# Efficient Modeling of ECG Waves for Morphology Tracking

R Dubois[1], P Roussel[1], M Vaglio[2], F Extramiana[3],
F Badilini[2], P Maison-Blanche[3], G Dreyfus[1]

[1]Laboratoire d'Electronique, ESPCI-ParisTech, Paris, France
[2]AMPS-llc, New-York, NY, USA
[3]Hopital Lariboisière, APHP, université Paris 7, Paris, France

**Abstract**

*We propose a new approach to fully automatic ECG wave extraction and morphology tracking. It is based on Generalized Orthogonal Forward Regression (GOFR), which allows decomposing a one-dimensional signal into a set of appropriate parameterized functions. Two applications of GOFR to ECG modeling are presented. First, in order to delineate ECG characteristic waves, we make use of a specific function, called the Gaussian Mesa function (GMF). Secondly, we track the evolution of the T-wave morphology by introducing a Bi-Gaussian function (BGF).*

*The approach was validated on three experimental settings; the results confirm that the combination of GOFR and of an appropriate parametric function is remarkably efficient for ECG wave modeling.*

## 1. Introduction

Automatic ECG analysis is a significant topic in biomedical signal processing; several algorithms have been proposed, based on different techniques (eg.[1-2]). The approach described in this paper is a machine-learning algorithm that decomposes the ECG into a sum of parameterized functions specifically designed to fit the cardiac characteristic waves. Therefore, it combines the power of a machine-learning algorithm with the insight of the experts through the design of specific modeling functions. The first part of Section II is devoted to the mathematical baseline of the GOFR (Generalized Orthogonal Forward Regression) learning algorithm, used for fitting parameterized functions to the ECG. Subsequently, we describe two types of parameterized function: GMF (Gaussian Mesa Function) and BGF (BiGaussian Function) that stem from experts knowledge and are used together with GOFR. Results are presented in section III on two databases devoted to three specific topics: (i) wave delineation with GMF, (ii) T wave morphology changes during Sotalol intake, and (iii) T wave morphology characterization of LQT patients with BGF functions. Discussion and conclusion are subsequently presented.

## 2. Methods

### 2.1. Generalized Orthogonal Forward Regression

GOFR is an extension of the Orthogonal Forward Regression algorithm originally designed for regression and feature selection. Given a signal $s$, a parameterized function $G_w$, and a predefined parameters $N$, GOFR aims at finding $N$ vectors of parameters $\{W_i\}_{i=1..N}$ such that the sum of the $N$ functions $\{G_{Wi}\}_{i=1..N}$ is a good model of $s$.

$\tilde{s}(X) = \sum_{i=1..N} G_{W_i}(X)$ such that $\int_{\Re_X} \left(s(X) - \tilde{s}(X)\right)^2 dX$ small

where $\tilde{s}$ is the model of $s$.

A full mathematical description of the procedure, together with a demonstration of the efficiency of GOFR as compared to standard algorithms can be found in [3] but we provide here an overview of the procedure for a 1d signal. First, we construct a library of $d$ functions $G_{Wi}$ with various values of the parameter vector $W_i$.

$D = \left\{G_{W_i}\right\}_{i=1..d}$

The model $\tilde{s}$ is obtained through $N$ iterations of the following procedure (Figure 1):

(i) select the function in $D$ that is most correlated to $s$ :

$G_{W_1}$ such that $\int s(t).G_{W_1}(t)\,dt = \max_i \int s(t).G_{W_i}(t)\,dt$

(ii) fine-tune function $G_{W_1}$ : find the parameter vector $W_1$ such that function $G_{W_1}$ best fits the signal $s$. This is performed by minimizing the mean squared error function $J$ computed between $s$ and $G_{W_1}$. Since $J$ is non linear with respect to the parameters, a second order non-linear optimization iterative algorithm is used for the minimization step.

Given $J(W_1) = \int \left(s(t) - G_{W_1}(t)\right)^2 dt$

find $W_1^*$ such that $J\left(W_1^*\right) = \min_{W_1} J\left(W_1\right)$

(iii) orthogonalize the functions of the library $D$ and the signal $s$ with respect to the tuned function $G_{W_1^*}$. This step prevents from choosing further in the procedure a function $G_{W_j}$ that is close to the previously chosen functions.

In order to construct a model composed of $N$ functions $G$, the above procedure is iterated $N$ times.

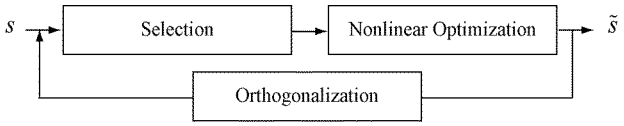


Figure 1: Overview of the 3-step procedure performed at each iteration of the GOFR algorithm.

## 2.2. Wave delineation

Wave delineation of a given heartbeat consists in (i) locating the characteristic waves (P, QRS, and T) and (ii) positioning markers at the beginning and at the end of these waves. To address the first problem, we designed a family of specific parameterized functions $G_W$ named Gaussian Mesa Functions (GMF); they are able to model the standard shapes of typical ECG waves. Each function is composed of two half-Gaussian functions linked by a horizontal segment (Figure 2). The variability in shape of such a function is large since it can model either a R wave, a ST abnormal segment, or a T wave, depending on the value of its 5-dimensional parameter vector $W=[\mu, \sigma_1, \sigma_L, \sigma_2, A]^T$, where $\mu$ is the location of the function in time, $\sigma_1$ and $\sigma_2$ the widths of the first and second Gaussian functions respectively, $\sigma_L$ the length of the horizontal segment, and $A$ the amplitude of the GMF.

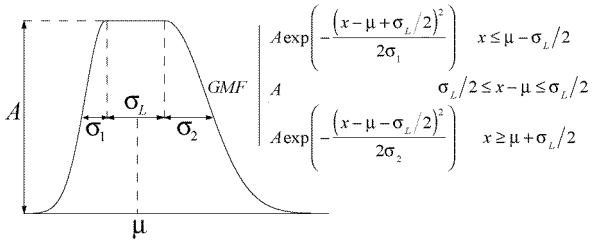


Figure 2: A Gaussian Mesa Function (GMF) is a 5-parameter function composed of two half Gaussian functions linked by a horizontal segment.

When applying the GOFR algorithm together with GMF on a heartbeat, the obtained model is a sum of $N$ GMFs, each of which models a specific part of the heartbeat. The purpose of the method is to model each wave of the heartbeat by a single GMF. Thus, for wave delineation, we decided to use $N$=6 in order to model the 5 characteristic waves and possibly a biphasic wave, or noise (Figure 3).

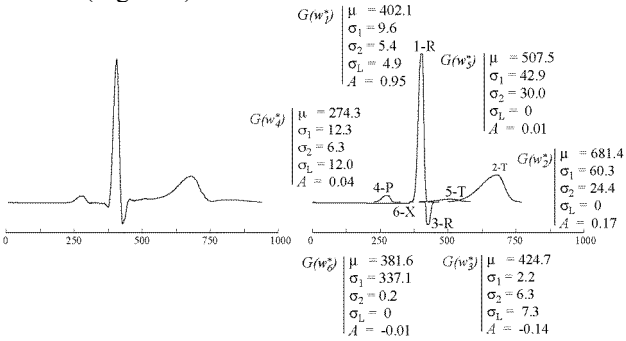


Figure 3: Obtained model with GOFR together with GMF. Each GMF is assigned a label with a medical significance.

The $N$ GMFs of the model are subsequently assigned a medical label P, R, T, or possibly $X$ if the GMF does not correspond to any of the ECG waves. This task is performed by three neural network classifiers (NNC). Each NNC has been trained to recognize one of the three labels from the GMF parameter vector. The output of each NNC is the probability for a given GMF to model the wave related to this NNC. Thus for each GMF, this process provides three probabilities: one per label. The label with the highest probability is assigned to the GMF (see example Table 1).

| | $G\left(w_1^*\right)$ | $G\left(w_2^*\right)$ | $G\left(w_3^*\right)$ | $G\left(w_4^*\right)$ | $G\left(w_5^*\right)$ | $G\left(w_6^*\right)$ |
|---|---|---|---|---|---|---|
| P-NNC | 0.00 | 0.00 | 0.00 | **0.99** | 0.00 | 0.00 |
| R-NNC | **0.99** | 0.00 | **0.15** | 0.00 | 0.00 | 0.00 |
| T-NNC | 0.05 | **0.91** | 0.05 | 0.00 | **0.31** | 0.00 |

Table 1: Example of the probabilities obtained by the 3 NNCs on the above example for each GMF of the model. Each GMF is assigned the label of the NNC that outputs the highest probability (boldface figures).

For each wave, the markers for the beginning and the end of each wave are set according to the parameters of the GMF associated to that particular wave:

$$M_{onset} = \mu - 2\sigma_1 - \sigma_L/2$$

$$M_{offset} = \mu + 2\sigma_2 + \sigma_L/2$$

In case of multiple GMFs being associated to a single wave, the markers are computed according to the relative position of the GMF [4].

## 2.3. T wave modeling

The T wave does not exhibit the same variability in shape as in the previous study; in particular, the top of a T wave is never flat, but amplitudes can be different from

one side to the other, as is the case for abnormal ST segments for example. Thus, the parameterized function used here consists of two half-gaussian functions with no flat segment and with different amplitudes $W=[\mu, A_1, \sigma_1, A_2, \sigma_2]^T$ (Figure 4); in the following, this function is referred to as a Bi-Gaussian function (BGF).

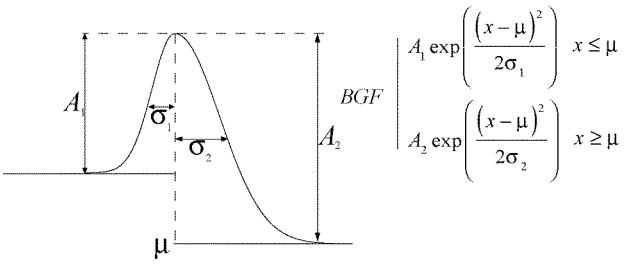


Figure 4: The bi-Gaussian function (BGF) is a 5-parameter function made of two half-Gaussian functions with different amplitudes

## 3. Results

The efficiency of the GOFR algorithm was tested on two databases. The first one (Sotalol database) has been described in [5]. It contains 12-lead ECG Holter recordings taken on 38 healthy subjects for 3 three consecutive days. On the first day, the subjects were drug-free; they took a single dose of Sotalol (160 mg) on the second day; a double dose (320mg) was delivered on the third day to a subset of the initial group. From each 24 hour ECG, a snapshot of 10 seconds was automatically extracted every 10 minutes using the Antares software (AMPS llc) [6]; a representative beat was subsequently computed from each 10-second snapshot. QRS onset/offset, and T wave offset were manually assigned by a cardiologist to these representative beats.

The second database (LQT database) is composed of 100 ECGs from patients with confirmed genotype of long-QT syndrome (50 LQT1 and 50 LQT2) [7].

### 3.1. Wave delineation

GOFR combined with GMF was used to automatically annotate the Sotalol Database. Since the representative beats are 12-lead ECG, a preprocessing step was performed to extract a 1-d signal from the 12 channels. It consists in a Principal Component Analysis computed on the 8 independent leads, the resulting 1-d signal being the projection of the original heartbeat onto the principal component.

According to the methodology discussed above, the heartbeat was modeled by GMFs, and QRS onset, T offset markers were set and compared to the reference. The results are summarized in Table 2 and a scatter plot for QT estimation is shown on Figure 5.

| | $QRS_{on}$ | $T_{off}$ | QT |
|---|---|---|---|
| Size of the database | 6377 | 6377 | 6377 |
| Number of markers | 6377 | 6377 | 6377 |
| Mean error (ms) | -9.4 | 14.7 | 24.3 |
| Std (ms) | 7.6 | 10.4 | 11.9 |

Table 2 : Wave delineation performance for the Sotalol database.

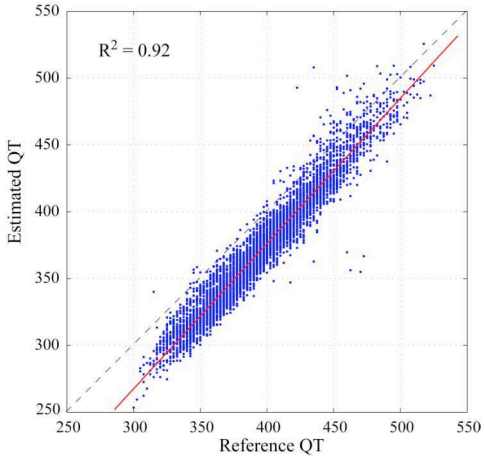


Figure 5: Scatter plot of the estimated QT values measured by the automatic delineation versus the QT intervals from manually annotated heartbeats. The correlation coefficient is $R^2$=0.92.

### 3.2. T wave analysis

The performances for the T wave analysis were first estimated on the Sotalol database. A 1-d representation of the T-wave was extracted from the 12-lead ECG using the delineation of the T-wave presented above and a PCA from the 8 independent signals of the ECG. Each T-wave was subsequently modeled by a single BGF and its morphology was described by the values of the parameters of the BGF (Figure 6 & 7, Table 3) [8,9].

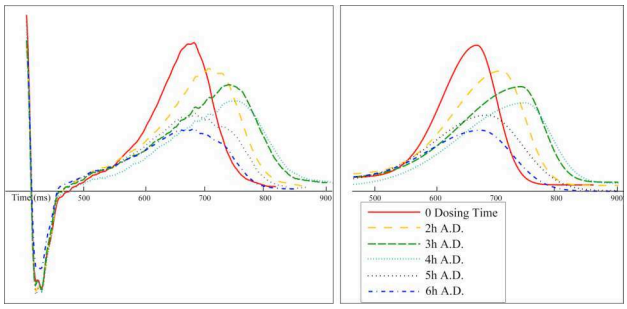


Figure 6: (left) T waves after intake of 320mg of Sotalol. (right) Corresponding BGF models.

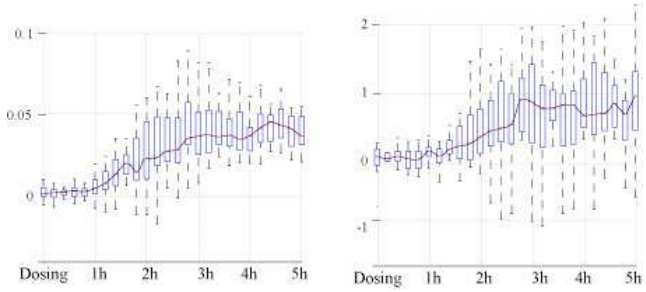


Figure 7: Trend of BGF parameters $\sigma_1$ and $\sigma_1/\sigma_2$ for double dosing of Sotalol (320 mg) paired to baseline model. (Left) Trend for parameter S1: on each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points. (Right) ratio of $\sigma_1$ to $\sigma_2$.

| | Baseline | Single Dose | Double Dose |
|---|---|---|---|
| Time peak | | 2h50±47' | 2h44±52' |
| $\mu + 2\,\sigma_2$ (ms) | 346±28 | 420±32$^{\ddagger}$ | 441±37$^{\ddagger\star}$ |
| $\sigma_1$ (ms) | 68±16 | 99±19$^{\ddagger}$ | 116±24$^{\ddagger\star}$ |
| $\sigma_2$ (ms) | 30±5 | 40±13$^{\ddagger}$ | 45±11$^{\ddagger\star}$ |
| $\sigma_1/\sigma_2$ | 2.3±0.8 | 2.6±0.6 | 2.6±0.5$^{\ddagger}$ |

Table 3: Results for BGF parameters at the peak Sotalol concentration [Extramiana]. ($^{\ddagger}p<0.05$ vs baseline, $^{\star}$ p<0.05 vs Single Dose).

Secondly, the performance of the algorithm was validated on the LQTS database. Here again, the 12-lead representation of the T-waves was projected onto a 1-dimensional space by PCA, and the obtained 1-d T-wave was modeled by a single BGF [9]. The parameter vectors of the models were compared to baseline ECGs from the Sotalol database (Table 4).

| | Control | $LQT_1$ | $LQT_2$ |
|---|---|---|---|
| Number of ECGs | 2351 | 50 | 49 |
| $\mu + 2\,\sigma_2$ (ms) | 338±24 | 401±67 | 410±64 |
| $\sigma_1$ (ms) | 52±6 | 61±12$^{\ddagger}$ | 69±21$^{\ddagger}$ |
| $\sigma_2$ (ms) | 29±3 | 29±7 | 35±13$^{\ddagger\star}$ |
| $\sigma_1/\sigma_2$ | 1.8±0.3 | 2.4±1.7$^{\ddagger}$ | 2.1±0.8$^{\ddagger}$ |
| $A_1$ (μV) | 1073±361 | 929±431$^{\ddagger}$ | 757±502$^{\ddagger\star}$ |
| $A_2$ (μV) | 1165±406 | 974±502$^{\ddagger}$ | 761±378$^{\ddagger\star}$ |

Table 4: Result for BGF parametres for LQT database ($^{\ddagger}p<0.05$ vs baseline, $^{\star}$ p<0.05 vs Single Dose)

## 4. Discussion and conclusions

The GOFR algorithm is an efficient algorithm for ECG modeling. Together with specific functions (GMF or BGF) it provides highly informative models whose parameters can be used to track the morphology of the waves. In wave delineation, the accuracy of this method is high. The stability of the markers can be estimated via the standard deviation of the error and is in a satisfactory range. The mean value of the error on the markers might be corrected by a more accurate choice of the marker position estimation from GMF parameters. Nevertheless, the QT estimated interval is highly correlated to the expert's reference. In T-wave modeling, in both databases, the values of the parameters are strongly related to the shape of the T-wave. Therefore, GOFR with appropriate parameterized functions can be used efficiently for fully automatic wave morphology characterization and tracking.

## References

[1] Martinez J, Almeida R, Olmos S, Rocha A, Laguna P. A wavelet-based ECG delineator: evaluation on standard databases. IEEE Trans Biomed Eng. 2004;51(4):570-81.

[2] Suppappola S, Sun Y, Chiaramida SA. Gaussian Pulse Decomposition: An intuitive Model of Electrocardiogram Waveforms. Annals of biomedical engineering. 1997;25:252-60.

[3] Dubois R, Quenet B, Faisandier Y, Dreyfus G. Building meaningful representations for nonlinear modeling of 1d- and 2d-signals: applications to biomedical signals. Neurocomputing. 2006;69(16-18):2180-92.

[4] Dubois R, Maison-Blanche P, Quenet B, Dreyfus G. Automatic ECG wave extraction in long-term recording using Gaussian mesa function models and nonlinear probability estimators. Computer Methods and Programs in Biomedicine. 2007;88:217-33.

[5] Sarapa N, Morganroth J, Couderc JP, Francom SF, Darpo B, Fleishaker JC, et al. Electrocardiographic identification of drug-induced QT prolongation: assessment by different recording and measurement methods. Ann Noninvasive Electrocardiol. 2004 Jan;9(1):48-57.

[6] Badilini F, Vaglio M, Sarapa N. Automatic Extraction of ECG Strips from Continuous 12-lead Holter Recordings for QT Analysis at Prescheduled versus Optimized Time Points. Annals of Noninvasive Electrocardiology. 2009;14(s1):S22-S9.

[7] Imboden M, Swan H, Denjoy I, Van Langen IM, Latinen-Forsblom PJ, Napolitano C, et al. Female Predominance and Transmission Distortion in the Long-QT Syndrome. N Engl J Med. 2006 December 28, 2006;355(26):2744-51.

[8] Extramiana F, Dubois R, Vaglio M, Roussel P, Dreyfus G, Badilini F, et al. The time course of new T-wave ECG descriptors following single and double dose administration of Sotalol in healthy subjects. Annals of Noninvasive Electrocardiology. in press.

[9] Badilini F, Vaglio M, Dubois R, Roussel P, Sarapa N, Denjoy I, et al. Automatic analysis of cardiac repolarization morphology using Gaussian mesa function modeling. J Electrocardiol. 2008 Nov-Dec;41(6):588-94.

Address for correspondence
Rémi Dubois, remi.dubois@espci.fr
Laboratoire d'Électronique,ESPCI-ParisTech
10 rue Vauquelin, 75005 Paris-France