14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A TIME-FREQUENCY ADAPTIVE SIGNAL MODEL-BASED APPROACH FOR PARAMETRIC ECG COMPRESSION

N. Ruiz-Reyes, P. Vera-Candeas, P.J. Reche-López and F. Cañadas-Quesada

Electronic and Telecommunication Engineering Department, University of Jaén
Polytechnic School, 23700 Linares, Jaén, SPAIN
phone: + (34) 953648554, fax: + (34) 953648508, email: {nicolas,pvera,pjreche}@ujaen.es

ABSTRACT

A preliminary investigation of an atomic model-based algorithm for the compression of single lead ECG is presented. The paper presents a novel coding scheme for ECG signals based on time-frequency atomic signal representations. Signal-adaptive parametric models based on overcomplete dictionaries of time-frequency atoms are considered. Such overcomplete expansions are here derived using the matching pursuit algorithm.

The compression algorithm has been evaluated with the MIT-BIH Arrhythmia Database. Our algorithm was compared with several well-known ECG compression algorithms and was found to be superior at all tested bit rates. An average compression rate of approximately 140 bps (compression ratio of about 28:1) has been achieved with a good reconstructed signal quality (PRD of about 7 %).

1. INTRODUCTION

Considering the great volume of ECG data which is generated each year, the ability to efficiently manage the storage and retrieval of this data for comparison and evaluation mandates the need for ECG compression techniques. Effective storage of ECG information is required in the intensive coronary care unit, or in the long-term (24-48 hours) wearable monitoring tasks (Holter). For good diagnostic quality, each ECG lead provided by the Holter should be sampled at a rate of 250-500 Hz with 12 bits resolution. The information rate is thus approximately 11-22 Mbits/hour/lead. If efficient compression methods are employed, memory requirements may drastically drop to make the monitoring device commercially feasible. ECG compression is also of practical importance for other aspects of electrocardiography. Transmitting the ECG signal through telephone lines or mobile radio, for example, may save a crucial time and unnecessary difficulties in emergency cases. Real-time heart rhythm analysis algorithms require ECG data compression. Compression parameters may also be valuable tools for developing pattern recognition schemes and automatic diagnostic algorithms.

In practice, efficient data compression may be achieved only with lossy compression techniques. In ECG compression algorithms the goal is to achieve a minimum information rate, while retaining the relevant diagnostic information in the reconstructed signal.

Data compression methods have been mainly classified into three mayor categories: a) direct data compression, b) transformation methods, and c) parameter extraction techniques. Most of the existing data compression techniques for ECG signals lie in two of the three categories described: direct methods and transformation methods. Direct methods are realized by irregular sampling and quantization of original waveforms in the time-domain. Transform methods are based on orthogonal transforms, such as Fourier, Karhunen-Loeve (KL), DCT or wavelets.

An excellent overview of direct and transform-based ECG compression techniques before 1990 is reported in [1]. AZTEC, Fan/SAPA, TP, CORTES, DPCM, Peak-Picking and Cycle-to-Cycle are well-known examples of direct ECG compression schemes. Regarding to transformation methods, the wavelet-packet transform has received a great deal of attention over the past year (and even by now), being successfully applied to several problems in electrocardiology, including data compression [2, 3, 4, 5, 6]. The success of wavelets for ECG compression is due to their time-frequency localization capability. High compression ratio are also achieved by the Karhunen-Loeve transform [7] at the expense of a meaningful computational cost.

Other relevant works published in last years for ECG data compression are [8, 9, 10, 11, 12, 13, 14].

Recently, a ECG data compression approach based on atomic models and matching pursuit has been proposed [15]. A signal is decomposed into atoms that are included in an overcomplete dictionary. The dictionary hence can be defined to best match the signal structure. It is expected that the time-frequency localization capability of matching pursuit can be superior to orthogonal transforms. Furthermore, a few waves appear during one heart-beat period in an ECG signal. It is expected that a small number of atoms can approximate the ECG waveforms.

In this paper, signal-adaptive parametric atomic models based on overcomplete dictionaries of wavelet functions have also been applied to ECG waveform compression. Such overcomplete expansions are derived using the matching pursuit algorithm. The resulting representations are signal-adaptive in that the atoms for the model are chosen to match the signal behavior; furthermore, the models are parametric in that the atoms can be described in terms of simple parameters.

Atomic decompositions and matching pursuit are revised in section 2. Section 3 describes in detail the proposed scheme for ECG compression. Experimental results are presented in section 4. Finally, conclusions are resumed in section 5.

2. ATOMIC DECOMPOSITIONS AND MATCHING PURSUIT

2.1 Principles of atomic modelling

Time-frequency atomic signal representations have been of ongoing interest since their introduction by Gabor several decades ago. The fundamental notions of atomic modelling are that a signal can be decomposed into elementary functions that are localized in time-frequency and that such decompositions are useful for applications such as signal analysis and coding. Here, an overview of the computation and properties of atomic models is presented. The overview is based on an interpretation of atomic modelling as a linear algebraic inverse problem.

A signal model of the form

$$x[n] = \sum_{i=1}^{M} \alpha_m g_i[n] \tag{1}$$

can be expressed in matrix notation as

$$\mathbf{x} = \mathbf{D} \cdot \alpha \quad \text{with} \quad \mathbf{D} = [g_1 g_2 \cdots g_i \cdots g_M] \tag{2}$$

where the signal $\mathbf{x}$ is a column vector (N x 1), α is a column vector of expansion coefficients (M x 1), and $\mathbf{D}$ is an (N x M) matrix whose columns are the expansion functions $g_i[n]$. In this framework, derivation of the model coefficients is an inverse problem.

When the functions $g_i[n]$ constitute a basis, such as in Fourier and wavelet decompositions, the matrix $\mathbf{D}$ is square (N = M) and invertible, and the expansion coefficients α for a signal $\mathbf{x}$ are uniquely given by

$$\alpha = \mathbf{D}^{-1} \cdot \mathbf{x} \tag{3}$$

While this ease of computation is an attractive feature, basis expansions are not generally useful for signal modelling, because they do not provide compact models of arbitrary signals [16]. To overcome the difficulties of basis expansions, signals can instead be modelled using overcomplete set of atoms that exhibit a wide range of time-frequency behaviors. Such overcomplete expansions allow for compact representation of arbitrary signals for the sake of compression and analysis. With respect to the interpretation of signal modelling as an inverse problem, when the functions $g_i[n]$ constitute an overcomplete or redundant set ($M > N$), the dictionary matrix $\mathbf{D}$ is of rank N and the linear system in equation (2) is underdetermined. The null space of $\mathbf{D}$ then has nonzero dimension and there are an infinite number of expansions of the form of equation (1).

There are a wide variety of approaches for deriving overcomplete signal expansions, which differ in the structure of the dictionary and the manner in which dictionary atoms are selected. Examples include best basis methods and adaptive wavelet packet, where the overcomplete dictionary consists of a collection of bases. Signal decomposition schemes using more general overcomplete sets can also be considered. Such approaches can be roughly grouped into two categories: a) parallel methods, such as the method of frames, basis pursuit, and FOCUSS, in which computation of the various expansion components is coupled and derive exact solutions; b) sequential methods, such as matching pursuit and its variations, in which models are computed one component at a time and derive sparse approximate solutions according to suboptimal criteria. All these methods can be interpreted as approaches to solving inverse problems.

Since sparse approximate solutions are of interest for compact signal modelling, matching pursuit is the chosen method for deriving overcomplete signal expansions in the proposed ECG compression scheme. Furthermore, it provides a framework for deriving such expansions by successive refinements with low computational cost.

2.2 Matching pursuit

Matching pursuit [17] is a greedy iterative algorithm that offers a sub-optimal solution for decomposing a signal $x[n]$ in terms of unit-norm expansion functions $g_i[n]$ chosen from an overcomplete dictionary D. When a well-designed overcomplete dictionary is used in matching pursuit, the nonlinear nature of the algorithm leads to compact signal-adaptive parametric models [17, 18].

At the first iteration, the atom $g_i[n]$ which gives the largest inner product with analyzed signal $x[n]$ is chosen. The contribution of this vector is then subtracted from the signal and the process is repeated on the residual. At the m-th iteration, the residue is:

$$r^m[n] = \begin{cases} x[n] & m = 0 \\ r^{m+1}[n] + \alpha_{i(m)} \cdot g_{i(m)}[n] & m \neq 0 \end{cases} \tag{4}$$

where $\alpha_{i(m)}$ is the weight associated to optimum atom $g_{i(m)}[n]$ at the m-th iteration, and $i(m)$ the dictionary index of the optimum atom chosen at the m-th iteration.

The orthogonality principle gives the weight α_i^m associated to each atom $g_i[n] \in D$ at the m-th iteration:

$$\begin{aligned} &\langle r^{m+1}[n], g_i[n] \rangle = \langle r^m[n] - \alpha_{i(m)} \cdot g_{i(m)}[n], g_i[n] \rangle = 0 \\ &\Longrightarrow \alpha_i^m = \frac{\langle r^m[n], g_i[n] \rangle}{\langle g_i[n], g_i[n] \rangle} = \frac{\langle r^m[n], g_i[n] \rangle}{\|g_i[n]\|^2} = \langle r^m[n], g_i[n] \rangle \end{aligned} \tag{5}$$

where the last step follows from restricting the atoms to be unit-norm.

The l^2 norm of $r^{m+1}[n]$ can be expressed as:

$$||r^{m+1}[n]||^2 = ||r^m[n]||^2 - \frac{|\langle r^m[n], g_i[n] \rangle|^2}{||g_i[n]||^2} = ||r^m[n]||^2 - |\alpha_i^m|^2 \tag{6}$$

which is minimized by maximizing

$$|\alpha_i^m|^2 = |\langle r^m[n], g_i[n] \rangle|^2 \tag{7}$$

Therefore, the optimum atom $g_{i(m)}[n]$ (and its weight $\alpha_{i(m)}$) at the m-th iteration are obtained from (8):

$$g_{i(m)}[n] = \arg\min_{g_i \in D} \|r^{m+1}[n]\|^2 = \arg\max_{g_i \in D} |\langle r^m[n], g_i[n] \rangle|^2 \tag{8}$$

This is simply equivalent to choosing the atom whose inner product with the signal has the highest magnitude.

To enable representation of a wide range of signal features, a large dictionary of time-frequency atoms is used in matching pursuit. The computation of correlations $\langle r^m[n], g_i[n] \rangle$ for all $g_i[n] \in D$ at each iteration is highly computational consuming. As derived in [17], this computation can be substantially reduced using an updating formula based on equation (4). The correlations at the m-th iteration are given by:

$$\langle r^{m+1}[n], g_i[n] \rangle = \langle r^m[n], g_i[n] \rangle - \alpha_{i(m)} \cdot \langle g_{i(m)}[n], g_i[n] \rangle \tag{9}$$

where the only new computation required for the correlation updating corresponds to the cross-correlation term $\langle g_{i(m)}[n], g_i[n]\rangle$, which can be pre-calculated and stored, if enough memory is available, once set D has been determined.

3. THE COMPRESSION ALGORITHM

The ECG signal may be considered a quasi-periodic signal. The main redundancies in the ECG signal exist in the form of correlation between adjacent or past beats (inter-beat correlation) and correlation between adjacent samples (intra-beat correlation).

The inter-beat correlation suggests the idea of using a Long-Term Predictor (LTP) [10]. The frequent existence of abnormal beats in some pathological cases suggests using a beat codebook. The codebook is used to store ”typical” past beats. The intra-beat correlation suggests using a Short-Term Predictor (STP). With LTP, STP and a beat codebook, a predicted beat can be estimated, and a residual signal, which has lower variance, can be calculated [13].

Our approach is somewhat different to the one proposed in [13]. Inter-beat correlation is reduced by pattern matching between two consecutive beats in an analysis by synthesis framework. Nevertheless, intra-beat correlation is reduced by operating over the difference between consecutive beats. This residual signal is modelled using matching pursuit over an overcomplete dictionary of time-frequency atoms.

Figure 1 shows the encoding stage of the proposed ECG coder and figure 2 shows the decoding stage of the same system.

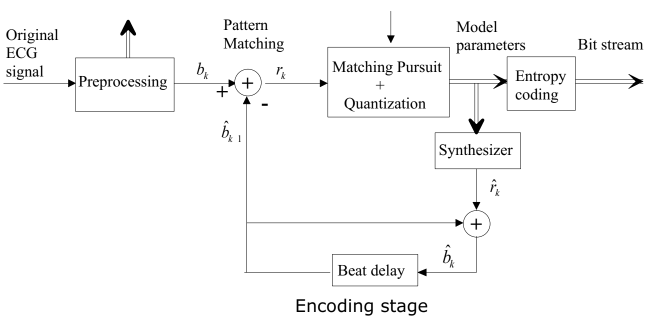

Figure 1: Encoding stage of the proposed ECG compression system

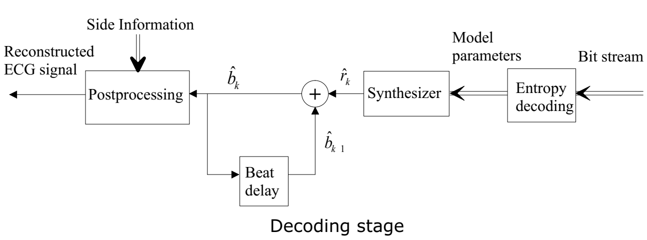

Figure 2: Decoding stage of the proposed ECG compression system

The general scheme of the proposed ECG compression method consists of three main subsystems: 1) preprocessing, 2) encoding (pattern matching and residual signal coding), 3) decoding. These subsystems are described below.

3.1 Preprocessing

The ECG signal is preprocessed prior to compression. Preprocessing consists of segmentation, alignment with respect to the fiducial point (R wave), nonuniform filtering and baseline removal.

Segmentation divides the ECG signal into beats (complexes), and every beat is further divided into three sections: P section, QRS section and T section. The motivation for such beat segmentation arises from the fact that every one of the three sections has a different diagnostic meaning and a different power spectral density.

The alignment with respect to the fiducial point between adjacent beats involves sending side information to the decoder, which must be taken into account in the last step at the decoder (postprocessing).

The nonuniform filtering consists of two different FIR filters. The P and T waves are filtered with a 0.01-50 Hz bandpass FIR filter, and the QRS section with a 0.1-100 Hz bandpass FIR filter. The filters are switched according to segmentation. The last step of preprocessing is baseline removal using cubic splines.

3.2 The encoding subsystem

The encoder matches the current preprocessed beat b_k with the previous synthesized one $\hat{b}_{k-1}$, and computes a difference signal r_k, taking into account that b_0 is an N-length zero vector ($b_0 = \hat{b}_0 = \mathbf{0}$):

$$\begin{array}{ll} r_k = b_k & k = 1 \\ r_k = b_k - \hat{b}_{k-1} & k \neq 1 \end{array} \tag{10}$$

If the length of the synthesized beat $\hat{b}_{k-1}$ is different from that of the current beat b_k, the last one is cut or zero-padded at the edges. The difference or residual signal r_k is represented by atomic modelling using matching pursuit with a dictionary of orthogonal wavelet-based atoms and efficiently coded. The current synthesized ECG signal $\hat{b}_k$ is obtained by adding the current decoded residue $\hat{r}_k$ to the previous synthesized ECG signal $\hat{b}_{k-1}$:

$$\hat{b}_k = \hat{b}_{k-1} + \hat{r}_k \tag{11}$$

At each segment, the matching pursuit algorithm is iterated, extracting atoms from the original beat (first segment) or the current residue (remaining segments), until the difference between the original ECG signal b_k and the synthesized one $\hat{b}_k$ reaches a predefined value of the PRD measure.

In order to achieve the same PRD value at the encoder and the decoder, the optimum weight $\alpha_{i(m)}$ at each iteration of matching pursuit must be quantized and the reconstructed value $\hat{\alpha}_{i(m)}$ applied to achieve the residue:

$$r^{m+1}[n] = r^m[n] - \hat{\alpha}_{i(m)} \cdot g_{i(m)}[n] \quad m \neq 0 \tag{12}$$

Lemarie wavelets have been chosen because we have found that they best match to the waves within each ECG beat. Orthogonal wavelets are considered to speed up the correlation updating procedure within the matching pursuit, as indicated in [19]. The overcomplete dictionary D is made up with those functions which give rise to a J-depth wavelet decomposition, being N the frame length and $W = \sum_{i=0}^{J-1} \frac{N}{2^i}$ the wavelet dictionary size.

Once the residual signal is completely modelled and the parameters of the model quantified, they are finally entropy coded.

3.3 The decoding subsystem

This subsystem exists at the transmission side as well as the receiver side, as expected of an analysis by synthesis based ECG coding scheme. The decoder recovers the residual signal from its quantized parameters, and obtains the current reconstructed beat $\hat{b}_k$ by adding the recovered residue $\hat{r}_k$ to the previous synthesized beat $\hat{b}_{k-1}$ (see equation (11)). This process is repeated beat to beat until the ECG signal is completely coded.

It must be noted that the current reconstructed beat $\hat{b}_k$ must be finally post-processed at the decoder in order to undo the alignment with respect to the fiducial point between adjacent beats performed at the encoder, which involves decoding the received side information (beat time duration and fiducial point location).

4. RESULTS AND DISCUSSION

The MIT-BIH Arrhythmia database [20] was used to evaluate the proposed compression algorithm and compare it with other known ECG compression methods (SAPA2 [21] and LTP [10]). These compressors were chosen because SAPA2 is very often referred for comparison in the literature, and LTP is one of the best ECG compressors available today.

Quantitative tests were performed using rate-distortion curves for each one of the compression algorithms to be compared. The rate was chosen to be expressed in terms of bits/s of the compressed ECG signal, and the distortion was chosen to be the PRD (in percentage units) between the reconstructed signal and the original one.

Figure 3 shows an example of reconstructed ECG signals, which were compressed by the proposed ECG compression scheme at two different PRD values.

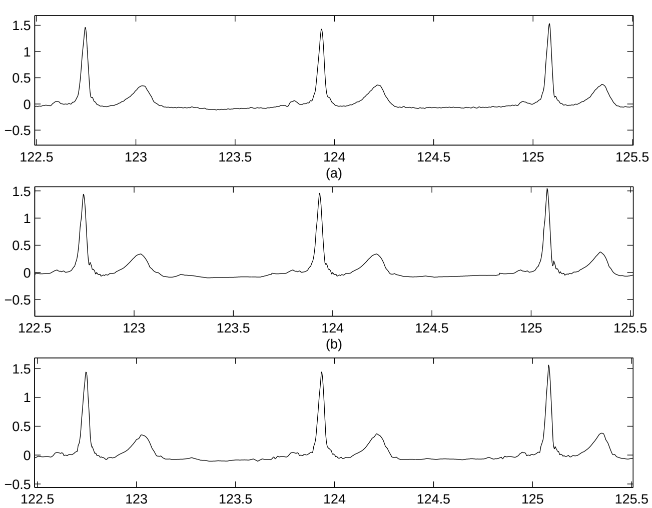

Figure 3: Example of reconstructed ECG signals (record 202 of MIT-BIH Arrhythmia database). (a) Original signal; (b) Reconstructed signal with PRD = 12% (bit rate = 51.48 bits/s); (c) Reconstructed signal with PRD = 7% (bit rate = 97.80 bits/s)

For the quantitative tests, a minute of 8 MIT-BIH records were processed: 100, 101, 103, 119, 202, 205, 207 and 209. The signals to be processed were chosen to show the performance of the proposed ECG coder for a wide variety of cases.

Figure 4 compares the distortion-rate curve obtained by the proposed algorithm with those of the SAPA2 and LTP ECG coders. Each line is a polynomial fit of the resulting points for each compression method.

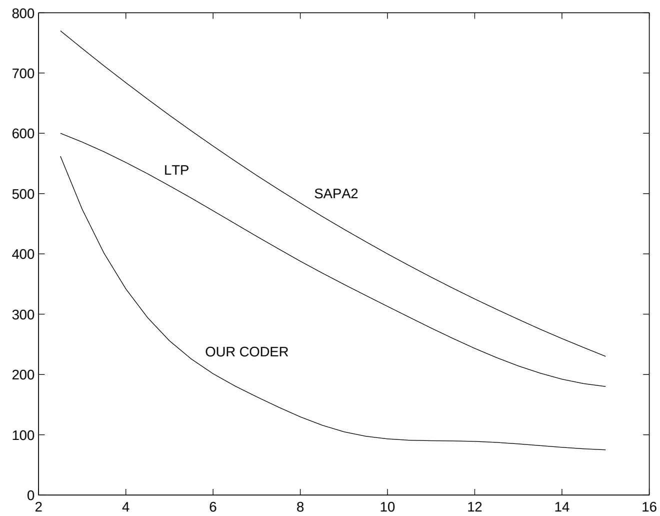

Figure 4: PRD-rate curves for SAPA2, LTP and the proposed ECG coder

Figure 4 suggests that the proposed ECG coding scheme is a profitable alternative to other existing ECG coders when the PRD measure is above a threshold of about 3%. Below that value the ECG coder performance is not good enough due to matching pursuit extracts too many atoms, most of them representing noise-like components, mainly for noisy ECG signals.

5. CONCLUSIONS

The proposed ECG coder allows to achieve low transmission rates (100-200 bits/s) while maintaining a good reconstructed signal quality (PRD of 6-10%), being an interesting alternative to other existing ECG coders. The best performance with regard to the ECG coders chosen for comparison was found with a PRD of 4-10%),

The ECG coder complexity is low owing to the use of orthogonal wavelet atoms, that make possible a fast correlation updating procedure in matching pursuit. Therefore, the compression system can be real-time implemented using inexpensive DSP chips.

In atomic models based on matching pursuit, signal adaptivity is achieved by choosing expansion functions that match the time-frequency behavior of the signal. By choosing a dictionary with a parametric structure, such as the selected wavelet one, the resultant ECG coder is both signal-adaptive and parametric, as the proposed one.

REFERENCES

[1] S. Jalaleddine, C. Hutchens, R. Strattan, and W. Coberly, ”ECG data compression techniques - A uni-

fied approach", *IEEE Trans. Biomedical Engineering*, vol. 37, no. 4, 1990, pp. 329-343.

[2] B. Bradie, "Wavelet packet-based compression of single lead ECG", *IEEE Trans. Biomedical Engineering*, vol. 43, no. 5, 1996, pp. 493-501.

[3] M.L. Hilton, "Wavelet and wavelet packet compression of electrocardiograms", *IEEE Trans. Biomedical Engineering*, vol. 44, no. 5, 1997, pp. 394-402.

[4] J. Chen, "A wavelet transform-based ECG compression method guaranteeing desired signal quality", *IEEE Trans. Biomedical Engineering*, vol. 45, no. 12, 1998, pp. 1414-1419.

[5] S. Miaou and C. Lin, "A quality-on-demand algorithm for wavelet-based compression of electrocardiogram signals", *IEEE Trans. Biomedical Engineering*, vol. 49, no. 3, 2002, pp. 233-239.

[6] B. A. Rajoub, "An efficient coding algorithm for the compression of ECG signals using the wavelet transform", *IEEE Trans. Biomedical Engineering*, vol. 49, no. 4, 2002, pp. 355-362.

[7] S. Olmos, I. Rios J. Garcia, R. Jane and P. Laguna, "ECG data compression with the Karhunen-Loeve transform", *Computers in Cardiology*, 1996, pp. 253-256, IEEE Computer Society Press.

[8] P.S. Hamilton and W.J. Tompkins, "Compression of the ambulatory ECG by average beat subtraction and residual differencing", *IEEE Trans. Biomedical Engineering*, vol. 38, no. 3, 1991, pp. 253-259.

[9] W. Philips and G. de Jonghe, "Data compression of ECG's by high-degree polinomial approximation", *IEEE Trans. Biomedical Engineering*, vol. 39, no. 4, 1992, pp. 330-337.

[10] G. Nave and A. Cohen, "ECG compression using long-term prediction", *IEEE Trans. Biomedical Engineering*, vol. 40, no. 9, 1993, pp. 877-885.

[11] W. Philips, "ECG data compression with time-warped polynomials", *IEEE Trans. Biomedical Engineering*, vol. 40, no. 11, 1993, pp. 1095-1101.

[12] H. Lee and K. M. Buckley, "ECG data compression using cut and align beats approach and 2-D transforms", *IEEE Trans. Biomedical Engineering*, vol. 46, 1999, pp. 556-564.

[13] Y. Zigel, A. Cohen and A. Katz, "ECG signal compression using analysis by synthesis coding", *IEEE Trans. Biomedical Engineering*, vol. 47, no. 10, 2000, pp. 1308-1316.

[14] R. Nygaard, G. Melnikov and A. K. Katsaggelos, "A rate distortion optimal ECG coding algorithm", *IEEE Trans. Biomedical Engineering*, vol. 48, no. 1, 2001, pp. 28-40.

[15] M. Nakashizuka, K. Niwa and H. Kikuchi, "ECG data compression by matching pursuits with multiscale atoms", *IEICE Trans. Fundamentals*, vol. E84-A, no. 8, 2001, pp. 1919-1932.

[16] M. M. Goodwin, *Adaptive Signal Models. Theory, Algorithms and Audio Applications*, 1998, Kluwer Academic Publishers.

[17] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries", *IEEE Trans. on Signal Processing*, vol. 41, no. 12, 1993, pp. 3397-3415.

[18] S. Qian and D. Chen, "Signal processing using adaptive normalized Gaussian functions", *Signal Processing*, vol. 36, no. 1, 1994, pp. 1-11.

[19] P. Vera, N. Ruiz, M. Rosa, D. Martinez and F. Lopez, "Transient Modeling by Matching Pursuits with a Wavelet Dictionary for Parametric Audio Coding", *IEEE Signal Processing Letters*, vol. 11, no. 3, March 2004.

[20] G. B. Moody, "The MIT-BIH Arrythmia Database", *Hardvard-MIT Division of Health Sciences and Technology*, 1992, Available on CD-ROM.

[21] M. Ishijima, S. B. Shin, G. H. Hostetter, and , J. Sklansky, "Scan Along Polygon Approximation for data compression of electrocardiograms", *IEEE Trans. Biomedical Engineering*, vol. BME-30, 1983, pp. 723-729.