Parametric linear and non-linear modeling techniques for estimating abnormal intra-QRS potentials in the high resolution ECG

P. Gomis,[a] P. Lander,[b] P. Caminal[c]
[a]*Universidad Simón Bolívar, Departamento de Tecnología Industrial, Camurí Grande, P.O. Box 314, La Guaira, Venezuela*
[b]*University of Oklahoma Health Science Center, Oklahoma City, USA*
[c]*Institut de Cibernètica, Barcelona, Spain*

Abstract

This paper presents a modeling procedure to measuring abnormal *intra*-QRS signals in the high resolution electrocardiogram (HRECG). Abnormal *intra*-QRS potentials (AIQP) are defined as very low-level notches and slurs in the high resolution QRS complex. AIQP are isolated using (i) parametric linear and (ii) non-linear modeling techniques for representing the predictable, smooth normal, part of the QRS from the original waveform. The first approach considers HRECG signals as time domain phenomena, although using the frequency domain for computational convenience. A discrete cosine transform (DCT) of the QRS complex is first obtained, in order to form an energy compacted description of the signal. The DCT of the QRS is considered as the impulse response of a linear model, estimated with ARX (autoregressive model with exogenous input) and OE (output error) structures. The second approach uses a non-linear ARX (NARX) structure, parametrizated by an artificial neural network, for estimating the smooth normal part of the QRS. AIQP were quantified using the residual of the modeling procedure. The problem of model order selection of the linear models and neural networks structure were solved empirically, in a data-dependent manner, using a training dataset. The modeling processes are capable of separating relatively predictable (normal, smooth) and unpredictable (abnormal) components of the HRECG. A clinical study strongly suggests that AIQP amplitudes are significantly greater in patients with ventricular tachycardia than in those with no-events. AIQP can be used as a new predictive index of arrhythmic events.

1 Introduction

The high resolution electrocardiogram (HRECG) is a noninvasive technique used to detect low amplitude and high frequency cardiac signals, not observable in the standard electrocardiogram. The main application of the HRECG has been the detection of ventricular late potentials in the terminal QRS complex

that correlate with reentrant ventricular tachycardia (VT) following myocardial infarction (MI) [1]. Late potentials (LP) represent activation of injured but viable cells within and surrounding infarct regions. They are defined as abnormal signals that outlast the normal QRS period during normal sinus rhythm. Abnormal signals resulting from disruptions of ventricular activation are present within the QRS, but are detected and interpreted only in the terminal portion. LP have been commonly characterized in the time domain by measurement of total QRS duration (QRSD) and root mean squared (RMS) amplitude of the terminal 40 ms of the filtered QRS (RMS40) [2]. Several studies suggest that the reentrant activity may be wholly contained within the normal duration of the QRS [3][4]. In these cases a VT reentry circuit can not be detected. On the other hand, late potentials are not necessarily a marker of a VT reentry circuit [1]. This lack of specificity produces a low positive predictive value for VT detection (10-25%) in studies of post MI patients.

These limitations have motivated the idea of characterizing abnormal signals linked with VT reentry circuits occurring anywhere within the period of ventricular activation. Several attempts have been made to analyze abnormal *intra*-QRS signals in the frequency domain and time-frequency plane [5]-[6]. These approaches have implicit limitations due to spectral resolution and technique-induced artifacts. We have recently proposed the concept of *abnormal intra-QRS potentials* (AIQP) with the purpose of enhancing the predictive value of the HRECG for arrhythmic events [7]. AIQP are defined as very low-level notches and slurs in the high resolution QRS complex. These potentials have long been associated with scarring from myocardial infarction [8], but not with arrhythmogenesis. However, like conventional late potentials, AIQP arise from myocardial infarct regions of scarring and are a potential marker of reentry. The pathophysiological basis of AIQP and their relationship to arrhythmogenesis has recently been studied [9].

Our objective is to characterize abnormal intra-QRS potentials in the time domain for extracting new indices of arrhythmic events. Section 2 describes two approaches for modeling the predictable, smooth normal, part of the QRS for estimating AIQP. The first ones uses parametric linear structures. In the second approach, the modeling QRS is obtained through non-linear structures parametrized by feedforward neural networks. The performance of the methods is considered in Section 3, where AIQP indices are applied to a population including VT and nonevent subjects.

2 Methods

2.1 Data Acquisition and High Resolution ECG Analysis

A dataset consisting of 16 VT and 24 no-event subjects, from a previously analyzed dataset [10], were used to develop parametric modeling techniques as described in the next sections. The high resolution ECGs were recorded using

orthogonal XYZ leads [2] with signal-averaged techniques performed with the Predictor system (Corazonix Corp., Oklahoma City). Details of the acquisition and processing protocol have been published elsewhere [2][10].

Abnormal intra-QRS potentials indices were estimated from each individual unfiltered, high resolution XYZ lead. In addition, two standard time-domain indices were calculated from the bi-directional filtered vector magnitude: the total QRS duration (QRSD) and the root mean squared amplitude of the terminal 40 ms of the QRS (RMS40).

2.2 Parametric Modeling of the HRECG

Abnormal intra-QRS potentials were calculated by the residual of parametric modeling processes. Each individual-lead HRECG QRS complex is presented unfiltered to be mathematically modeled. These signals can be considered as the response of the cardiac system to an stimulus. Therefore, computer acquired QRS waveforms modeling can be based on the identification of a discrete-time system. Considering an input stimulus *u(t)* and an output (QRS waveform) *y(t)* observed at sampling instants $t = 1,2,\ldots,N$, a general model of a discrete-time dynamic system can be expressed as

$$y(t) = f(y(t-1), y(t-2), \ldots, u(t), u(t-1), \ldots) + \nu(t) = f(\varphi(t), \theta) + \nu(t) \quad (1)$$

where $f(\cdot)$ is some linear or nonlinear function; the additive term $\nu(t)$ represents the fact that the output $y(t)$ is not an exact function of past observed data. The function $f(\cdot)$ represents a general regression model structure, and it is expressed as a function of a finite-dimensional parameter vector θ and a finite-dimensional *regression* vector $\varphi(t)$,

$$\varphi(t) = \varphi(y(t-1), \ldots, y(t-n_y), u(t), \ldots, u(t-n_u)) \quad (2)$$

where n_y and n_u are the maximum lags considered for the output and input respectively. Parameters are selected from $\theta = \hat{\theta}_N$ so that a loss function V_N is minimized in terms of the error between the model and the observed data:

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta) = \arg \min_{\theta} \frac{1}{N} \sum_{t=1}^{N} (y(t) - f(\varphi(t), \theta))^2 \quad (3)$$

where arg min denotes the minimizing argument. Our work will be choosing appropriate model structures $f(\cdot)$ and estimating procedures for adjusting the predictive QRS waveform $\hat{y}(t)$,

$$\hat{y}(t) = f(\varphi(t), \hat{\theta}_N) \quad (4)$$

to the normal smooth part of the QRS complex. The modeling procedures isolates the predictable part of the QRS from the original waveform. Then, the modeled QRS signal is subtracted from the original data. The difference is a

residual signal which represents the part that could not be modeled. This signal represents the abnormal intra-QRS potentials. AIQPs were quantified by computing the RMS amplitude between the QRS limits. Only an approximate knowledge of the QRS onset and offset is needed.

2.3 Linear Modeling structures

Linear model structures constitute the most common class of modeling a discrete-time dynamic system. Previous attempts to approximate a time domain ECG signal, as an impulse response of a pole-zero model, have lead to system structures of very high model order [11]. To overcome this limitation, the original HRECG is pre-processed with the discrete cosine transform (DCT) prior to the modeling procedure. This energy compacting transformation produces damped cosinusoid waveshapes well suited for estimation by linear modeling structures. The DCT of a discrete signal $x(t)$ of N samples (for $t = 0,1,\ldots,N-1$) is defined as

$$X(k) = \sqrt{\frac{2}{N}} C_k \sum_{t=0}^{N-1} x(t) \cos\left[\frac{(2t+1)\pi}{2N}\right] \qquad k = 0,1,\ldots,N-1 \qquad (5)$$

where $C_k = 1/\sqrt{2}$, for $k = 0$ and $C_k = 1$, for $k > 0$. Once the modeled QRS is obtained, it is restored to the time domain, applying the inverse DCT .

A general, linear, time-discrete parametric model can be written as [12]

$$y(t) = G(q,\theta)u(t) + \nu(t) \qquad (6)$$

where $G(q,\theta)$ represents a rational transfer function in the backward shift operator q, i.e. $q^{-1}y(t) = y(t-1)$, such that

$$G(q,\theta) = \frac{B(q)}{A(q)} = \frac{b_0 + b_1 q^{-1} + \cdots + b_{nu} q^{-nu}}{1 + a_1 q^{-1} + \cdots + a_{ny} q^{-ny}} \qquad (7)$$

The unpredictable term $\nu(t)$ can also be expressed as a function of a rational transfer operator $H(q,\theta)$ and some white noise sequence $e(t)$: $\nu(t) = H(q,\theta)e(t)$, where $H(q,\theta)=C(q)/D(q)$. Several linear structures are used in practice, where the main problems consist in selecting an efficient estimation technique and choosing a correct model order.

2.3.1 ARX modeling

The most simple linear modeling is the autoregressive model with an exogenous input (ARX). This structure uses the same denominator for $G(q)$ and $H(q)$: $A(q)=D(q)$, and has $C(q) = 1$, that is : $A(q)y(t) = B(q)u(t) + e(t)$, or

$$y(t) = -\sum_{i=1}^{ny} a_i y(t-i) + \sum_{j=0}^{nu} b_j u(t-j) + e(t) = \varphi^T(t)\theta + e(t) \qquad (8)$$

where $\theta = \left[a_1 \cdots a_{ny}\ b_0 \cdots b_{nu}\right]^T$, and $\varphi(t) = \left[-y(t-1) \cdots -y(t-ny)\ u(t) \cdots u(t-nu)\right]^T$. This structure has the advantage that the predicted signal is defined by a linear regression, $\hat{y}(t) = \varphi^T(t)\theta$, and its parameters can be estimated analytically, minimizing the loss function described in equation (3),

$$V_{ARX}(\theta) = \sum_{n=p+1}^{N}\left[y(n) - \varphi^T(n)\theta\right]^2 \tag{9}$$

where p is the greater value between ny and nu. The data of N samples include a pre-windowing of p zeros.

2.3.2 OE modeling

The second linear model used was the output error (OE) structure. This model is described by a linear relationship between the non-disturbed output of the system, $y_{nd}(t)$, and the input. The output signal is obtained by adding $y_{nd}(t)$ to a white noise sequence $e(t)$ $(H(q)=1)$, i.e.

$$A(q)y_{nd}(t) = B(q)u(t) \tag{10}$$

$$y(t) = y_{nd}(t) + e(t) \tag{11}$$

This structure represents a special case of the general linear model structure of eq. (6), where the noise sequence, $e(t) = v(t)$, will then be the difference (error) between the observed and the non-disturbed output. The parameters of $A(q)$ and $B(q)$, in this structure, cannot be minimized by analytical methods. They were estimated using iterative optimization techniques for minimizing

$$V_{OE}(\theta) = \sum_{n=p+1}^{N}\varepsilon_{OE}^2(n) = \sum_{n=p+1}^{N}\left[y(n) - \frac{\hat{B}(q)}{\hat{A}(q)}u(n)\right]^2 \tag{12}$$

A Gauss-Newton algorithm, as described in the next section, was used to search iteratively for the parameters values [12].

2.3.3 Model order selection

The fit of the normal part of the QRS is constrained by the chosen model order. Lead-specific model order were used. A low $A(q)$ and $B(q)$ polynomial order extracts notches and slurs, but it also causes normal components of the QRS appear in the residual. An optimal model order, according to Akaike's information criteria (AIC) [13], accurately models both the normal and abnormal components of the QRS and overestimates the necessary number of parameters. Model order were selected empirically using a training dataset from the 2 groups of patients: 10 VT and 10 no-event subjects were randomly chosen. For each lead independently, the model which maximized the ratio of mean AIQP amplitudes between the VT and no-event groups were chosen. Amplitudes of AIQP were found to be significantly higher in the VT than in the no-event training group. Each lead-specific model order selected is given in table I. These model orders were used for every patient of the complete dataset.

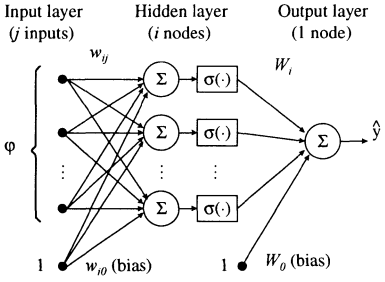

Figure 1: Feedforward network with one hidden layer

2.4 Non-Linear Modeling of the HRECG using Neural Networks

It has been proved that any continuous non-linear function can be approximated by a multilayer feedforward neural network with sufficient nodes in the hidden layers [14][15]. This has motivated the use of a non-linear model structure based on neural networks for modeling HRECG signals. Following the nomenclature for linear models, several non-linear modeling structures have been developed for representing the function $f(\cdot)$ in eq. (1), depending on the regression vector selected; such as the non-linear autoregressive moving average with exogenous inputs (NARMAX) model, the non-linear ARX (NARX) structure or the non-linear OE (NOE) model [16].

We used a NARX model structure, parametrized by a multilayer feed-forward neural network (multilayer perceptron). This model consists of an input and two processing layers (see figure 1). The input layer is formed by the elements of the regression vector. The inputs are propagated forward through two processing layers before the output sample is calculated. The processing layers are formed by a hidden and an output layer, which consist of a number of computing nodes based on McCulloch-Pitts neurons. Each node makes the operation of a weighted sum of the incoming signals and a bias term, fed through an activation function $\sigma(\cdot)$, resulting in the output value of the neuron.

The dimension n of the regression vector, $\varphi(t) = [\varphi_1 \ \cdots \ \varphi_n]$, determines the number of inputs in the network; where $n = ny+nu+1$. Each input is fed to each node of the hidden layer. A generic ith hidden node, h_i, with n inputs and one output, has an input-output relationship as

$$h_i = \sigma\left(\sum_{j=1}^{n} w_{ij}\varphi_j + w_{i0}\right) \tag{13}$$

where w_{ij} are the weights of the ith node given to each jth input and w_{i0} is the bias associated with each node. The activation function of the hidden nodes was chosen as an hyperbolic tangent, $\sigma(x) = \tanh(x) = (1 - e^{-2x})/(1 + e^{-2x})$. The

output layer contains one node with a linear activation function, resulting an estimating output of the model as

$$\hat{y}(t) = f(\varphi(t),\hat{\theta}_N) = \sum_{i=1}^{nh} W_i h_i + W_0 = \sum_{i=1}^{nh} W_i \tanh(\sum_{j=1}^{n} w_{ij}\varphi_j + w_{i0}) + W_0 \qquad (14)$$

where nh is the number of hidden nodes, W_i are the weights given to each ith output of the hidden nodes and W_0 is the output node bias. W_i, W_0, w_{ij} and w_{i0} are the model parameters and they are calculated through a training process.

For the non-linear modeling procedure we considered the QRS waveform, *y(t)*, as the system response to a ramp input, *u(t)*. The NARX structure network is trained for predicting each value of $\hat{y}(t)$, based on observed output values *y(t)*, as the target, and its regression vector $\varphi(t)$ as the network input. $\varphi(t)$ is defined by the same regressor as for the linear ARX model. This network has the advantage of possessing a static predictor ; i.e., none component of the regression vector depends on past outputs of the model. Other non-linear structures as NARMAX or NOE models need a recurrent neural network.

2.4.1 Model structure selection

To settle the structure of the neural network based NARX model, the dimension of $\varphi(t)$ (the model order) and the number of hidden nodes had to be chosen. Lead-specific network structures were used. The model order and the number of hidden nodes, *nh*, will decide the model abilities for approximating the normal part of the QRS complex. After having tested several alternatives for each lead, using the training dataset, we chose a model structure with $nh = 6$ in the X and Z leads and $nh = 7$, in the Y lead; and a model order given by $ny=1$ and $nu=1$, i.e. $\varphi(t) = \left[-y(t-1) \quad u(t) \quad u(t-1)\right]$.

2.4.2 Parameter estimation algorithm

The training is done in a batching operation using a regression matrix, formed by the regression vectors $\varphi(t)$, as the network input and the observed QRS, *y(t)*, as the target output. The objective of training is then to determine a mapping from the set of training data to the set of possible parameters $\hat{\theta}_N$, minimizing the loss function described by eq. (3). The search for the minimum cannot be computed analytically so it has to be done by some optimization method with an iterative scheme according to

$$\hat{\theta}_N^{(i+1)} = \hat{\theta}_N^{(i)} - \mu^{(i)} g^{(i)} \qquad (15)$$

where $\hat{\theta}_N^{(i)}$ is the parameter estimate after iteration number i, $g^{(i)}$ is a search direction function based on the gradient of $V_N(\hat{\theta}_N^{(i)})$ and $\mu^{(i)}$ is a step size. $g^{(i)}$ is usually expressed in terms of the loss function gradient and a search direction matrix $R^{(i)}$, $g^{(i)} = \left[R^{(i)}\right]^{-1} V_N'(\hat{\theta}_N^{(i)})$. For $R^{(i)} = I$ (identity matrix), a simple gradient direction search is used as the steepest descent algorithm. Back-propagation error rules are steepest descent-type algorithms. They suffer from a

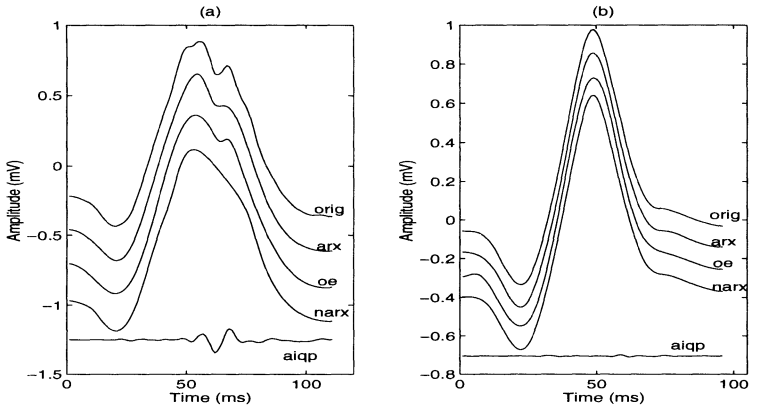

Figure 2: Examples in the X lead of the modeling techniques and estimates of AIQP. (a) A VT subject with a notch. (b) A no-event subject.

slow convergence rate. This can be improved using a *Newton* search direction, given by the loss function Hessian, $R^{(i)} = V_N''(\hat{\theta}_N^{(i)})$. This function may be quite costly for computing. In practice, approximations of the Hessian, $H^{(i)}$, are normally preferred, as the Gauss-Newton methods and its Levenberg-Marquardt modification. The Gauss-Newton search direction is defined by $R^{(i)} = H^{(i)}$ and $\mu^{(i)}$ is usually set to 1. In the Levenberg-Marquardt algorithm the Hessian is approximated by $H^{(i)}$ adding some small positive scalar δ, given a search direction defined by $R^{(i)} = H^{(i)} + \delta^{(i)} I$ (see details in [12][16]).

Parameters then were estimated with a Levenberg-Marquardt algorithm using MATLAB. $\delta^{(i)}$ was initially set to 1 and it was adjusted adaptively by the algorithm according to the loss function decrease ratio.

3 Results

The modeling processes to extract abnormal signals in the HRECG were applied to each XYZ leads of both VT and non-VT groups. The performance of the linear and non-linear procedures is depicted in figure 2. Panel (a) shows the original (top trace) and modeled X lead from a VT subject with a QRSD of 107 ms. From top to bottom, the traces are the original, ARX, OE and NARX modeled signals. The bottom trace corresponds to the residual (AIQP signal) formed by subtracting the NARX modeled QRS from the original. A visible notch is seen in the AIQP waveform. The RMS amplitude of the AIQP signal depends on the model structure used: $AIQPx_{ARX} = 13.85\ \mu V$, $AIQPx_{OE} = 16.19\ \mu V$ and $AIQPx_{NARX} = 16.67\ \mu V$. Following the same format, panel (b) shows the X lead from a non-VT subject. The AIQP signal has no distinct features or transient events. The RMS amplitudes are: $AIQPx_{ARX} = 5.84\ \mu V$, $AIQPx_{OE} = 9.45\ \mu V$ and $AIQPx_{NARX} = 2.20\ \mu V$.

HRECG Index	Non-VT (24) (μV RMS)	VT (16) (μV RMS)	p value
$AIQPx_{ARX}$ [*ny*=7, *nu*=8]	6.49 ±2.12	12.52 ±7.04	0.004
$AIQPy_{ARX}$ [*ny*=8, *nu*=3]	34.45 ±22.96	49.11 ±26.60	0.084
$AIQPz_{ARX}$ [*ny*=5, *nu*=15]	7.69 ±4.21	13.28 ±7.31	0.011
$AIQPx_{OE}$ [*ny*=3, *nu*=8]	11.11±3.68	18.41±8.44	0.004
$AIQPy_{OE}$ [*ny*=9, *nu*=3]	21.25±8.70	28.84±18.24	0.137
$AIQPz_{OE}$ [*ny*=7, *nu*=8]	6.45±3.98	10.80±6.20	0.021
$AIQPx_{NARX}$ [*ny*=1, *nu*=1], *nh*=6	5.60±3.94	10.27±7.82	0.039
$AIQPy_{NARX}$ [*ny*=1, *nu*=1], *nh*=7	6.32±4.31	9.45±7.11	0.128
$AIQPz_{NARX}$ [*ny*=1, *nu*=1], *nh*=6	7.58±4.09	20.30±17.79	0.012
RMS40	29.84 ±22.11	38.15 ±38.91	0.447
QRSD	105.1 ±20.5 ms	113.8 ±22.1 ms	0.219

Table I: Mean AIQP and LP values in VT and non-VT groups.
A p value < 0.05 indicates a significant difference between groups.

The mean RMS amplitudes of AIQP in the XYZ leads for the VT and non-VT groups, for the different modeling techniques, are shown in Table I. The model structures selected for each modeling procedure is shown for each AIQP index. For both linear an nonlinear modeling methods, AIQP amplitudes are significantly greater ($p<0.05$) in patients with VT than in those without VT. Statistical analysis of the data was performed using a 2-tailed Student *t* test. Both linear modeling structures in lead X give a highly significant difference ($p<0.01$) between groups. Differences in QRSD and RMS40 are small.

4 Discussion and Conclusions

Abnormal *intra*-QRS potential signals in the HRECG have been extracted using parametric linear and non-linear modeling techniques. Similar results were obtained from the different methods, selecting a proper model order or neural network structure. Linear modeling with an ARX structure is the first option for clinical application because the estimation of the parameter values is straightforward by linear regression. However, linear structures requires a pre-processing of the original signal for energy compacting. AIQP allows measurement of potential pathophysiologic signals contained wholly within the normal QRS period. HRECG indices based in the AIQP amplitude significantly enhance VT and no-event groups classification. This preliminary clinical study strongly suggests that AIQP can improve the sensitivity and positive predictive value of the HRECG for VT detection. The modeling techniques do not require exact determination of QRS limits, reducing dependence on different algorithms and filtering techniques for estimating the QRS onset and offset.

Key words : Non-linear modeling, High Resolution ECG, Signal Processing, Neural Networks

References

1. El-Sherif, N. & Turrito, G. (ed). *High Resolution Electrocardiography*, Futura Publishing Company, Mt Kisco, New York, 1992.
2. Berbari, E.J., & Lazzara, R. An introduction to high-resolution ECG recordings of cardiac late potentials, *Arch. Intern. Med.*, **148**, 1859-1863, 1988.
3. Vaitkus, P.T., Kindwall, K.E., Marchlinski, F.E *et al.* Differences in electrophysiological substrate in patients with coronary artery disease and cardiac arrest or ventricular tachycardia. Insights from endocardial mapping and signal-averaged electrocardiography, *Circulation*, **84**, 672-678, 1991.
4. Hood, M., Pogwidz, S., Patrick, J. & Cain, M. Contribution of myocardium responsible for ventricular tachycardia to abnormalities detected by analysis of signal-averaged electrocardiograms, *Circulation*, **86**, 1888-1901, 1991.
5. Cain, M.E., Ambos, H., Witkowski, F. & Sobel B.E, Fast-Fourier transform analysis of signal-averaged electrocardiograms for identification of patients prone to sustained ventricular tachycardia, *Circulation*, **69**, 711-724, 1984.
6. Lander, P., Albert, D.E. & Berbari, E.J. Spectrotemporal analysis of ventricular late potentials, *J. Electrocardiol.*, **23**, 95-108, 1990.
7. Gomis, P., Jones, D.L., Caminal, P., Berbari, E.J. & Lander, P. Analysis of abnormal signals with the QRS complex of the high resolution ECG, *IEEE Transactions on Biomedical Engineering,* submitted.
8. Flowers, N.C., Horan, L., Thomas J. *et al.* The anatomic basis for high frequency components in the electrocardiogram, *Circulation*, **39,** 531-539, 1969.
9. Lander, P., Gomis, P., Caminal, P., Berbari Pathophysiological insights into abnormal intra-QRS signals in the high resolution ECG, in *Computers in Cardiology, IEEE Computer Society Press*, pp 273-275, 1995.
10. Steinberg, J.S., Regan, A., Sciacca, R.R., Bigger, J.T., Fleiss, J.L., Salvatore, D.E., Fosina, M. & Rolnitzky, L.A. Predicting arrhythmic events after acute myocardial infarction using the signal-averaged electrocardiogram, *The American Journal of Cardiology*, **69**, 13-21, 1992.
11. Caminal, P., Gomis, P., Laguna, P. & Jané, R. Application of parametric models to classify ECG signals, in *Proc. Second Eur. Conf. Eng. Med., Technology and Health Care,* Elsevier Sci. Pub., **2**, 354-355, 1993.
12. Ljung, L. *System Identification: Theory for the user* Prentice-Hall, Englewood Cliffs, New Jersey, 1987.
13. Akaike, H. A new look at the statistical model identification, *IEEE Trans. Automat. Contr.*, **19**, 716-723, 1974.
14. Cybenko, G. Approximation by superposition of a sigmoidal function, *Mathematics of Control, Signals and Systems*, **2**, 183-192, 1989.
15. Barron, A.R. Universal approximation bounds for superposition of a sigmoidal function, *IEEE Transactions on Information Theory*, **39**, 930-945, 1993.
16. Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P., Hjalmarsson, H. & Juditsky A. Nonlinear black-box modeling in system identification : a unified overview, *Automatica*, **31**, 1691-1724, 1995.