ONCONET: ORIGINAL ARTICLE
UDC: 616.12-073.97:681.322

Arch Oncol 2004;12(3):152-8.

Predictive ECG coding using linear time-invariant models

Aleksandar Bošković[1], Miroslav Despotović[2], Dragana Bajić[2]

[1]JP "Elektrovojvodina", Novi Sad, Serbia and Montenegro, [2]Faculty of Engineering, University of Novi Sad, Novi Sad, Serbia & Montenegro; Address correspondence to: Aleksandar Bošković, JP "Elektrovojvodina", Novi Sad, Serbia & Montenegro, E-mail : aleksandar.boskovic@ev.co.yu; The manuscript was received: 11.06.2004, Provisionally accepted: 23.06.2004, Accepted for publication: 23.06.2004

ABSTRACT

Electrocardiogram (ECG) signal compression suffers of lack of standards for analogue-digital conversion. Results of this study have shown that 8 bits/sample, although frequently in use, does not satisfy quality criteria for medical doctors. This paper also presents predictive technique for lossless ECG compression using linear time-invariant models. Tests on clinically measured ECG signals confirm a very good performance in terms of compression ratio.

KEY WORDS: *Electrocardiography; Signal Processing, Computer Assisted; Analog-Digital Conversion; Linear Models*

INTRODUCTION

A typical ECG signal requires less storage capacity than for medical images. However, ECG monitor devices are used frequently, even on a regular medical exams, while the medical images are recorded only when is necessary. Nowadays, electrocardiograph represents a necessary diagnostic device. Consequently, large amounts of data have to be stored. A need for efficient coding of ECG signals is continually increasing with modern use of long-term monitoring and telemedicine. Modern medical telemetry systems with low bit-rate channels require signal compression for efficient functioning.

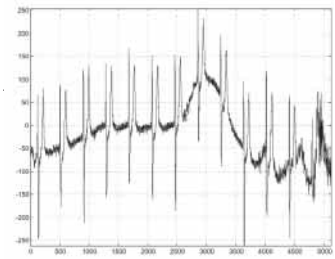

Figure 1. ECG signal

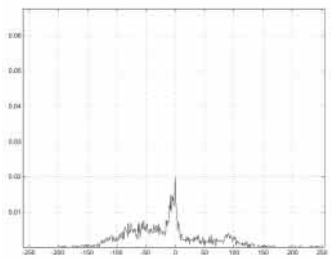

Figure 2. Probability distribution of ECG signal

For example, high-resolution electrocardiogram monitoring device, records 12-chanells ECG with 11-bit resolution and sampling rate of 1000 samples per second, and therefore it generates over 56 MB per hour (about 1360 MB per day), or charges the network with constant flow of 132 kb/s. Thus, the ECG signal compression is not only desirable and useful, but also necessary.

Signal compression methods fall into two common categories: lossy and lossless. Lossless ECG compression is essential for storage and transmission of electrocardiographs. The purpose of ECG compression should not be only to transmit or store the signal with fewer bits, but also to preserve the clinically significant information. According to the law regulations in many countries, medical signals after lossy compression cannot be used in diagnostics.

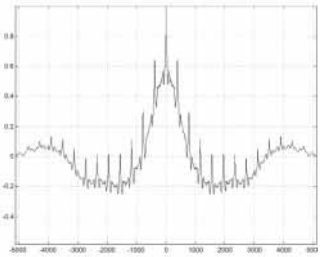

Figure 3. Covariance of ECG signal

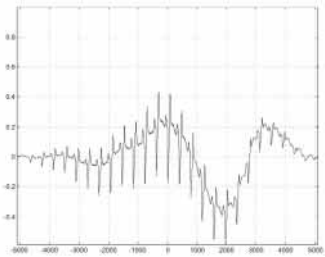

Figure 4. Cross-covariance of ECG signals

Predictive methods are a subclass of the lossless techniques. These methods exploit redundancy between samples, beats and leads of ECG signal, so that only new information has to be coded. Correlation of ECG samples is significant, and it is illustrated in Figures 3. and 4. In this paper, predictive techniques for lossless ECG compression are introduced. In combination with entropy coding, they achieve very good results.

The Massachusetts Institute of Technology and Beth Israel Hospital (MIT-BIH) ECG Compression Test Database and the MIT-BIH Arrhythmia Database were used for testing the different compression methods. The MIT-BIH ECG Compression Test Database contains 168 short ECG two-lead recordings (20.48 seconds each). The recordings were digitized at 250 samples per second per lead with 12-bit resolution over a 10 mV range. The MIT-BIH Arrhythmia Database contains 48 half-hour excerpts of two-lead ambulatory ECG recordings. The recordings were digitized at 360 samples per second per lead with 11-bit resolution over a 10 mV range.

The amount of compression is often expressed with the compression ratio (CR)

$$CR = \frac{b_{orig.}}{b_{comp.}}$$

that is defined as the ratio between the bit rate of the original ECG signal and the bit rate of the compressed one. In this way, compression ratio shows how much the ECG data is being compressed compared to the original.

In addition, possible amount of compression could be estimated based on the entropy. Entropy of ECG signal is defined as the average information per symbol:

$$H = -\sum_{y} p_y \log_2 p_y \left[bit / symbol \right]$$

where p_y is probability of symbol y. Entropy tells us what is the average number of bits required to represent a sample of ECG signal.

Krzyszof Duda, Pawel Turcza and Tomasz P. Zielinski (1) published a comparison of medium, maximum and minimum compression ratios for different lossless ECG compression methods. The results are presented in Table 1. For testing the methods, a MIT-BIH Arrhythmia Database has been used.

Table 1. Comparison of medium, maximum and minimum compression ratios for different methods

	CRmean	CRmax	CRmin
DPCM, VL-no, EE=0	2.772	3.333	2.201
DPCM, VL-no, OEE=1	2.541	3.057	1.986
DPCM, VL-yes, OEE=0	2.861	3.463	2.223
DPCM, VL-yes, OEE=1	3.093	3.682	2.366
LWT, VL-no, OEE=0	2.990	2.593	2.434
LWT, VL-no, OEE=1	2.710	3.432	2.099
LWT, VL-yes, OEE=0	3.135	3.709	2.525
LWT, VL-yes, OEE=1	3.114	3.648	2.429
ALWT, VL-no, OEE=0	3.009	3.605	2.437
ALWT, VL-no, OEE=1	2.732	3.442	2.124
ALWT, VL-yes, OEE=0	3.156	3.722	2.528
ALWT, VL-yes, OEE=1	3.131	3.655	2.431
SPb, VL-no, OEE=0	2.964	3.547	2.420
SPb, VL-no, OEE=1	2.732	3.419	2.141
SPb, VL-yes, OEE=0	3.069	3.621	2.485
SPb, VL-yes, OEE=1	3.073	3.583	2.403
SPc, VL-no, OEE=0	3.009	3.578	2.521
SPc, VL-no, OEE=1	2.777	3.458	2.193
SPc, VL-yes, OEE=0	3.131	3.675	2.613
SPc, VL-yes, OEE=1	3.126	3.617	2.510

DPCM – differential pulse code modulation, LWT – lifting wavelet transform, ALWT – adaptive lifting wavelet transformation, SPb – S+P transform with predictor B, SPc – S+P transform with predictor C, OEE – order of entropy encoder, VL – MS-VLI

ELECTRICAL ACTIVITY OF THE HEART

The cardiovascular system is a circular system that propels oxygen and nutrients to the body cells, and picks up wastes and other substances. Cardiovascular system is made up of heart, blood and blood vessels.

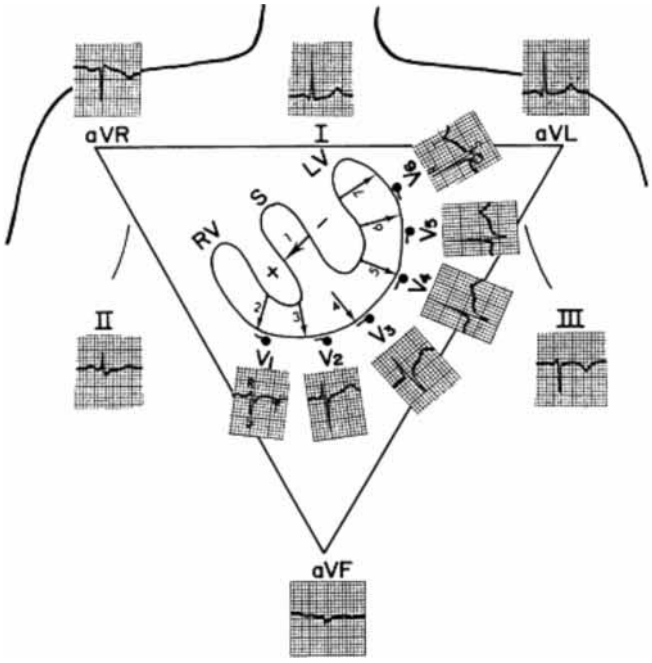

Figure 5. The standard 12-leads ECG

The heart is the transport system pump, the delivery routes are the blood vessels, and blood is the transport medium (2).

The intrinsic conduction system sets the basic rhythm of the beating heart. It consists of autorhythmic cardiac cells that initiate and distribute electrical impulses throughout the heart. The coordinated contractions of the heart result from electrical changes that take place in cardiac cells. Cardiac cells are capable of generating and conducting electrical impulses that are responsible for the contraction and relaxation of myocardial cells.

The electrocardiogram - ECG - is a graphic representation of the heart's electrical activity, formed as the cardiac cells depolarize and repolarize. Electrodes attached to the skin may detect this electrical activity. The cardiac electrical activity is recorded through 12 surface leads named I, II, III, aVR, aVL, aVF, V_1, V_2, V_3, V_4, V_5, and V_6.

The electrodes recording the first six are located on the limbs, while leads V1 to V6 are recorded from the chest (Figure 5).

PARAMETERS FOR DIGITALIZATION OF ECG SIGNAL

ECG signal compression requires existence of a digitalized signal. The process of digitalization is called analogue-digital conversion. During the analogue-digital conversion, levels of signal are sampled and converted to a discrete bit pattern. Each sample (... 1.00V, -1.34V ...) is represented with the binary word of prescribed length (... 00000001, 10000101 ...). Consequently, digital signal approximates an analog one. Basic parameters of digitalization are sampling rate and number of quantization levels. Number of quantization levels depends on number of bits (resolution) that are used for representing the sample.

Thus, it is obvious that the analogue-digital conversion essentially presents a lossy compression. Example of analog and digitalized ECG signal is illustrated in Figures 6 and 7, respectively.

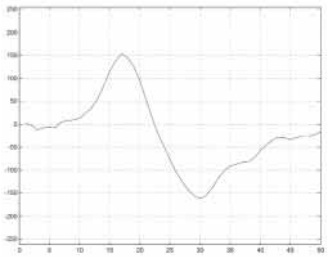

Figure 6. Segment of ECG signal

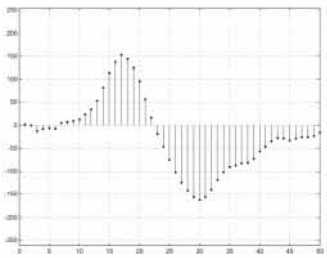

Figure 7. Digitalized segment of ECG signal

Paradoxical, standard for ECG digitalization does not exist. Commercial devices use sampling rate from 125 up to 500, and they can even use 1000 samples per second. Resolution of digitalization can be from 8 up to 12 bits. Thus, bit-rate can vary from 1 Kb/s up to 12Kb/s. Since the compression occurs after the analogue-digital conversion, it is very important what the initial parameters are.

Therefore, comparing the compression methods is not explicit job, because it does not include signals digitalized in the same way. Additionally, there is no standard comparison method, because there are several measures that express the compression gain: compression ratio, number of bits per sample and number of bits per second.

Obviously, comparison of ECG compression methods is difficult, but possible - general review is given in the paper of Jalaleddine, Hutchens, Strattan and Coberly, [6] 1990, but up to date methods are not included. Problem is discussed in the papers of Miroslav Despotovic [4] and Dragana Bajic [5], but only unified databases like MIT-BIH ECG Compression Test Database and MIT-BIH Arrhythmia Database allow objective comparison. Quantitative measures of quality for ECG digitalized signal could be acceptable or not for medical doctors, so we performed a little survey to recognize a medical expert opinion on parameters for digitalization of ECG signals. This was done in effort to have a reliable indi-

cator of ECG quality that preserves important details significant for correct diagnosis.

Initially, for this purpose, we have chosen a set of ECG signals with the following characteristic events: premature atrial contraction, premature ventricular contraction, paced rhythm, atrial bigeminy, ventricular bigeminy, ventricular trigeminy, ventricular tachycardia, atrial fibrillation, ventricular fibrillation, atrial flutter, ventricular flutter and first-degree AV block.

With the assistance of medical doctors, characteristic subset of this set was chosen for the survey. ECG records of this subset are:

- ECG record of female person, 75 years old, on whose record premature atrial contraction is occurred;
- ECG record of female person, 24 years old, on whose record ventricular bigeminy is occurred;
- ECG record of male person, 64 years old, on whose record atrial fibrillation is occurred.

After that, each of these signals, in original form with 360 samples per second, 11 bits per sample (I), is processed and for each one, three new signals are generated (Appendix A):

- 240 samples per second, 10 bits per sample (II),
- 180 samples per second, 9 bits per sample (III) and
- 180 samples per second, 8 bits per sample (IV).

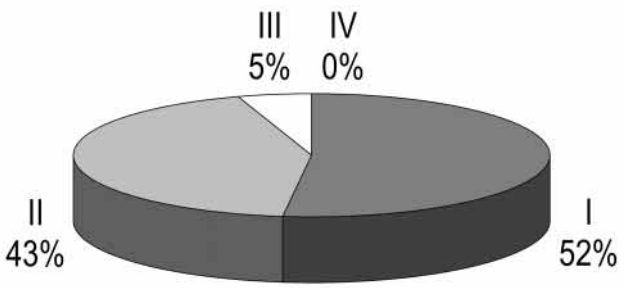

Figure 8. Questionnaire results

Five medical doctors were interviewed. In general, doctors rejected signal IV (8 bits per sample), although this type of conversion is used in numerous commercial ECG devices. They were satisfied with other signals and they did not make a big difference. However, signal I was their favorite (Figure 8).

According to these results, it would be appropriate to use analogue-digital conversion with high sampling rate and a higher number of quantization levels. Results indicate that the signal IV does not satisfy medical needs, but signal II is as good as signal I. Choice of signal II parameters reduces required number of bits to 60% of case I. This is a significant reduction of required resources.

ECG SIGNAL PREDICTION

The fundamental principle of predictive techniques is to use previously coded samples to generate a predicted value for the current sample.

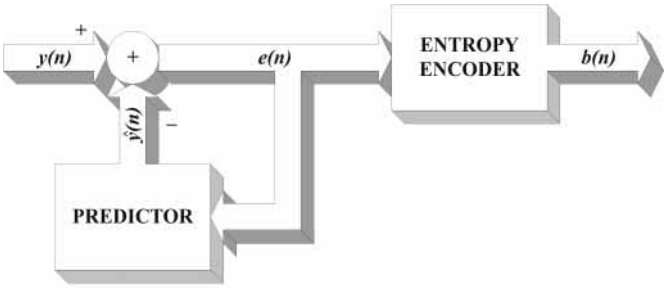

Figure 9. ECG signal compression

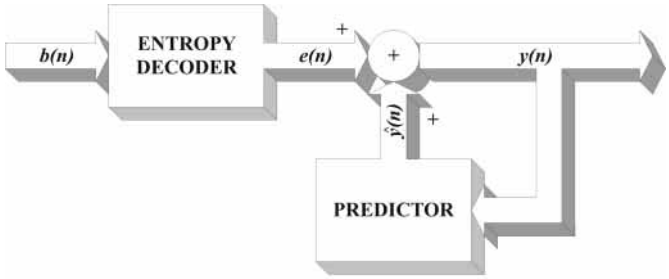

Figure 10. ECG signal decompression

This follows by coding the difference between the predicted and the actual values of the current sample. During a slow motion period, most consecutive samples will be similar, whereas the difference will be greater when the motion increases. Assumption is that entropy coding of ECG signal prediction error will achieve better compression ratio then entropy coding of original ECG signal. ECG signal compression and decompression scheme is illustrated in Figure 9 and 10.

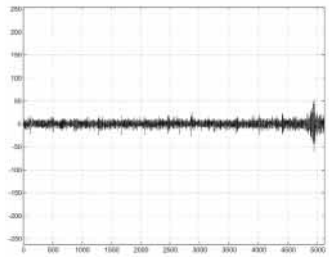

Figure 11. ECG signal prediction error

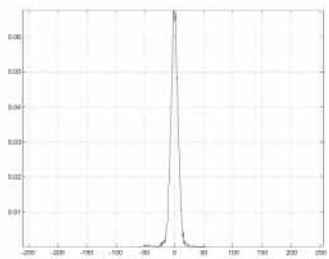

Figure 12. Probability distribution of ECG signal prediction error

Now, consider the ECG signal *y(n)*, which has been coded up to sample *n* - 1, and let *y(n - 1)*, *y(n - 2)*,... be the values of the ECG signal up to that moment. Then, an estimate of ECG signal is predicted from previously samples:

$$\hat{y}(n) = \psi(y(n-1), y(n-2), \ldots, u(n-1), u(n-2), \ldots)$$

where Ψ denotes the prediction rule.

Having the ECG signal prediction it is enough to code prediction error *ε(n)* that is illustrated in Figure 11. A reasonable prediction criterion is to minimize the prediction error probability. Probability distribution of prediction error is depicted in Figure 12. It is obvious that amplitudes of the prediction error are less then amplitudes of the original ECG signal (Figures 1 and 11). In addition, prediction error signal has significantly smaller range of values making it more suitable for compression, so consequently, it requires fewer resources for storage and transmission.

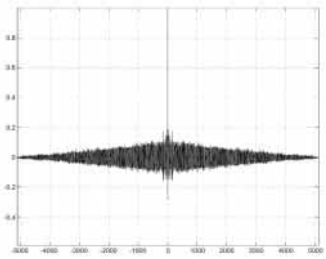

Figure 13. Covariance of ECG signal prediction error

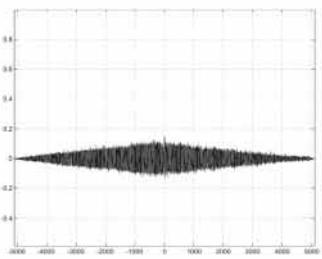

Figure 14. Cross-covariance of ECG signal prediction error

Prediction error represents unexpected component of ECG signal. Coding process continues recursively in this manner. Correlation of prediction error samples, illustrated on Figures 13 and 14, is much less then correlation of original ECG samples.

MODELS OF LINEAR TIME-INVARIANT SYSTEMS

One way of ECG signal prediction is to characterize ECG signal lead as the output of linear system which inputs are other leads of the same signal. In the following analysis, we shall discuss single-input single-output systems, because they correspond to ECG signals from test databases. Generalization to multiple-input system is straightforward. Time-invariant linear systems are the most important class of dynamical systems. System is time-invariant if its response to a certain input signal does not depend on absolute time. System is linear if its output response to a linear combination of inputs is the same linear combination of the output responses of the individual inputs. System is causal if the output at a certain time instant depends on the input up to that time instant only.Consider a system with an input signal *u(n)*, output signal *y(n)*, and disturbance $\nu(n)$ (Figure 15).

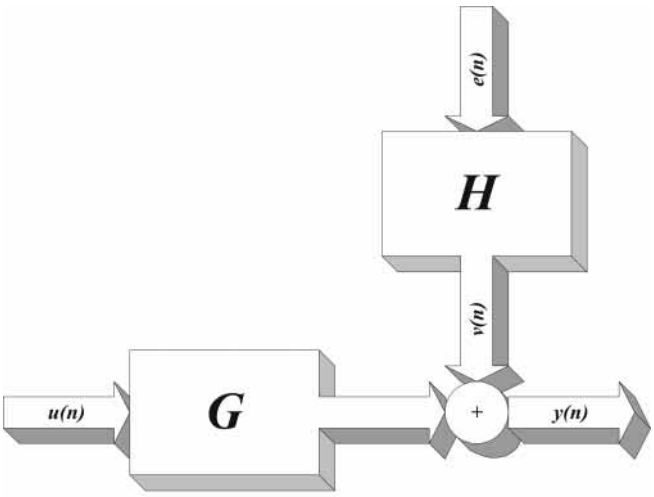

Figure 15. Dynamical system with disturbance

Linear time-invariant causal system can be described by its impulse response *g(k)* as follows:

$$y(n) = \sum_{k=1}^{\infty} g(k)u(n-k) + \nu(n) \qquad n = 0,1, \ldots$$

The impulse response is a complete characterization of the system.

Let the disturbance $\nu(n)$ be given as

$$\nu(n) = \sum_{k=0}^{\infty} h(k)e(n-k) \qquad n = 0,1, \ldots$$

where *e(n)* is the white noise with zero mean and variance λ.

This description is versatile enough for most practical purposes. Then, the disturbance $\nu(n)$ has also zero mean, and the covariance

$$R_\nu(n) = \lambda \sum_{k=0}^{\infty} h(k)h(k-n)$$

It will be convenient to introduce a shift operator *q* of a certain signal *s(n)* as follows:

$qs(n) = s(n+1)$, $q^{-1}s(n) = s(n-1)$.

Then, linear time-invariant causal system could be described as

$y(n) = G(q)u(n) + H(q)e(n)$,

where *G(q)* is transfer function of linear system.

Let filter *H(q)* be monic *h(0)* = 1, stable

$\sum |h(k)| < \infty$, and invertible $\quad \nu(n) = \tilde{H}(q)\nu(n) = \sum_{k=0}^{\infty} \tilde{h}(k)\nu(n-k)$

with $\quad \sum_{k=0}^{\infty} |\tilde{h}(k)| < \infty, \tilde{H}(q) = H^{-1}(q)$

Since *H* is monic and invertible, and white noise *e(n)* has zero mean, the conditional expectation of disturbance $\nu(n)$ is

$$\hat{\nu}(n) = \sum_{k=1}^{\infty} h(k)e(n-k) = \left[1 - H^{-1}(q)\right]\nu(n) = \sum_{k=1}^{\infty} -\tilde{h}(k)\nu(n-k) \Leftrightarrow H(q)\hat{\nu}(n) = \left[H(q) - 1\right]\nu(n)$$

Then, the conditional expectation of output signal *y(n)* is given by

$$\hat{y}(n) = G(q)u(n) + \hat{\nu}(n) = G(q)u(n) + \left[1 - H^{-1}(q)\right]\nu(n) = G(q)u(n) + \left[1 - H^{-1}(q)\right]\left[y(n) - G(q)u(n)\right]$$

Collecting the terms gives, (3)

$$\hat{y}(n) = H^{-1}(q)G(q)u(n) + \left[1 - H^{-1}(q)\right]y(n) \Leftrightarrow H(q)\hat{y}(n) = G(q)u(n) + \left[H(q) - 1\right]y(n)$$

The prediction error is given by

$$\varepsilon(n) = y(n) - \hat{y}(n) = -H^{-1}(q)G(q)u(n) + H^{-1}(q)y(n)$$

and represents that part of the output *y(n)* that cannot be predicted from past data. For this reason, it is also called the innovation at time n.

In the following discussion, several models of time-invariant linear system are presented.

AR model

The AR model for a single-output system with an output signal *y(n)* can be written as:

$$y(n) + a_1 y(n-1) +,,, + a_n y(n-n_a) = y(n) + \sum_{k=1}^{n_a} a_k y(n-k) =$$

$$y(n) + \left[\sum_{k=1}^{n_a} a_k q^{-k}\right] y(n) = e(n) \Leftrightarrow A(q)y(n) = e(n)$$

where

$$A(q) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} = 1 + \sum_{k=1}^{n_a} a_k q^{-k}$$

AR model parameter vector has form $\theta = \left[a_1 \ldots a_{n_a}\right]^T$

AR model predictor is given by difference equation,

$\hat{y}(n|\theta) = \left[1 - A(q)\right]y(n)$ and requires knowledge of $y(0)\ldots y(1-n_a)$

Prediction error of output signal is then given by $\varepsilon(n,\theta) = A(q)y(n)$

AR model state vector has a form $\varphi(n) = [-y(n-1)\ldots -y(n-n_a)]^T$

Since the calculation of the predictor from past data not depends on the parameter vector, predictor defines linear regression.

ARMA model

Output signal *y(n)* of the ARMA model linear system can be presented by linear difference equation:

$$y(n) + a_1 y(n-1) + \ldots + a_{n_a} y(n-n_a) = e(n) + c_1 e(n-1) + \ldots + c_{n_a} e(n-n_c) \Leftrightarrow$$

$$y(n) + \sum_{k=1}^{n_a} a_k y(n-k) = e(n) + \sum_{k=1}^{n_c} c_k e(n-k) \Leftrightarrow y(n) + \left[\sum_{k=1}^{n_a} a_k q^{-k}\right] y(n) =$$

$$e(n) + \left[\sum_{k=1}^{n_c} c_k q^{-k}\right] e(n) \Leftrightarrow A(q)y(n) = C(q)e(n)$$

ARMA model parameter vector has form $\theta = \left[a_1,,, a_{n_a} c_1 \ldots c_{n_c}\right]^T$

ARMA model predictor is given by difference equation

$$\hat{y}(n|\theta) = \left[1 - \frac{A(q)}{C(q)}\right] y(n) \Leftrightarrow C(q)\hat{y}(n|\theta) = \left[C(q) - A(q)\right]y(n)$$

and requires knowledge of $\hat{y}(0|\theta)\ldots\hat{y}(1-n_c|\theta), y(0)\ldots y(1-\max(n_a,n_c))$

Adding

$\left[1 - C(q)\right]\hat{y}(n|\theta)$ to both sides gives

$$\hat{y}(n|\theta) = \left[1 - A(q)\right]y(n) + \left[C(q) - 1\right]\left[y(n) - \hat{y}(n|\theta)\right] = \left[1 - A(q)\right]y(n) + \left[C(q) - 1\right]\varepsilon(n,\theta)$$

Prediction error of output signal is then given by

$$\varepsilon(n,\theta) = \frac{A(q)}{C(q)} y(n)$$

ARMA model state vector has a form

$$\varphi(n,\theta) = \left[-y(n-1)... - y(n-n_a)\varepsilon(n-1,\theta)...\varepsilon(n-n_c,\theta)\right]^T$$

Since the calculation of the predictor from past data depends on the parameter vector, predictor defines pseudo-linear regression.

ARX MODEL

ARX model is also known as "equation error model". The ARX model for a single-output system with input signal *u(n)* and output signal *y(n)* can be written as:

$$y(n) + a_1 y(n-1) + ... + a_{n_a} y(n-n_a) = b_1 u(n-1) + ... + b_{n_b} u(n-n_b) + e(n) \Leftrightarrow$$

$$y(n) + \sum_{k=1}^{n_a} a_k y(n-k) = \left[\sum_{k=1}^{n_b} b_k q^{-k}\right] u(n) + e(n) \Leftrightarrow A(q)y(n) = B(q)u(n) + e(n)$$

ARX model parameter vector has form $\theta = \left[a_1...a_{n_a} b_1...b_{n_b}\right]^T$

ARX model predictor is given by difference equation

$\hat{y}(n|\theta) = B(q)u(n)\left[1 - A(q)\right]y(n)$ and requires knowledge of

$y(0)...y(1-n_a), u(0)...u(1-n_b)$

Output signal prediction error is then given by $\varepsilon(n,\theta) = A(q)y(n) - B(q)u(n)$.

ARX model state vector has a form

$$\varphi(n) = \left[-y(n-1)... - y(n-n_a)u(n-1)...u(n-n_b)\right]^T$$

Since the calculation of the predictor from past data not depends on the parameter vector, predictor defines linear regression.

ARMAX MODEL

Another very common, and more general, model is the ARMAX model. The basic disadvantage of the ARX model is the lack of adequate freedom in describing the properties of disturbance. The disturbance can be described as a moving-average. Output signal *y(n)* of ARMAX model linear system can be presented by linear difference equation [3]:

$$y(n) + a_1 y(n-1) + ... + a_{n_a} y(n-n_a) =$$

$$b_1 u(n-1) + ... + b_{n_b} u(n-n_b) + e(n) + c_1 e(n-1) + ...c_{n_c} e(n-n_c) \Leftrightarrow$$

$$y(n) + \sum_{k=1}^{n_a} a_k y(n-k) = \sum_{k=1}^{n_b} b_k u(n-k) + e(n) + \sum_{k=1}^{n_c} c_k e(n-k) \Leftrightarrow$$

$$\left[\sum_{k=1}^{n_b} b_k q^{-k}\right] u(n) + e(n) + \left[\sum_{k=1}^{n_c} c_k q^{-k}\right] e(n) \Leftrightarrow A(q)y(n) = B(q)u(n) + C(q)e(n)$$

ARMAX model parameter vector has form $\theta = \left[a_1...a_{n_a} b_1...b_{n_b} c_1...c_{n_c}\right]^T$

ARMAX model predictor is given by difference equation

$$\hat{y}(n|\theta) = \frac{B(q)}{C(q)} u(n) + \left[1 - \frac{A(q)}{C(q)}\right] y(n) = B(q)u(n) + \left[C(q) - A(q)\right]y(n)$$

and requires knowledge of

$\hat{y}(0|\theta)....\hat{y}(1-n_c|\theta), y(0)...y(1-\max(n_c,n_a)), u(0)...u(1-n_b)$

Adding

$\left[1 - C(q)\right]\hat{y}(n|\theta)$ to both sides gives

$$\hat{y}(n|\theta) = B(q)u(n) + \left[1 - A(q)\right]y(n) + \left[C(q) - 1\right]\varepsilon(n,\theta)$$

Prediction error of output signal is then given by

$$\varepsilon(n,\theta) = \frac{A(q)}{C(q)} y(n) - \frac{B(q)}{C(q)} u(n)$$

ARMAX model state vector has a form

$$\varphi(n,\theta) = \left[-y(n-1)... - y(n-n_a)u(n-1)...u(n-n_b)\varepsilon(n-1,\theta)...\varepsilon(n-n_c,\theta)\right]^T$$

Since the calculation of the predictor from past data depends on the parameter vector, predictor defines pseudo-linear regression.

ARARX MODEL

ARARX model is also known as "generalized least squares model". Instead of describing the disturbance as a moving-average, it can also be described as an autoregression. Output signal *y(n)* of the ARARX model linear system can be presented by linear difference equation:

$$A(q)y(n) = B(q)u(n) + \frac{1}{D(q)} e(n)$$

$$D(q) = 1 + d_1 q^{-1} + ... + d_{n_d} q^{-n_d} = 1 + \sum_{k=1}^{n_d} d_k q^{-k}$$

ARARX model parameter vector has form $\theta = \left[a_1...a_{n_a} b_1...b_{n_b} d_1...d_{n_d}\right]^T$

ARARX model predictor is given by difference equation

$$\hat{y}(n|\theta) = D(q)B(q)u(n) + \left[1 - D(q)A(q)\right]y(n)$$

and requires knowledge of $y(0)...y(1-n_a-n_d), u(0)...u(1-n_b-n_d)$

Prediction error of output signal is then given by

$$\varepsilon(n,\theta) = D(q)\left[A(q)y(n) - B(q)u(n)\right]$$

It is convenient to introduce the auxiliary variable

$$\nu(n,\theta) = A(q)y(n) - B(q)u(n)$$

Output signal prediction error is then given by

$$\varepsilon(n,\theta) = D(q)\nu(n,\theta)$$

ARARX model state vector has a form

$$\varphi(n,\theta) = \left[-y(n-1)... - y(n-n_a)u(n-1)...u(n-n_b)\nu(n-1,\theta)...\nu(n-n_d,\theta)\right]^T$$

where the auxiliary variable and prediction error are

$$\nu(n,\theta) = y(n) + a_1 y(n-1) + ... + a_{n_a} y(n-n_a) - b_1 u(n-1) - ... - b_{n_b} u(n-n_b)$$

$$\varepsilon(n,\theta) = \nu(n,\theta) + d_1 \nu(n-1,\theta) + ... + d_{n_d} \nu(n-n_d,\theta)$$

Since the calculation of the predictor from past data depends on the parameter vector, predictor defines pseudo-linear regression.

PARAMETER ESTIMATION

After a model has been selected, it is necessary to parameterize it. Parameter estimation is based on solving linear or pseudo-linear regression, depending on chosen model. It tends to minimize prediction error.

Linear regression

Linear regression is very useful for describing linear and nonlinear systems. Linear regression employs a predictor in form $\hat{y}(n|\theta) = \varphi^T(n)\theta$

where $\phi(t)$ is state vector, and θ is parameter vector.

List squares criterion for the linear regression is given by

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{2}\left[y(n) - \varphi^T(n)\theta\right]^2$$

Since the list squares criterion is a quadratic function in θ, it can be minimized analytically, which gives, provided the indicated inverse exists, least-squares estimate:

$$\hat{\theta}_N^{LS} = \arg\min V_N(\theta, Z^N) = \left[\frac{1}{N}\sum_{n=1}^{N} \varphi(n)\varphi^T(n)\right]^{-1} \frac{1}{N}\sum_{n=1}^{N} \varphi(n)y(n)$$

It is convenient to introduce the matrix

$R(N) = \frac{1}{N}\sum_{n=1}^{N} \varphi(n)\varphi^T(n)$ and vector $f(N) = \frac{1}{N}\sum_{n=1}^{N} \varphi(n)y(n)$

Then is $\hat{\theta}_N^{LS} = R(N)^{-1} f(N)$

An alternative is to view

$\hat{\theta}_N^{LS}$ as the solution of linear equations $R(N)\hat{\theta}_N^{LS} = f(N)$

These equations are known as the normal equations. There are several different methods

for the construction of matrix *R*. An efficient way is using QR factorizations.

Pseudo-linear regression

In general, prediction error $\varepsilon(n,\theta)$ should be independent of past data. If the prediction error is correlated with a past data, then there was more information available in past data about output signal $y(n)$ than picked up by the predictor $y(n|\theta)$. The predictor then is not ideal. This leads to the characterization of a good model as one that produces prediction errors that are independent of past data.

Let predictor be given in form

$\hat{y}(n|\theta) = \varphi^T(n,\theta)\theta$ If the state vector $\varphi(n,\theta)$ does not depend on θ this relationship would be a linear regression. From this, the term pseudo-linear regression for predictor is derived. The state vector $\varphi(n,\theta)$ contains relevant past data, partly reconstructed using the current model. Thus, it is reasonable to require from the model that the resulting prediction errors be uncorrelated with state vector $\varphi(n,\theta)$ (3]:

$$\hat{\theta}_N^{PRL} = sol\left\{\frac{1}{N}\sum_{n=1}^{N}\varphi(n,\theta)\left[y(n) - \varphi^T(n,\theta)\theta\right] = 0\right\}$$

This is called the PLR estimate.

There are several approaches for solving this problem. They are basically numerical solutions by iterative search methods. One of these methods is bootstrap method.

It is necessary to solve

$$f_N(\theta, Z^N) = \frac{1}{N}\sum_{n=1}^{N}\varphi(n,\theta)\varepsilon(n,\theta) = 0$$

in the special case where the prediction error is given by

$$\varepsilon(n,\theta) = y(n) - \varphi(n,\theta)^T\theta$$

Let $\hat{\theta}_N^{(i-1)}$

be a current iterate. Then is natural to determine the next one by solving

$$\frac{1}{N}\sum_{n=1}^{N}\varphi(n,\hat{\theta}_N^{(i-1)})\left[y(n) - \varphi^T(n,\hat{\theta}_N^{(i-1)})\theta\right] = 0$$

This is a linear problem and can be solved as:

$$\hat{\theta}_N^{(i)} = \left[\frac{1}{N}\sum_{n=1}^{N}\varphi(n,\hat{\theta}_N^{(i-1)})\varphi^T(n,\hat{\theta}_N^{(i-1)})\right]^{-1}\left[\frac{1}{N}\varphi(n,\hat{\theta}_N^{(i-1)})y(n)\right] = 0$$

Solving of this linear problem is essentially a least-squares problem with proper definitions of *R(N)* and *f(N)*. The algorithm is known as a bootstrap method, since it alternates between computing θ and forming new vectors.

RESULTS

In this paper, usage of simple AR and ARX models is analyzed. For obtaining parameters of the predictor, these models require solving linear regression. In addition, complex methods such as ARMA, ARMAX and ARARX are discussed. They require solving of pseudo-linear regression. Using more complex methods result with better compression ratio, but computing is much more time consuming having iterative functions for parameterization. General review of results for these models is given in Tables 2 and 3. These results were achieved using predictive techniques in combination with techniques of entropy coding (non-adaptive zero-order arithmetic coding using standard 32-bit integer arithmetic).

Analysis of results indicates that difference in achieved compression ratios for simpler linear time-invariant models, such as AR and ARX model, and for more complex models, such as ARMA, ARMAX, and ARARX model, is not significant and it is about 2.5 %.

On the other hand, AR and ARMA models do not exploit correlation between different leads of the same ECG signal, while ARX, ARMAX and ARARX model use this correlation. In addition, results indicate that as order of model increase, potential ECG compression ratio increases, with increased complexity too.

Therefore, we recommend using ARX model. This model gives good compression ratios, it is not too complex, and it uses correlation between leads of the same ECG signal.

Table 2. Results of ECG signal prediction - MIT-BIH ECG Compression Test Database

	H_{min}[b/sym]	H_{max}[b/sym]	H_{sr}[b/sym]	CR_{min}	CR_{max}	CR_{sr}
AR	3.37	5.70	4.87	2.11	3.56	2.46
ARMA	3.33	5.53	4.77	2.17	3.61	2.52
ARX	3.37	5.63	4.85	2.13	3.56	2.47
ARMAX	3.29	5.59	4.78	2.15	3.65	2.51
ARARX	3.19	5.61	4.76	2.14	3.76	2.52

Table 3. Results of ECG signal prediction - MIT-BIH Arrhythmia Database

	H_{min}[b/sym]	H_{max}[b/sym]	H_{sr}[b/sym]	CR_{min}	CR_{max}	CR_{sr}
AR	3.26	4.53	4.02	2.43	3.37	2.74
ARMA	3.25	4.48	4.01	2.46	3.38	2.75
ARX	3.25	4.55	4.01	2.42	3.38	2.74

CONCLUSION

This paper presents predictive techniques for lossless ECG compression. Described models are custom in control theory and they are rarely used in data compression. This paper proposes using of these models in ECG signal compression.

Based on presented results, we recommend using simple models, such as AR and ARX that do not require iterative solving of pseudo-linear regression. After prediction, we recommend using of entropy coding (Huffman [8] or arithmetic coding [9]) on prediction error signal.

Disadvantage of predictive methods is weak utilization of correlation between neighbor beats of the same ECG signal. Used methods have possibility to include this correlation, but then they become too complex.

According to the results of survey among medical experts, we suggest using analogue-digital conversion with high sampling rate and a higher number of quantization levels. Sampling rate of 240 samples per second and resolution of 10 bits per sample offer good compromise between quality and storage capacity.

Acknowledgement

We thank to Robert Jung M.D., Čanković M.D., Velimir Seratlić M.D., and Slobodan Jovanović M.D. from Novi Sad and Nina Žigon M.D., Ph.D. and Milošević M.D., Ph.D. from Belgrade, for fruitful discussions and their help and suggestions in the process of evaluation of different ECG digitalization schemes.

REFERENCES

1. Duda K, Turcza P, Zielinski TP. Lossless ECG Compression with Lifting Wavelet Transform, IEEE Instrumentation and Measurement Technology Conference, Budapest, Hungary, May 21-23, 2001. p. 640-4.
2. Elaine Marieb, Human Anatomy and Physiology, 3rd ed. Pearson Benjamin Cummings; 1995. p. 615-33.
3. Lennart Ljung, System Identification -Theory for the User, 2nd ed. Prentice Hall, Upper Saddle River, NJ; 1999.
4. Despotovic M, Žigon N, Bajić D. ECG signal compression techniques. Annals of the Academy of Studenica 2001;(4).
5. Bajić D. ECG standards?, MEDNET 2001, The 6th World Congress on the Internet in Medicine, Udine, Italy, 29 Nov - 2Dec 2001.
6. Jalaleddine SMS, Hutchens CG, Strattan RD, Coberly WA. ECG data compression techniques - a unified approach. IEEE Transactions on Biomedical Engineering 1990;37(4).
7. Shahabi C, Ortega A, Kolahdouzan MR. A Comparison Of Different Haptic Compression Techniques, Integrated Media Systems Center, University of Southern California
8. Huffman DA. A Method for the Construction of Minimum-Redundancy Codes, Proceedings of the I.R.E., 40 (September 1952). p. 1098-101.
9. Rissanen JJ, Langdon GC. Arithmetic coding, IBM Journal of Research and Development 1979; 23(2):149-62.

Appendix A ECG signal for different parameters of digitalization

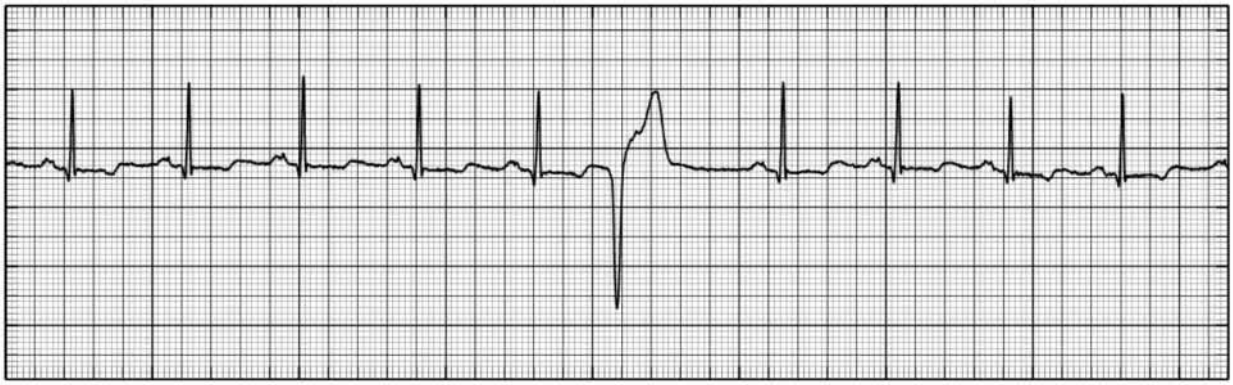

Figure I. ECG signal I, 360 samples per second, 11 bits per sample

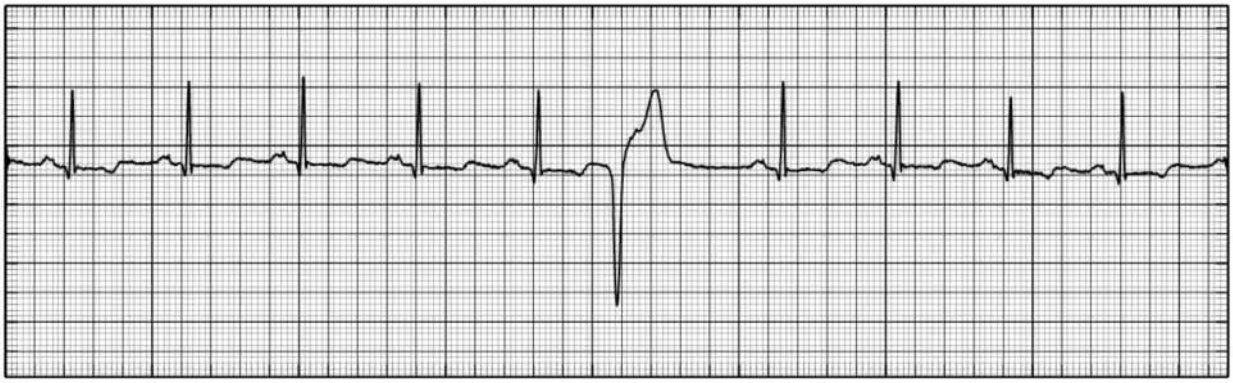

Figure II. ECG signal II, 240 samples per second, 10 bits per sample

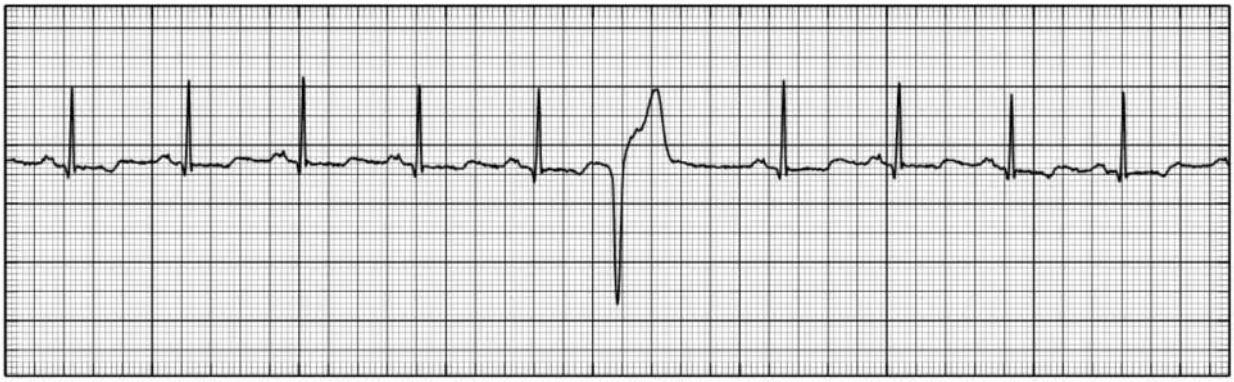

Figure III. ECG signal III, 180 samples per second, 9 bits per sample

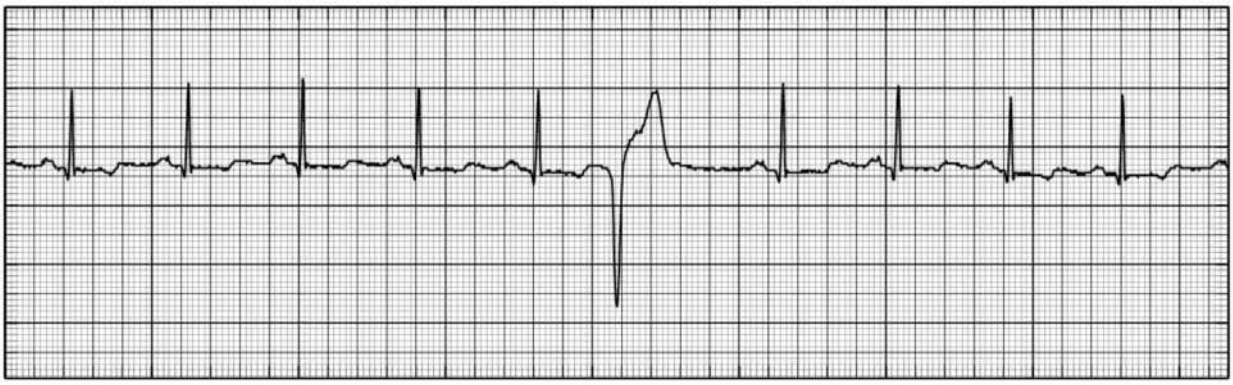

Figure IV. ECG signal IV, 180 samples per second, 8 bits per sample

Iistitute of Industrial...; Address correspondence to: Ilia V. Yarmoshenko, Ph. D. Senior Researcher, Institute of Industrial Ecology of Ural Branch of Russian Academy of Sciences, 20A S. Kovalevskoy St., Ekaterinburg 620219 Russia, E-mail ivy@ecko.uran.ru, The manuscript was received: 22. 10. 2003, Accepted for publication: 31. 10. 2003