A NOVEL METHOD TO REPRESENT ECG SIGNALS VIA PREDEFINED PERSONALIZED SIGNATURE AND ENVELOPE FUNCTIONS

B. S Yarman, H. Gürkan, Ü. Güz, B. Aygün
Department of Electronics Engineering, İşık University, 80670 Maslak, Istanbul, Turkey
Tel: +90.212.286 29 60, Fax: +90.212. 285 28 75
E-mail: [yarman@isikun.edu.tr] [bakan@isikun.edu.tr] [guz@isikun.edu.tr]

Abstract-In this paper, a new method to model ECG signals by means of "Predefined Personalized Signature and Envelope Functions" is presented. ECG signals are somewhat unique to a person. Moreover, it presents quasi-stationary behavior. Therefore in this work, on a frame basis, personal ECG signals $X_i(t)$ is modeled by the form of $X_i(t) = C_i \phi(t) \alpha(t)$. In this model, $\phi(t)$ is defined as the Personalized Signature Function (PSF); $\alpha(t)$ is referred to as Personalized Envelope Function (PEF) and $C_i$ is called the Frame-Scaling Coefficient (FSC). It has been demonstrated that for each person, the sets $\Phi = \{\phi(t)\}$ and $A = \{\alpha(t)\}$ constitute a "Predefined Personalized Functional Bases or Banks (PPFB)" to describe any measured ECG signal. Almost optimum forms of (PPFB), namely $\{\alpha(t)\}$, $\{\phi(t)\}$ pairs are generated in the Least Mean Square (LMS) sense. Thus, ECG signal for each frame is described in terms of the two indices "$R" and "K" of PPFB and the frame-scaling coefficient $C_i$. It has been shown that the new method of modeling provides significant data compression. Furthermore, once PPFB are stored on each communication node, transmission of ECG signals reduces to the transmission of indexes "$R" and "K" of $[\alpha(t), \phi(t)]$ pairs and the coefficients $C_i$, which also result in considerable saving in the transmission band.

Keywords - Models, Compression, Transmission of ECG Signals

I. INTRODUCTION

Biomedical signals are used widespread in diagnosis and therapy of many diseases. The processing, storing and transfer of biological signals is gaining ever more importance in today's information era. However, since such signals comprise huge amounts of data, the storage, transfer and reconstruction of biological signals create certain limitations. One way to overcome this problem is the compression of the signals, provided that the information covered by the signal is not lost. Moreover, such signals have to be repeatedly received and evaluated during the course of the illness in order to verify the diagnosis, determine the treatment method(s) and allow follow up of the therapy to avoid abnormal patterns and complications. Therefore, compression of signals with acceptable loss is inevitable.

Diagnosis, compression and speedy transmission of the ECG signals may be achieved by means of appropriate models with least number of parameters. In this case, the measured ECG data may be described in terms of the parameters of the selected model. Usually, linear models are preferred for the sake of simplicity. In the existing literature, efforts have been made to represent the ECG signals in terms of the linear combination of somewhat orthogonal functions such as complex exponentials, trigonometric functions, wavelets or pre-selected, signature functions, which best describe the signal [1-7]. For example, in [5-6], over a finite time interval, limited number of ad-hoc sinusoidal waveforms with arbitrary frequencies is selected as signature functions. Then, these forms are orthogonalized to constitute a functional base to represent the original ECG signal. Eventually, the ECG data is approximated by the weighted sum of these, so called signature base functions obtained from the previous step. In these techniques however, the optimum form of the signature base functions and the expected minimum numbers of terms in the summation, are in question. The difficulties of the above-mentioned techniques have been overcome with the novel modeling method proposed in this paper. The new method is based on the generation of the so called the "Predefined-Personalized Functional Bases or Banks - PPFB". The PPFB consist of the two major functional banks, namely the "Envelope Functions Bank - EFB" and the "Signature Functions Bank - SFB". The measured ECG data is then; modeled as the multiplication of two appropriate functions, which are retrieved within a constant from the above-mentioned banks respectively. In the following sections first the new technique to model the measured ECG data is introduced. Then, SFB, EFB and reconstruction algorithms are presented. Finally, an example is given to exhibit the utilization of the new modeling technique.

II. NEW METHOD

ECG signals are somewhat unique to a person. Moreover, on the personnel basis, it presents a quasi-stationary like behavior. Therefore, it would be appropriate to extract the statistical features of the ECG signals over a reasonable length of time. For the sake of practicality, we present the new technique on the discrete time domain since all the measurements are made with digital equipment. Let $X(n)$ be the discrete time domain representation of the measured ECG signal of a person with length $N$. Let this signal be analyzed frame by frame. Let $X_i(n)$ represents the portion of the original signal within frame "i". Then, we make the following "Main Statement" which constitutes the basis of the proposed technique in this work.

Main Statement:
(a) For any time frame "i", the measured ECG signal which is given by the vector $X_i$, can be expressed or approximated as
**Title and Subtitle**
A Novel Method to Represent ECG Signals Via Predefined Personalized Signature and Envelope Functions

**Performing Organization Name(s) and Address(es)**
Department of Electronics Engineering Istanbul University 80670 Maslak, Istanbul Turkey

**Sponsoring/Monitoring Agency Name(s) and Address(es)**
US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500

**Abstract**
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-26, 2001 held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom. The original document contains color images.
where $C_i$ is a real constant, $\psi_T = [\varphi_1, \varphi_2, \varphi_3 \ldots \varphi_{LF}]$ is a row vector. The vector $C_i \varphi_i$ carries "almost maximum energy" of $X_i$ in the LMS sense. That is to say, $C_i \varphi_i$ is the best approximation of $X_i$ with one term (i.e. $X_i = C_i \varphi_i$) that minimizes the sum of square error point by point, over the frame under consideration.

The $(L_F \times 1)$ diagonal matrix $\alpha_i = \begin{bmatrix} a_{i1} & 0 & 0 \cdots & 0 \\ 0 & a_{i2} & 0 \cdots & 0 \\ \vdots & \vdots & \vdots \cdots & \vdots \\ 0 & 0 & a_{il_F} \end{bmatrix}$ acts as an envelope term, which may satisfy the equality or further reduce the LMS error in (1a). The integer $L_F$ designates the total number of elements in a frame "i".

(b) In continuous time domain, (1a) takes the form of

$$X_i(t) = C_i \varphi_i(t) \alpha_i(t)$$

(1b)

Based on the main statement we can make the following definitions.

**Definition 1a:** The vector $\varphi_i$ (or equivalently sequence $\varphi_i$) is called the Personalized Signature Vector (PSV) or Signature Sequence since it carries almost maximum energy of the ECG frame vector $X_i$ with a constant $C_i$.

**Definition 1b:** In a similar manner to that of Definition 1a, the continuous time domain function $\varphi_i(t)$ is called the Personalized Signature Function (PSF).

**Definition 2a:** The diagonal matrix $\alpha_i$ (or equivalently sequence $\alpha_i$) is called the Personalized Envelope Matrix (PEM) or Envelope Sequence since it matches the envelope of $C_i \varphi_i$ to the original ECG frame vector $X_i$.

**Definition 2b:** In a similar manner to that of Definition 2a, the continuous function $\alpha_i(t)$ is called the Personalized Envelope Function (PEF).

**Definition 3:** The real constant $C_i$ is called the Frame Scaling Coefficient (FSC).

In order to verify the main statement, let us proceed as follows.

A discrete ECG signal $x(n)$, can be written as

$$x(n) = \sum_{i=1}^{N} x_i \delta_i(n-i)$$

(2a)

In this equation, $\delta_i(n)$ represents the unit sample; $x_i$ designates the amplitude of the sequence $x(n)$ of length $N$. $x(n)$ can also be given employing the vector/matrix notation.

$$X = [x(1) \ x(2) \ \cdots \ x(N)] = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix}$$

(2b)

$X$ is called the "Main Frame vector" and it is divided into sub-frames with equal lengths, having, for example, 8,16 or 32 samples etc. The Frame Matrix that is represented by $M_F$ is obtained by means of the sub-frame vectors:

$$M_F = \begin{bmatrix} X_1 & X_2 & \cdots & X_{N_F} \end{bmatrix}$$

(3)

where

$$X_i = \begin{bmatrix} s_{i1} \ldots s_{iL_F+1} \\ \vdots \\ s_{i1} \ldots s_{iL_F+2} \\ \vdots \\ s_{iL_F} \end{bmatrix}, \ i=1\ldots N_F$$

(4)

$N_F = N/L_F$ designates the total number of frames in $X$. It can be shown that each frame sequence or vector $X_i$ can be spanned to a vector space formed by the orthonormal vectors $\{V_{ki}; k=1,2,3\ldots LF\}$, such that

$$X_i = \sum_{k=1}^{LF} c_k V_{ki}$$

(5)

$$c_k = (X_i)^T V_{ki}$$

(6)

$V_{ki}$ are determined by minimizing the expected value of the error vector $e = X_i - \sum_{k=1}^{LF} c_k V_{ki}$ with respect to $V_{ki}$. This process is called the "determination of $V_{ki}$ in the LMS sense". Eventually $V_{ki}$ are computed as the Eigenvectors of the Autocorrelation Matrix $R_i$ of the sub-frame sequence $X_i$. It is straightforward to obtain the autocorrelation matrix $R_i$ as

$$R_i = \begin{bmatrix} r_{i1}(1) & r_{i1}(2) & \cdots & r_{i1}(L_F) \\ r_{i2}(1) & r_{i2}(2) & \cdots & r_{i2}(L_F-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{i(L_F-1)}(1) & r_{i(L_F-1)}(2) & \cdots & r_{i(L_F-2)} \\ \vdots & \vdots & \ddots & \vdots \\ r_{i(L_F-1)}(L_F-1) & r_{i(L_F-1)}(L_F-2) & \cdots & r_{i(L_F-1)}(1) \end{bmatrix}$$

(7)

where

$$r_{ij}(d+1) = \frac{1}{L_F - d} \sum_{j=[(i-1)L_F+1]}^{[iL_F-1]} x_j x_{j+d}, d=0,1,2\ldots L_F-1$$

(8)

It should be noted that $R_i$ is a positive-semi definite, real symmetrical and toepplitz matrix. The above-mentioned LMS process results in the following eigen-value problem

$$R_i V_{ki} = \lambda_{ki} V_{ki} \quad ; \quad k=1,2,\ldots L_F$$

(9)

Obviously, $\lambda_{ki}$ and $V_{ki}$ are the eigen-values and the eigenvectors of the problem under consideration. It is well known that the eigen-values of the $R_i$ are also real and non-negative. Moreover, the eigenvectors $V_{ki}$ are all orthonormal. Hence, (5-6) follows.

Let eigen-values be sorted in descending order such that $\lambda_{k1} \geq \lambda_{k2} \geq \lambda_{k3} \ldots \geq \lambda_{kL_F}$ with corresponding eigenvectors.

Total energy of the sub-frame "i" is given by

$$X_i^T X_i = \sum_{k=1}^{LF} c_k^2 = \sum_{k=1}^{LF} \frac{L_F}{N_F} \lambda_{ki}$$

(10)

The eigenvectors that have the highest energy associated with the highest value of the eigen-values represent the directions of greatest variations of the signal and they are also called the principle components. Therefore, (5) may be truncated by taking the first "p" principal components, which have the highest energy of the original signal such that

$$X_i \approx \sum_{k=1}^{p} c_k V_{ki}$$

(11)
The simplest form of (10) can be obtained by setting \( p = 1 \). The eigenvector \( V_{ii} \) is called "the major signature vector". That is to say, the major signature vector, which has the highest energy in the LMS sense, may approximate each frame that belongs to the original ECG signal. Thus,

\[
X_i = c_i V_{ii}
\]

In the following section, the new ECG modeling algorithms, which utilize predefined personalized envelope and signature banks is presented.

### III. ALGORITHMS

The new modeling method is implemented within two major algorithmic steps. First the predefined personalized envelope and signature banks (PPFB) are generated (Algorithm 1). Then, any ECG signal is reconstructed utilizing the (PPFB) (Algorithm 2). Therefore, in this section, first Algorithm 1 is introduced. Then Algorithm 2 is outlined.  

**Algorithm 1: Generation of the PPFB**  
**Inputs:**
- Main Frame sequence of the ECG Signal
- Total Number of samples in the sub-frames: \( (L_F) \)

**Step 1:** Compute the total number of frames \( N_F = N/L_F \)

**Step 2:** Sort the Main frame \( X \) into sub-frames \( X_i \)

**Step 3:** For each sub-frame \( X_i \) compute the \( R_i \)

**Step 4:** For each \( R_i \) compute the eigen-values \( \lambda_i \) and the corresponding eigenvectors \( V_{ii} \)

**Step 5a:** Store the eigenvector \( V_{ii} \) which is associated with the maximum eigen-value \( \lambda_i = \max \{ \lambda_{i1}, \lambda_{i2}, \lambda_{i3}, ..., \lambda_{iLF} \} \) and simply refer this vector with the frame index, such as \( V_{ii} \).

**Step 5b:** Find the Frame Scaling Coefficient \( C_i \) in the LMS sense to approximate \( X_i = C_i V_{ii} \).

**Step 6:** Do step 5 for all the sub-frames \( (i=1,2,...,N_F) \). At the end of this step, you will be collecting the eigenvectors, which have the maximum energy for each frame.

**Step 7:** Compare all the collected eigenvectors in Step 6, with an efficient algorithm, and eliminate the similar ones. Thus, generate the Signature Bank \( \{ \Phi_{n,s}(n) : n_s=1,2,...,N_S \} \) with reduced number of eigenvectors \( V_{ii} \). Here, \( N_S \) designates the total number of original signature patterns after the reduction.

**Step 8:** Compute the Diagonal Envelope Matrix \( A_i = \{ \Phi_{n,s}(n) : n_s=1,2,...,N_S \} \) of each \( C_i V_{ii} \) computed in step 5b such that \( a_{ir} = x_{ir}/C_i v_{ir} ; r=1,2,...,L_F \).

**Step 9:** As in Step 7, eliminate the similar patterns of envelope sequences with an efficient algorithm and create the envelope bank with reduced number original sequences \( \{ \alpha_{n,s}(n) : n_s=1,2,...,N_E \} \); Here, \( N_E \) denotes the total number of original envelope patterns in the envelope bank.

The above algorithm is utilized for many different ECG signals for a person to enhance the formation of the personalized signature and envelope banks.

**Algorithm 2: Reconstruction of ECG Signals in terms of PPFB**  
**Inputs:**
- ECG signal \( \{X(n), n=1,2,...,N\} \) to be modeled.
- PPFB created utilizing the algorithm 1.
- \( L_F \) : Number of Samples in each sub-frame.

**Step 1:** Sort the sub-frames \( X_i \) as in algorithm 1.

In this research work, for a person, many ECG signals are unique to a person and exhibits quasi-repetitive similarities. Above-mentioned patterns somewhat unique to a person. Therefore, we say that ECG signals are unique to a person and exhibits quasi-repetitive similarities.

Patterns obtained by plotting \( a_{ir}(n) = (a_{ir} \text{ vs. sub-frame index-} n=1,2,...,L_F) \) and \( v_{ir}(n) = (v_{ir} \text{ vs. sub-frame index-} n=1,2,...,L_F) \) exhibits repetitive similarities. Above-mentioned patterns somewhat unique to a person. Therefore, we say that ECG signals are unique to a person and exhibits quasi-repetitive similarities. It has been observed that ECG signals are unique to a person and exhibits quasi-repetitive similarities. It is deduced that similar patterns can be eliminated; and we can create two types of bases or stationary like behavior. It has been observed that ECG signals are unique to a person and exhibits quasi-repetitive similarities. Above-mentioned patterns somewhat unique to a person. Therefore, we say that ECG signals are unique to a person and exhibits quasi-repetitive similarities. It has been observed that ECG signals are unique to a person and exhibits quasi-repetitive similarities.
Step 2:  
(a) For each sub-frame “i” pull an appropriate signature sequence $\varphi_k$ such that the distance or the total error 
$$\delta = \| X_i - C_k \varphi_k \|^2$$ is minimum for all $k=1,2,...,N_k$. 
(b) Store the index number $K$ that refers to $\varphi_k$. Hence, $X_i = C_k \varphi_k$ 

Step 3:  
(a) Pull the appropriate envelope sequence (or diagonal matrix) $\alpha_k$ such that the error is further minimized for all $r=1,2,...,N_E$. 
$$\delta_r = \min \| X_i - C_k \varphi_k \alpha_r \|^2 ; r=1,2,...,N_E.$$ 
(b) Store the index number $R$ that refers to $\alpha_R$.

It should be noted that at the end of this step, the best envelope $\alpha_R$ and signature $\varphi_K$ patterns are found by appropriate selections. Hence, the sub-frame $X_i$ is best described in terms of the patterns of $\alpha_R$ and $\varphi_K$, i.e. $X_i = \varphi_K \alpha_R$.

Step 4: Having fixed $\alpha_R$ and $\varphi_K$, compute the new frame coefficient $C_i$ to find the global minimum of the error
$$\delta_{Global} = \min \| X_i - C_i \varphi_K \alpha_R \|^2$$ and store it. At this step, the frame sequence is approximated as $X_i \approx C_i \varphi_K \alpha_R$.

Step 5: Repeat the above steps for each frame to reconstruct the original ECG sequence $X(n)$.

Eventually, the above algorithm yields the frame coefficient $C_i$ and the index numbers $R$ and $K$ for the best envelope and the signature sequences respectively for each frame. Thus, each frame $X_i$ is referred by means of the FSC-C_i and the indices $R$ and $K$ with an acceptable error
$$\delta_{Global} = \| X_i - C_i \varphi_K \alpha_R \|^2.$$

Hence, in the reproduction of ECG signals, substantial compression is achieved. Once the PPFB are stored in each communication node, then considerable amount of saving in the transmission band will be obtained.

Now, let us see the merits of the newly proposed ECG modeling technique with an example.

### IV. Example

In this example, first, several ECG signal data is recorded for a person. Each recording sampled with 500 Hz and contains 2048 samples. Then, employing the algorithm 1, we generated envelope and signature sequences bank. In the computations, the sub-frame length $N_E$ = 16 was selected. It was found that with an error $\delta_{Global} < 10^{-3}$, we were able to reconstruct all the measured ECG signals with 5 original signature (i.e. $N_S=5$) and 315 envelope (i.e. $N_E=315$) patterns. Let us presume that 12 bits represent each sample of the sub-frame $X_i$. Then, each frame of 16 samples will be represented with 12x16 = 192 bit. On the other hand, employing the newly proposed technique, each frame is described by means of $C_i$ (FSC) and two frame indices $R$ and $K$. In this example, $C_i$ is represented by 12 bits. $R$ and $K$ are represented by total of 12 bits; then, 24 bits will be good enough to represent each frame. Thus, compression rate of 192/24 = 8 is obtained. In Figure 1, original and reconstructed signals are shown. As can be seen from Figure 1 that original and reconstructed signal agrees within an error of $10^{-3}$.

![Original and Reconstructed ECG Signals](image)

**Fig. 1. Original and Reconstructed ECG Signals**

### V. Conclusion

In this paper, a novel method to represent ECG signals is presented. The proposed technique is based on the generation of the so-called, predefined personalized signature and envelope banks. It has been exhibited that the new method of modeling ECG signals, yields substantial compression and results in band saving in transmission. It is expected that detail characterization and classification of PPFB by disorders will lead to diagnosis of heart diseases in the follow-up research work.

### References