,0306	0,7173
S	0,440	-0,0015	0,0254	-0,3964
T	0,680	10	0,0555	0,1795
U	0,880	2	0,0555	0,095

Fig. 2.5 Non-linear model created to represent a heart beat, using coefficients in table 2.3.
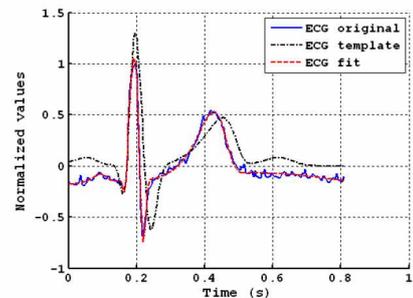

Fig. 2.6. An example of a nonlinear fit where is shown the acquired ECG (blue, solid line), the starting template (black, dash dot line) and the final fit (red, dashed line).

The iterative process uses as initialization the template mentioned above, described by the 24 parameters in table 2.3. Preceding this process, the template is scaled, dilated or contracted so that the initialization is the best possible, taking into account the previously assessed heartbeat interval. Considering $\underset{\sim}{\theta}^{(0)}$ the 24 parameter initialization, the iterative walk through the solution space is defined by the equation,

$$\underset{\%}{\theta}^{(i+1)} = \underset{\%}{\theta}^{(i)} + \underset{\%}{\nu}\left(\underset{\%}{\varepsilon}^{(i)}\right), \qquad (9)$$

where $\underset{\sim}{\varepsilon}^{(i)}$ defines the squared error linked to the iteration i and $\underset{\sim}{\nu}$ is the director function defined by,

$$\underset{\%}{\nu}\left(\underset{\%}{\varepsilon}^{(i)}\right) = \left(\underset{\%}{J}^{(i)^T} \cdot \underset{\%}{J}^{(i)} + \underset{\%}{\xi I}\right)^{-1} \cdot \underset{\%}{J}^{(i)^T} \cdot \underset{\%}{\varepsilon}^{(i)}, \quad (10)$$

being $\underset{\sim}{J}$ the Jacobian matrix of $T\left(\underset{\%}{\theta}, \underset{\%}{t}\right)$ (eq. 7) and ξ the adaptative coefficient whose behavior is determined by the algorithm in table 2.2. For the correct operation of the algorithm described in table 2.2 is also mandatory to compute the following linearity coefficient,

$$\rho_l = \frac{\varepsilon^{(i)^T} \cdot \varepsilon^{(i)} - \varepsilon^{(i+1)^T} \cdot \varepsilon^{(i+1)}}{\underset{\%}{\nu}\left(\varepsilon^{(i)}\right)^T \left[\underset{\%}{J}^{(i)^T} \cdot \varepsilon^{(i)} + \xi \cdot \underset{\%}{\nu}\left(\varepsilon^{(i)}\right)\right]}, \qquad (11)$$

which quantifies the quality of the linear approximation achieved in eq. (10). An example of a nonlinear fit can be seen in fig. 2.6 in which are represented the starting template, the acquired waveform and the final fit.

Clearly this approach will result in a very compact representation of the electrocardiogram since, for each period only 24 parameters are necessary. The level of compression in the representation is variable, depending on the sampling frequency of the signal acquisition system. Other interesting characteristic, intrinsic to the representation (error, residue), is the filtering obtained, which is depicted in the template, fig. 2.5.

The substantial shrinkage in the representation space of the electrocardiogram period has obvious advantages in the field of pattern recognition, since it eases the space of search and increases insensibility to parameters as sampling frequency.

An important question that might arise refers to the need of special templates for the cases not covered by the actual template (pathologies). As a matter of fact the solution to this problem lies on the robustness of the Levenberg-Marquardt algorithm. Since we are analysing signals composed by many cardiac periods, the 2 first periods of those signals can be used to create a new template. This means that, if one uses the previous non-linear fit as an initialization (template) to the following non-linear fit, after 2 periods a new template is created, which is more adequate for the subject under study. Genetic algorithms are also being considered for the estimation of the starting template to guarantee the global minimum is found.

3. Experimental Results

After the above explanation of the developed system hardware and signal processing algorithms we are now ready to analyze the extracted results.

Regarding the hardware experimental results the figure 3.1 depicts two different ECG signal periods, b) extracted from the high resolution ECG PTB [9, 10] database and a) extracted from one of the authors with the hardware setup implemented. Comparing the signals, and taking in consideration that the PTB system uses a 16 bit ADC to perform the signal acquisition it is possible to infer that the PTB signal only makes use of 1% of the total dynamic range whereas our signal employs 25% of the available dynamic range to represent the ECG signal.
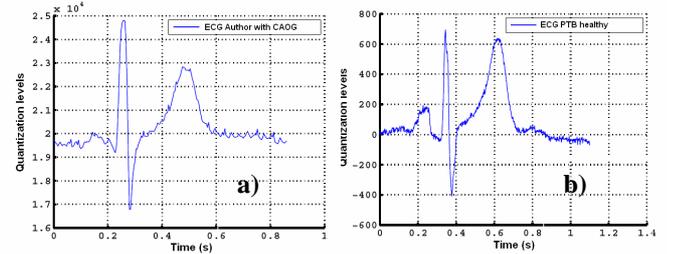

Fig. 3.1. Real Healthy ECG signals. Signal a) was acquired with the electronic setup described in section 2.1, with the AGOC connected and signal b) was extracted from the high reolution database PT B.

Analyzing a continuous acquisition of an ECG register one easily understands the mentioned problems caused by the base-line wander. In fig. 3.2 is possible to interpret two continuous registers. Signal a) belongs to one of the authors and was acquired with the AGOC system disconnected. Signal b) was extracted from a healthy patient present in the PTB database.
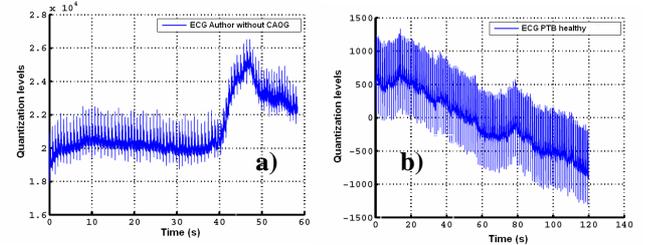

Fig. 3.2. Continuous acquisition of ECG signals. Signal a) belongs to one of the authors and it was acquired with the AGOC disconnected. Signal b) was extracted from PTB [14] database.

The behaviour of both signals is very similar, since the AGOC system is disconnected the signal acquired presents low levels of gain and no base-line wander compensation as like any of the PTB database signals.

When the AGOC system is used the signal acquired presents a level of effective resolution very high comparatively with the PTB high resolution signals. In figure 3.3 a) one can observe a continuous register presenting a controlled base-line wander and amplification level. In figure 3.3 b) is depicted a 10 second ECG signal (detail from figure 3.3 a) from the author. The ellipses exhibit clear cases of the offset compensation performed by the AGOC system.
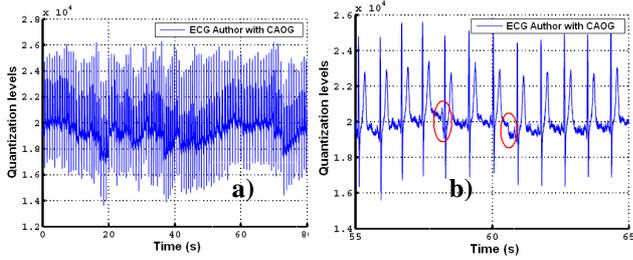
Fig. 3.3. Continuous ECG signal from the author, with AGOC connected a). Signal b) is a Zoom from signal a), where you can see in red circle the offset compensation working.

In terms of algorithmic work we attained very promising results. In figure 3.4 it's possible to analyze the non-linear fitting algorithm applied to two different ECG signal periods. Figure 3.4 a) and c) presents the original signal (solid line) we pretend to represent and the ECG template (dashed line), algorithm initialization. In figure 3.4 b) and d) one can verify the results of the non-linear fit, in solid line the original signal and in a dashed line is the resulting fit. In this case the original signal presents an offset relatively to the non-linear fit to improve the results visualization.
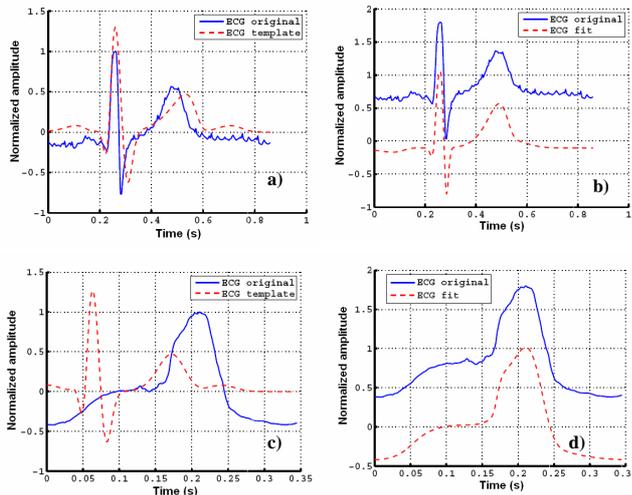

Fig. 3.4. Non-linear fits performed to real signals, the top row is an author's signal and the bottom row presents a bundle branch block signal. In figures a) and c) one has in dashed line the algorithm initialization and in continuous line the original wave. Figures b) and d) present the fit results in dashed line and the original in continuous, which is scaled for better visualization.

Analyzing the fitted signal it is obvious that a very selective filtering has been performed, and considering that for each period only 24 coefficients are required, an high compression can be achieved. This corresponds to a lossless compression and filtering, assuming that all the relevant information in a period is given by these features. The level of compression is obviously dependent on the sample frequency of the acquired signal. For example, if the signal sampling frequency is 1kHz (high resolution signal) and the cardiac period is 1s, the rate of compression is 20. In the fit case of figure 3.4 c) d) one can verify the robustness of the non-linear fit method, since the template used as initialization is completely inappropriate. Yet, the algorithm still manages to make an extremely good representation of the original pathological signal.

In figure 3.5 a) b) c) and d), are depicted the results from the application of the non-linear fit algorithm to continuous ECG record. After applying the segmentation process that allows for an accurate identification of the ECG R complexes, the non-linear fit is applied to each of the isolated signal periods. The outcome is a smoothed continuous ECG record with good levels of compression.
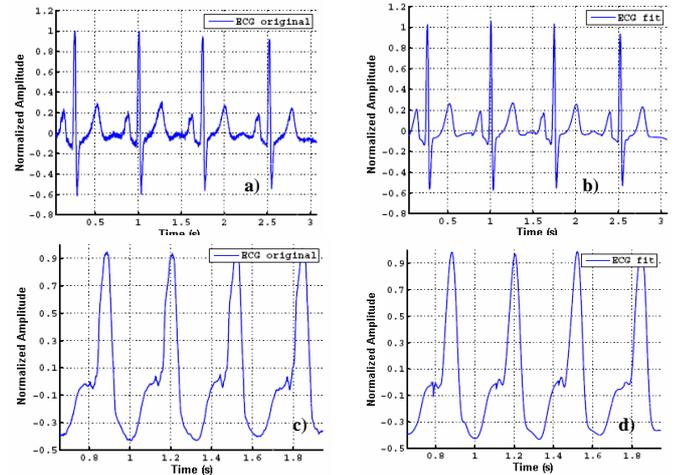

Fig. 3.5. First column (left a) and c)) depicts two signals (4 periods) taken from the PTB [13, 14] database, the top one is healthy and the other is a disease, bundle branch block. The right column is the result of applying the non-linear fitting algorithm to the PTB real signals (left column).

An electrical characterization of the acquisition channel was performed in which the total harmonic distortion plus noise (SINAD) was found to be 89.3 dB, close to the theoretically maximum of 92 dB for a 15 bit converter. The phase response was found to be linear from DC to 450 Hz.

4. CONCLUSIONS

The work presented here relates to two major areas, electronics and signal processing. In each of them is presented an original contribution.

The first novelty is associated with the experimental platform designed to acquire the ECG signal. This platform presents an alteration relatively to the common ones, a feedback loop that allows for the compensation of the signal base-line wander. This fact enables the use of higher levels of amplification in the amplification stage, avoiding saturation and allowing for an acquisition with enhanced effective resolution.

In the field of signal processing a method for signal representation and morphological identification was devised. This method makes use of 6 generalized Gaussians to represent each of the ECG period relevant features. To find the specific set of parameters for a given ECG period a non-linear fit based on the Marquardt-Levenberg algorithm is executed. This process presents signal filtering, good compression levels and an elegant way to model the ECG signal.

The reduced parameter space is ideally suited for pattern classification applications. There are, however, still some problems concerning the non-linear fit algorithm since in some cases the Gaussians shift amongst them, hampering the discrimination capabilities of a pattern classification system.

ACKNOWLEDGMENTS

REFERENCES

[1] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 192-202, 1999.

[2] H. Dickhaus and H. Heinrich, "Analysis of ECG late potentials using time-frequency methods," in *Time Frequency and Wavelets in biomedical signal Processing*, M. Akay Ed., IEEE Press, 1998.

[3] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo, "Clustering ECG complexes using Hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 838-848, 2000.

[4] Jyh-Jong Wei, Chuang-Jan Chang, Nai-Kuan Chou, Gwo-Jen Jan, "ECG data compression using truncated singular value decomposition," *IEEE Trans. on Information Technology in Biomedicine*, vol.5, no.4, pp.290-299, Dec 2001.

[5] M. Potter, W. Kinsner, "Competing ICA techniques in biomedical signal analysis," Proceedings of the Canadian Conference on Electrical and Computer Engineering, vol. 2, pp.987-992, 2001.

[6] Barros A, Mansour A, Ohnishi N., "Removing artifacts from ECG signals using independent components analysis," *Neurocomputing*, vol. 22, pp. 173–18, 1998.

[7] M. Strintzis, X. Magnisalis, G. Stalidis, and N. Maglaveras, "Use of neural networks for electrocardiogram (ECG) feature extraction recognition and classification," *Neural Network World J.*, vols. 3–4, pp. 313–327, 1992.

[8] Rangaraj M. Rangayyan, "Biomedical Signal Analysis; A case study approach," IEEE Press, John Wiley & Sons, pp. 137–176, 2002.

[9] Physikalisch-Technische Bundesanstalt, the National Metrology institute of Germany
http://www.physionet.org/physiobank/database/ptbdb

[10] MIT-BIH Database Distribution. Physiobank archive index. http://www.physionet.org/physiobank/database/#ecg.

[11] D. C. Rife and G. A. Vincent, "Use of the discrete Fourier transform in the measurement of frequencies and levels of tones," *Bell Syst. Tech. J.,* vol. 49, pp. 197-228, 1970.

[12] D. Agrez, "Frequency Estimation of the Non-Stationary Signals Using Interpolated DFT," *Proceedings of IMTC,* vol. 2, pp. 925-930, 2002.

[13] S. Thurner, M. C. Feuerstein, and M. C. Teich, "Multiresolution wavelet analysis of heartbeat intervals discriminates healthy patients from those with cardiac pathology," *Phys. Rev. Lett.*, vol. 80, no. 7, pp. 1544-1547, 1998.

[14] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, pp. 164–168, 1944.