Multiple knot spline approximation for ECG data compression*

Marta Karczewicz, Matti Hämäläinen and Moncef Gabbouj
Tampere University of Technology, Signal Processing Laboratory
P.O. Box 553, 33101 Tampere, Finland

ABSTRACT — **This paper presents a new adaptive compression method for ECG signals based on B-spline expansion, allowing multiple knots at the same location. The algorithm tries to remove as many of these knots as possible to yield a high compression ratio. Low bit rates on the order of 160-200 bits/s are achieved with very good quality of the reconstructed signal. The algorithm is compared with other transform based schemes.**

1 Introduction

In recent years, many algorithms for ECG data compression have been suggested [5]. ECG signal compression is very important for two reasons: effective storage and effective real time transmission. Transform methods produce smooth compressed signals and are insensitive to noise. In general, transform techniques involve expanding a signal into a set of basis functions and properly encoding the transform coefficients. The compression ratio (CR), the ratio of the number of bits needed to represent the signal and the compressed data, depends heavily on the set of basis functions. Ideally, the functions should adapt to the nonstationary behavior of ECG signals; and the information describing the way they vary in time should require as small number of bits as possible. We propose to use B-splines as the set of basis functions. The number and shape of these basis functions are completely characterized by the location of the so-called knots and their local properties can be changed by varying the position of the knots. By reducing the number of knots, we reduce the number of basis functions. The position of the knots can be efficiently coded using run-length coding. A number of authors have considered the problem in which not only the coefficients but also the position of the knots are optimized [3]. In conventional methods, after the initial choice of knots, their locations are varied in order to determine the positions that minimize the least-square error. This nonlinear least square problem is nonconvex; thus, we can only expect a local minimum in the vicinity of the initial location. The number of knots should be known a priori. The other problem is computational complexity. Each function evaluation in the minimization algorithm requires calculations of a fixed knot least-squares splines involving the solution of an overdetermined system of equations. In [7], a strategy based on eliminating the knots was proposed. In this paper, we introduce a new, and efficient approach for removing knots.

*This work is supported in part by the Academy of Finland, Project FAVI.

2 B-Splines

Let n and k be positive integers with $k \leq n$, and $\mathbf{t} = (t_i)_{i=1}^{n+k}$ be a sequence of real numbers. There are n B-splines $(B_{i,k,t})_{i=1}^{n}$, associated with the knot sequence, which are normalized to sum to unity. We denote by $S_{k,t}$ the linear space of functions spanned by these B-splines, and we refer to elements of $S_{k,t}$ written in terms of this basis, as B-spline functions or simply splines. A function $f(x)$ in $S_{k,t}$ with coefficients $a_1, ..., a_n$ can be written as

$$f(x) = \sum_{i=1}^{n} a_i B_{i,k,t}(x). \tag{1}$$

In our algorithm we use $k = 4$.

3 Compression by knot removal

The idea behind our data reduction method is to start with an initial spline approximation given by a large number of knots. The knots are removed one by one without perturbing the initial spline by more than a given tolerance. The final number and location of these knots are determined automatically.

Let $f(x)$, an element of $S_{k,t}$, be the initial spline approximation to a given data, and ε a given nonnegative real number. We want to determine a subspace $S_{k,\tau}$ of $S_{k,t}$ of the lowest possible dimension and an element $g(x)$ of $S_{k,\tau}$ such that

$$\max_x |f(x) - g(x)| < \varepsilon. \tag{2}$$

Let $\boldsymbol{\tau}^l$ denote a subsequence of $\mathbf{t}$ obtained by removing l knots and by $g_l(x)$ an approximation to $f(x)$ defined on $\boldsymbol{\tau}^l$. Our strategy to determine $g(x)$ involves three main steps:

1. Initialize: $g_0(x) = f(x)$ and $\boldsymbol{\tau}^0 = \mathbf{t}$;
2. Decide which knot should be removed from $\boldsymbol{\tau}^l$ and create a new subsequence $\boldsymbol{\tau}^{l+1}$;
3. Compute the approximation g_{l+1} to the spline f.

The iterations are stopped when

$$\max_x |f(x) - g_{l+1}(x)| \geq \varepsilon, \tag{3}$$

and the approximation $g_l(x)$ is the output $g(x)$. The spline approximation $g_l(x)$ at each iteration is computed using the least-square method. Our algorithm can handle multiple knots at the same location. Removing knots one by one allows us to easily update the approximation to the initial spline and avoids problems related to interaction between neighboring knots.

3.1 Least-squares approximation with splines

The ECG signal is divided into intervals. In each of these, a spline approximation is used. To further improve compression ratio and avoid discontinuities, the first and the last point in each interval should be interpolated not approximated. Let us assume that we have a given set of data values y_j corresponding to data points x_j, $j = 1, \ldots, N$, with $x_1 = 0$, $x_N = b$. We want to find a spline $g_l(x)$ of degree k with given knots $\tau_1^l, \ldots, \tau_{m+k}^l$ such that the expression

$$\sum_{j=2}^{N-1} (y_j - g_l(x_j))^2, \tag{4}$$

where

$$g_l(x_j) = \sum_{i=1}^{m} c_i^l B_{i,k,\tau^l}(x_j) \tag{5}$$

is minimized subject to

$$g_l(x_1) = y_1, \quad g_l(x_N) = y_N. \tag{6}$$

If we choose coincident boundary knots $x_1 = \tau_1^l = \ldots = \tau_k^l$, $\tau_{m+1}^l = \ldots = \tau_{m+k}^l = x_N$ and replace the original data by

$$y_j' = y_j - \frac{y_N - y_1}{x_N - x_1}, \tag{7}$$

we will get $g_l(x_1) = c_1^l$, $g_l(x_N) = c_m^l$, $y_1' = 0$ and $y_N' = 0$. Thus, conditions (6) are satisfied when $c_1^l = 0$ and $c_m^l = 0$ and the first $B_{1,k,\tau^l}(x)$ and the last $B_{m,k,\tau^l}(x)$ basis functions can be removed from the representation of spline $g_l(x)$. Instead of Eq. (4) we can minimize

$$\sum_{j=2}^{N-1} \left(y_j' - \sum_{i=2}^{m-1} c_i^l B_{i,k,\tau^l}(x_j) \right)^2. \tag{8}$$

Therefore, our problem reduces to the determination of the B-spline coefficients c_i^l, $i = 2, \ldots, m-1$, as the solution, in the least-square sense, of the overdetermined linear system

$$\mathbf{E}_l \mathbf{c}_l = \mathbf{y}', \tag{9}$$

where

$$\mathbf{E}_l = \begin{bmatrix} B_{2,k,\tau^l}(x_2) & \ldots & B_{m-1,k,\tau^l}(x_2) \\ \vdots & & \vdots \\ B_{2,k,\tau^l}(x_{N-1}) & \ldots & B_{m-1,k,\tau^l}(x_{N-1}) \end{bmatrix}, \tag{10}$$

$$\mathbf{y}' = \begin{bmatrix} y_2' \\ \vdots \\ y_{N-1}' \end{bmatrix}, \quad \mathbf{c}_l = \begin{bmatrix} c_2^l \\ \vdots \\ c_{m-1}^l \end{bmatrix}. \tag{11}$$

To determine the least-square solution of the system (9) we use the orthogonalization method. Let us assume that we have QR factorization of the matrix $\mathbf{E}_l$

$$\mathbf{E}_l = \mathbf{Q}_l \mathbf{R}_l \tag{12}$$

and

$$\mathbf{R}_l = \begin{bmatrix} \mathbf{R}_{l1} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{z}_l = \begin{bmatrix} \mathbf{z}_{l1} \\ \mathbf{z}_{l2} \end{bmatrix} = \mathbf{Q}_l^T \mathbf{y}', \tag{13}$$

with $\mathbf{z}_{l1}$ having $m-2$ elements and $\mathbf{R}_{l1}$ is an upper triangular matrix of order $m-2$. The vector of spline coefficients $\mathbf{c}_l$ is the solution of the triangular system

$$\mathbf{R}_{l1} \mathbf{c}_l = \mathbf{z}_{l1} \tag{14}$$

which can be obtained by simple backsubstitution. The reason why we chose this method is that the QR factorization of the matrix $\mathbf{E}_{l+1}$ can be very efficiently obtained by updating the factorization of $\mathbf{E}_l$. The matrix $\mathbf{E}_{l+1}$ consisting of the new basis functions $B_{i,k,\tau^{l+1}}$ defined for the knot sequence $\boldsymbol{\tau}^{l+1} = \boldsymbol{\tau}^l \backslash \{\tau_p^l\}$ can be obtained from the matrix $\mathbf{E}_l$ [2]:

$$\mathbf{E}_{l+1} = \mathbf{E}_l \mathbf{B}. \tag{15}$$

From Eqs. (12) and (15)

$$\mathbf{E}_{l+1} = \mathbf{Q}_l \mathbf{R}_l \mathbf{B} = \mathbf{Q}_l \tilde{\mathbf{R}}. \tag{16}$$

The columns of the matrix $\tilde{\mathbf{R}}$ are either unchanged columns of the upper triangular matrix $\mathbf{R}_l$ or a linear combination of two consecutive columns of this matrix. Thus the matrix $\tilde{\mathbf{R}}$ is upper Hessenberg and the unwanted subdiagonal elements can be zerod by a sequence of Givens rotations of row vectors. The matrix $\mathbf{E}_l$ has the band structure with at most k non-zero adjacent elements in a row. As a consequence $\mathbf{R}_l$ and $\tilde{\mathbf{R}}$ have also the band strucure with k and $k+1$ adjacent non-zero elements, respectively. This further reduces the computational complexity because the Givens rotation can now be restricted to the vectors of length $k+1$. Matrix $\mathbf{Q}_l$ does not have to be updated. The vector $\mathbf{z}_{l+1}$ can be obtained by applying Givens transformations to the corresponding elements of $\mathbf{z}_l$ ($\mathbf{z}_l$ can be considered as an extra column of $\tilde{\mathbf{R}}$).

3.2 Ranking the knots

Next we describe the algorithm which allows us to estimate the significance of each interior knot in representing the spline function g_l given by Eq. (5) and decide which knot should be removed. Let us remove knot $z = \tau_p^l$. The function $h(x)$ defined on $\boldsymbol{\rho} = \rho_1, \ldots, \rho_{m+k-1} = \boldsymbol{\tau}^l \backslash \{z\}$ is given by:

$$h(x_j) = \sum_{i=1}^{m-1} d_i B_{i,k,\rho}(x_j), \quad j = 1, \ldots, N. \tag{17}$$

A natural approach is to let the distance between g_l and h, be the measure of the significance of the knot τ_p^l in representing spline g_l:

$$\min_{d_1,\ldots,d_{m-1}} \max_j |h(x_j) - g_l(x_j)|. \tag{18}$$

Only a rough estimate of the relative importance of the knots is needed. Thus, to avoid extensive computations, we are use approximate solution to the problem of minimizing Eq. (18). First we represent h using the same basis B_{i,k,τ^l} as function g_l:

$$h(x_j) = \sum_{i=1}^{m} \tilde{d}_i B_{i,k,\tau^l}(x_j), \quad j = 1, \ldots, N, \tag{19}$$

where

$$\tilde{d}_i = \begin{cases} d_i & \text{if } i \le p-k \\ \alpha_i d_i + (1-\alpha_i) d_{i-1} & \text{if } p-k+1 \le i \le p-1 \\ d_{i-1} & \text{if } p \le i \le m \end{cases} \tag{20}$$

$$\alpha_i = \frac{z - \rho_i}{\rho_{i+k-1} - \rho_i}. \tag{21}$$

Let us denote by $\tilde{\mathbf{d}} = \phi(\mathbf{d})$ the transformation $(d_i)_{i=1}^{m-1} \to (\tilde{d})_{i=1}^{m}$. We can find such transformation ϕ that all coefficients $c_i = \tilde{d}_i$ for all i exept $i = i_0$ for some i_0. From Eq. (20) we immediately have

$$d_i = \begin{cases} c_i^l & \text{if } i \le p-k \\ c_{i+1}^l & \text{if } p-1 \le i \le m-1 \end{cases} \tag{22}$$

The rest of the coefficients are computed as follows:

$$d_i = \frac{c_i^l - (1-\alpha_i) d_{i-1}}{\alpha_i} \quad i = p-k+1, \ldots, p-2 \tag{23}$$

or

$$d_{i-1} = \frac{c_i^l - (1-\alpha_i d_i)}{1-\alpha_i} \quad i = p-1, p-2, \ldots, p-k+2. \tag{24}$$

Using coefficients computed by Eq. (22) and Eq. (23) we have $\tilde{d}_i = c_i^l$ for all i but $i = p-1$ and by Eq. (22) and Eq. (24) $\tilde{d}_i = c_i^l$ for all i but $i = p-k+1$. The maximum error between g_l and h is given, respectively, by

$$\varepsilon_p^1 = \max_j |(c_{p-1}^l - \tilde{d}_{p-1}) B_{p-1,k,\tau^l}(x_j)| \tag{25}$$

and

$$\varepsilon_p^2 = \max_j |(c_{p-k+1}^l - \tilde{d}_{p-k+1}) B_{p-k+1,k,\tau^l}(x_j)|, \tag{26}$$

where $\tilde{d}_{p-1}$, $\tilde{d}_{p-k+1}$ are given by Eq. (20). The weight $w_p = \min(\varepsilon_p^1, \varepsilon_p^2)$ is assigned to the knot τ_p^l. This weight is a measure of the significance of this knot in representing the spline g_l. This procedure is repeated for all interior knots τ_p^l, $p = k+1, \ldots, m$. The knot with the smallest weight w_p is removed from $\boldsymbol{\tau}^l$.

To improve the computational efficiency first the linear approximation to the data is found, which requires only a small amount of computing time for each data point. The objective of this stage is to produce an approximation that well represents the details in the data, but might lack smoothness and has too many parameters. The output is treated as a linear spline $f = \sum_i a_i B_{i,2,t_1}$ which is subsequently represented as a cubic spline $f = \sum_i c_i B_{i,4,t_2}$ with triple knots. Then the knot removal is performed on f to obtain the final approximation.

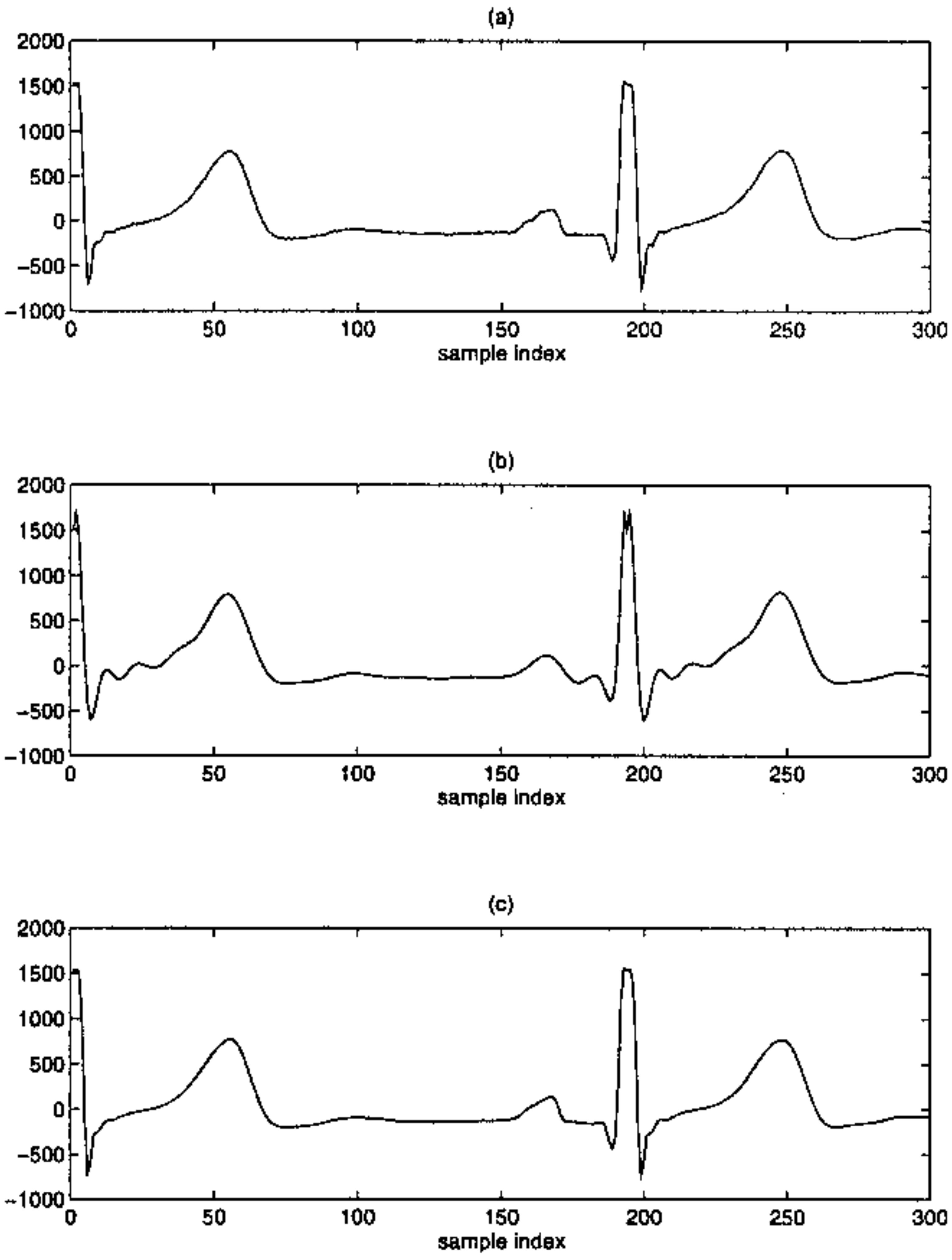

Figure 1: (a)Original ECG (200Hz). Its reconstruction using (b) 30 quantized coefficient of DLT and (c) 20 quantized coefficient of B-splines.

4 Compression of ECG

We use the fact that the ECG is a quasi-periodic signal with strong correlation between adjacent beats. The detection of P and T waves is not easy or sometimes not even feasible; thus, we chose the R-R interval to be the repetitive period. For the first interval, we find the knot sequence and compute coefficients of the approximating spline function using the procedure described in Section 3. For other intervals the knot sequence from the previous interval is used to compute approximation to the current data and the MSE is compared with the MSE for the interval for which the sequence was originally found. Only when the error increases more than some predefined threshold we search for a new knot sequence. This reduces not only the computation time but also the amount of overhead data. The coefficients c_i of the spline function are further

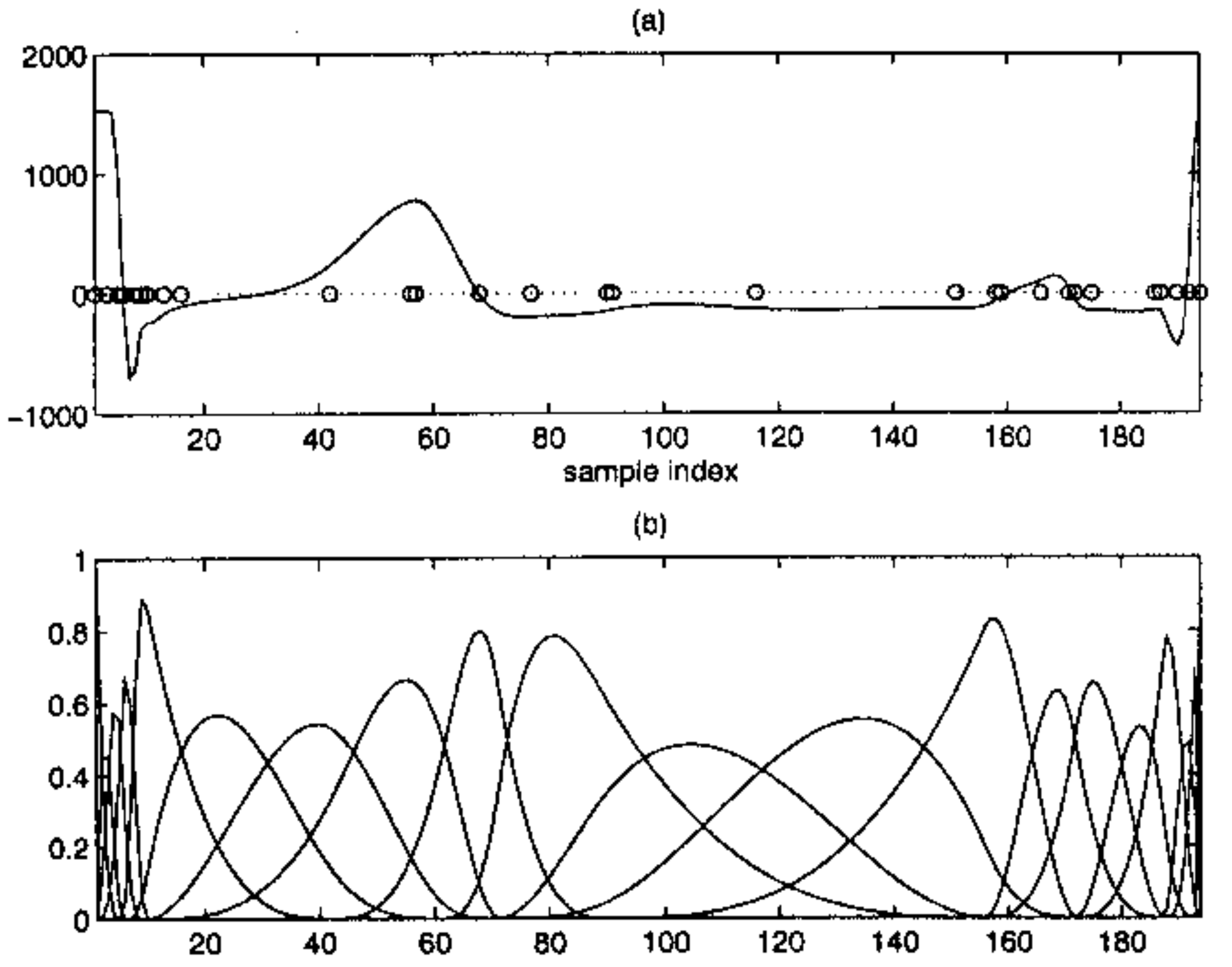

Figure 2: (a) The location of knots (marked as dots) for the R-R interval and (b) the resulting B-spline basis functions.

quantized, using a uniform quantization scheme. For simplicity, the sequence of knots is run-length coded. The coefficients in R-R intervals are usually strongly correlated. Thus, the CR can be further improved by quantizing the difference between the coefficients in the previous and current interval.

5 Results and discussion

Our method was tested using 10 non-standard ECG records (200Hz with 12 bits/sample resolution) and 10 records from MIT-BIH Arrhythmia database [8] (360Hz with 11 bits/sample resolution). Each record was 30 s in duration. The compression ratio (CR) is defined as the total number of bits in the original signal, divided by the number of bits needed to store the coefficients and overhead information, such as interval length or position of knots. As an objective measure of "goodness" we use percentage root-mean-square difference (PRD) calculated as

$$PRD = 100\sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i y_i^2}} \tag{27}$$

where y_i is the original ith sample and $\hat{y}_i$ is the reconstructed sample. We compare our method with discrete Legendre transform DLT [9], where it was shown that the CR for DLT is almost twice that of the discrete cosine transform (DCT) with the same reconstruction error. To make the comparison easier, the number N of basis function is set a priori and the same method has been used to code the coefficients in both algorithms. Tables 1 and 2 list the mean value of CR and PRD. The spline compression with 20 coefficients is comparable to the DLT with 40 coefficients as far as PRD is concerned, but the CR is about 60% higher. Fig.1 shows an original ECG, its approximation by 30 basis functions of DLT and by 20 B-spline basis functions. Very annoying ringing effect can be seen in the DLT reconstruction; while, the reconstruction

	SPLINES		DLT	
N	CR	PRD	CR	PRD
50	5.83	2.52	5.82	2.61
40	7.07	2.72	**7.19**	**4.26**
30	9.02	2.94	9.39	7.05
20	**12.77**	**4.27**	13.61	16.82

Table 1: Mean values of CR, PRD for ECG sampled at 200Hz with 12 bit/sample resolution.

	SPLINES		DLT	
N	CR	PRD	CR	PRD
50	11.39	3.02	12.73	4.27
40	14.01	3.11	**15.63**	**5.09**
30	17.36	3.47	19.86	7.67
20	**24.56**	**5.16**	29.96	17.16

Table 2: Mean values of CR, PRD for ECG sampled at 360Hz with 11 bit/sample resolution.

by spline approximation is still very good. The placement of the knots and the resulting basis functions for the first R-R interval of the ECG signal in Fig.1 is depicted in Fig. 2. We see that the local frequency of these basis functions resembles well the nonstationary behavior of the ECG signal. The quality of the ECG data compressed by DLT with less than 40 coefficients deteriorates rapidly. However, with splines ECG can be represented accurately using only 20 coefficients. This corresponds to a bit rate of 160 – 200 bits/s. Thus, higher CR can be achieved with spline approximation since it requires less basis functions for the same signal quality.

References

[1] C. de Boor, *A Practical Quide to Splines*, Applied Mathematical Science, Vol. 27. Springer-Verlag, 1978.

[2] W. Boehm, "Inserting new knots into B-spline curves," *Computer Aided Design*, vol. 12, pp. 199-201, 1980

[3] P. Dierckx, *Curve and Surface Fitting with Splines*. New York: Oxford University Press Inc., 1993

[4] G.H. Golub and C.F. van Loan, *Matrix Computations*, Baltimore: Johns Hopkins University Press, 1989

[5] S.M.S. Jalaleddine, C.G. Hotchens, R.D. Strattan, W.A. Coberly, "ECG Data Compression Techniques - A Unified Approach," *IEEE Trans. Biomed. Eng.* Vol. 37, pp. 329-343, 1990

[6] M. Karczewicz, M. Gabbouj and J. Astola,"Robust B-Spline Image Smoothing," accepted to *IEEE International Conference on Image Processing*, Austin, USA, November 1994

[7] T. Lyche and K. Morken, "A data-reduction strategy for splines with applications to the approximation of functions and data," *IMA Journal of Numerical Analysis*, vol. 8, pp. 185-208, 1988

[8] Massachusetts Institute of Technology, *The MIT-BIH Arrhythmia Database*, August 1980

[9] W. Philips and G. D. Jonghe, "Data Compression of ECG's by High-Degree Polynomial Approximation," *IEEE Trans. Biomed. Eng.* vol. 39, pp. 330-337, 1992