BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES
Vol. 55, No. 4, 2007

Bayesian and empirical Bayesian approach to weighted averaging of ECG signal

A. MOMOT[1], M. MOMOT[2], and J. ŁĘSKI[2,3]*

[1] Silesian University of Technology, Institute of Computer Science, 16 Akademicka St., 44-101 Gliwice, Poland
[2] Institute of Medical Technology and Equipment, 118 Roosevelt St., 41-800 Zabrze, Poland
[3] Silesian University of Technology, Institute of Electronics, 16 Akademicka St., 44-101 Gliwice, Poland

Abstract. One of the prime tool in non-invasive cardiac electrophysiology is the recording of an electrocardiographic signal (ECG) which analysis is greatly useful in the screening and diagnosis of cardiovascular diseases. However, one of the greatest problems is that usually recording an electrical activity of the heart is performed in the presence of noise. The paper presents Bayesian and empirical Bayesian approach to problem of weighted signal averaging in time domain which is commonly used to extract a useful signal distorted by a noise. The averaging is especially useful for biomedical signal such as ECG signal, where the spectra of the signal and noise significantly overlap. Using the methods of weighted averaging are motivated by variability of noise power from cycle to cycle, often observed in reality. It is demonstrated that exploiting a probabilistic Bayesian learning framework leads to accurate prediction models. Additionally, even in the presence of nuisance parameters the empirical Bayesian approach offers the method of theirs automatic estimation which reduces number of preset parameters. Performance of the new method is experimentally compared to the traditional averaging by using arithmetic mean and weighted averaging method based on criterion function minimization.

Key words: ECG signal, weighted averaging, Bayesian inference.

1. Introduction

In the most of biomedical signal processing systems (for example electrocardiographic signal, which is the main case of interest in this work) noise reduction plays very important role. Accuracy of all later operations performed on signal, such as detections, classifications or measurements, depends on the quality of noise-reduction algorithms. Using the fact that certain biological systems produce repetitive patterns, an averaging in the time domain may be used for noise attenuation. Traditional averaging technique assumes the constancy of the noise power cycle-wise, however the most types of noise are not stationary. In these cases a need for using weighted averaging occurs, which reduces influence of hardly distorted cycles on resulting averaged signal (or even eliminates them).

The paper presents new method for resolving of signal averaging problem which incorporates Bayesian and empirical Bayesian inference. By exploiting a probabilistic Bayesian framework [1], [2] and an expectation-maximization technique [3] it can be derived an algorithm of weighted averaging which application to electrocardiographic (ECG) signal averaging is competitive with alternative methods as will be shown in the later part of the paper.

2. Signal averaging methods

Let us assume that in each signal cycle $y_i(j)$ is the sum of a deterministic (useful) signal $x(j)$, which is the same in all cycles, and a random noise $n_i(j)$ with zero mean and variance for the ith cycle equal to σ_i^2. Thus, $y_i(j) = x(j) + n_i(j)$, where i is the cycle index $i \in \{1, 2, \ldots, M\}$, and the j is the sample index in the single cycle $j \in \{1, 2, \ldots, N\}$ (all cycles have the same length N). The weighted average is given by

$$v(j) = \sum_{i=1}^{M} w_i y_i(j), \tag{1}$$

where w_i is a weight for ith signal cycle and $v(j)$ is the averaged signal.

2.1. Traditional arithmetic averaging. The traditional ensemble averaging with arithmetic mean as the aggregation operation gives all the weights w_i equal to M^{-1}. If the noise variance is constant for all cycles, then these weights are optimal in the sense of minimizing the mean square error between $\boldsymbol{v}$ and $\boldsymbol{x}$, assuming Gaussian distribution of noise. When the noise has a non-Gaussian distribution, the estimate (1) is not optimal, but it is still the best of all linear estimators of $\boldsymbol{x}$ [4].

2.2. Weighted averaging method based on criterion function minimization. As it is shown in [8], for $\boldsymbol{y}_i = [y_i(1), y_i(2), \ldots, y_i(N)]^{\mathrm{T}}$, $\boldsymbol{w} = [w_1, w_2, \ldots, w_M]^{\mathrm{T}}$ and $\boldsymbol{v} = [v(1), v(2), \ldots, v(N)]^{\mathrm{T}}$ minimization the following scalar criterion function

$$I_m(\boldsymbol{w}, \boldsymbol{v}) = \sum_{i=1}^{M} (w_i)^m \rho(\boldsymbol{y}_i - \boldsymbol{v}), \tag{2}$$

with respect to the weights vector w yields

*e-mail: jacek.leski@polsl.pl

$$w_i = \frac{[\rho(\boldsymbol{y}_i - \boldsymbol{v})]^{(1-m)^{-1}}}{\sum\limits_{j=1}^{M} [\rho(\boldsymbol{y}_i - \boldsymbol{v})]^{(1-m)^{-1}}}. \quad (3)$$

for $i \in \{1, 2, \ldots, M\}$, where $\rho(\cdot)$ is a measure of dissimilarity for vector argument and $m \in (1, \infty)$ is a weighting exponent parameter. When the most frequently used quadratic function $\rho(\cdot) = \|\cdot\|_2^2$ is used, the averaged signal can be obtained as

$$\boldsymbol{v} = \frac{\sum\limits_{i=1}^{M} (w_i)^m \boldsymbol{y}_i}{\sum\limits_{i=1}^{M} (w_i)^m}, \quad (4)$$

for the weights vector given by (2) with the quadratic function. The optimal solution for minimization (2) with respect to $\boldsymbol{w}$ and $\boldsymbol{v}$ is a fixed point of (3) and (4) and it is obtained from the Picard iteration.

If m tends to one then the trivial solution is obtained where only one weight, corresponding to the signal cycle with the smallest dissimilarity to averaged signal, is equal to one. If m tend to infinity then weights tend to M^{-1} for all i. Generally, a larger m results in a smaller influence of dissimilarity measures. The most common value of m is 2 which results in greater decrease of medium weights [5].

2.3. Bayesian and empirical Bayesian weighted averaging methods. Given a data set $\boldsymbol{y} = \{y_i(j)\}$, where i is the cycle index $i \in \{1, 2, \ldots, M\}$ and the j is the sample index in the single cycle $j \in \{1, 2, \ldots, N\}$, there are made assumptions that $y_i(j) = x(j) + n_i(j)$, where a random noise $n_i(j)$ is zero-mean Gaussian with variance for the ith cycle equal to σ_i^2, and signal $\boldsymbol{x} = \{x(j)\}$ has also Gaussian distribution with zero-mean and covariance matrix $B = \mathrm{diag}(\eta_1^2, \eta_2^2, \ldots, \eta_N^2)$. The zero-mean assumption for the signal expresses the fact that no prior knowledge about the real distance from the signal to the isoelectric line. Thus, from the Bayes rule, the posterior distribution over $\boldsymbol{x}$ and the noise variance is proportional to

$$p(x, \alpha | x, \beta) = \frac{p(y|x,\alpha)p(x|\beta)p(\alpha)}{p(y)} \propto \left(\prod_{i=1}^{M} \alpha_i\right)^{\frac{N}{2}} \prod_{j=1}^{N} \beta_j^{\frac{1}{2}}$$
$$\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N} (y_i(j) - x(j))^2 \alpha_i - \frac{1}{2}\sum_{j=1}^{N} (x(j))^2 \beta_j\right), \quad (5)$$

where $\alpha_i = \sigma_i^{-2}$ and $\beta_j = \eta_j^{-2}$, because of assumption that the prior $p(\boldsymbol{\alpha})$ is approximately constant (for large M the influence of this prior is very small). The values $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ which maximize (5) are given by

$$\alpha_i = N \left(\sum_{j=1}^{N} (y_i(j) - x(j))^2\right)^{-1}, \quad (6)$$

$$x(j) = \frac{\sum\limits_{i=1}^{M} \alpha_i y_i(j)}{\beta_j + \sum\limits_{i=1}^{M} \alpha_i} \quad (7)$$

for $i \in \{1, 2, \ldots, M\}$ and $j \in \{1, 2, \ldots, N\}$. The conditions in equations (6) and (7) are obtained by differentiating logarithm of (5) with respect to $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ respectively and setting the results equal to zero. As can be calculated differentiating logarithm of (5) with respect to $\boldsymbol{\alpha}$ gives:

$$\frac{\partial}{\partial \alpha_k} \log \left(\left(\prod_{i=1}^{M} \alpha_i\right)^{\frac{N}{2}} \prod_{j=1}^{N} \beta_j^{\frac{1}{2}} \right.$$
$$\left. \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N} (y_i(j) - x(j))^2 \alpha_i - \frac{1}{2}\sum_{j=1}^{N} (x(j))^2 \beta_j\right)\right) =$$
$$= \frac{\partial}{\partial \alpha_k}\left(\frac{N}{2}\sum_{i=1}^{M} \log \alpha_i\right) -$$
$$-\frac{1}{2}\frac{\partial}{\partial \alpha_i}\left(\sum_{i=1}^{M}\sum_{j=1}^{N} (y_i(j) - x(j))^2 \alpha_i\right) =$$
$$= \frac{N}{2\alpha_k} - \frac{1}{2}\sum_{j=1}^{N} (y_k(j) - x(j))^2, \quad (8)$$

and with respect to $\boldsymbol{x}$ gives:

$$\frac{\partial}{\partial x(k)} \log \left(\left(\prod_{i=1}^{M} \alpha_i\right)^{\frac{N}{2}} \prod_{j=1}^{N} \beta_j^{\frac{1}{2}} \right.$$
$$\left. \exp\left(-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N} (y_i(j) - x(j))^2 \alpha_i - \frac{1}{2}\sum_{j=1}^{N} (x(j))^2 \beta_j\right)\right) =$$
$$= -\frac{1}{2}\frac{\partial}{\partial x(k)}\left(\sum_{j=1}^{N}\sum_{i=1}^{M} (y_i(j) - x(j))^2 \alpha_i\right) -$$
$$-\frac{1}{2}\frac{\partial}{\partial x(k)}\left(\sum_{j=1}^{N} (x(j))^2 \beta_j\right) =$$
$$= \sum_{i=1}^{M} (y_i(k) - x(k))\, \alpha_i - x(k)\beta_k =$$
$$= \sum_{i=1}^{M} y_i(k)\alpha_i - x(k)\left(\beta_k + \sum_{i=1}^{M} \alpha_i\right). \quad (9)$$

Since β_j could not be observed, the iterative EM algorithms is used like in [6]. Assuming the gamma prior (which is conjugate prior distribution for the inverse of the normal variance [1], [2]) with scale parameter λ and shape parameter p:

$$p(\beta_j) = \frac{\lambda^p}{\Gamma(p)}\beta_j^{p-1}\exp(-\lambda\beta_j), \qquad \lambda, p, \beta_j > 0 \tag{10}$$

for all j, as values of β_j it is taken

$$\begin{aligned}
E(\beta_j|x(j)) &= \int_0^\infty \beta_j p(\beta_j|x(j))d\beta_j = \\
&= \frac{\int_0^\infty \beta_j p(x(j)|\beta_j)p(\beta_j)d\beta_j}{\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j} = \\
&= \frac{\int_0^\infty \beta_j \frac{\sqrt{\beta_j}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x(j)^2\beta_j\right) \frac{\lambda^p}{\Gamma(p)}\beta_j^{p-1}\exp(-\lambda\beta_j)d\beta_j}{\int_0^\infty \frac{\sqrt{\beta_j}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x(j)^2\beta_j\right) \frac{\lambda^p}{\Gamma(p)}\beta_j^{p-1}\exp(-\lambda\beta_j)d\beta_j} = \\
&= \frac{\frac{\lambda^p}{\sqrt{2\pi}\Gamma(p)}\left(-\frac{1}{2}x(j)^2-\lambda\right)^{-1}\beta_j^{p+\frac{1}{2}}\exp\left(\left(-\frac{1}{2}x(j)^2-\lambda\right)\beta\right)\Big|_0^{+\infty}}{\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j} + \\
&+ \frac{-\left(p+\frac{1}{2}\right)\left(-\frac{1}{2}x(j)^2-\lambda\right)^{-1}\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j}{\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j} = \\
&= \frac{0+\left(p+\frac{1}{2}\right)\left(\frac{1}{2}x(j)^2+\lambda\right)^{-1}\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j}{\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j} \\
&= \frac{2p+1}{x(j)^2+2\lambda}.
\end{aligned} \tag{11}$$

The estimate $\hat{\lambda}$ of hyperparameter λ can be calculated by applying empirical method [7]. The probability distribution function $p(x(j)|\lambda)$ can be written in the form

$$\begin{aligned}
p(x(j)|\lambda) &= \int_0^\infty \frac{\sqrt{\beta_j}}{\sqrt{2\pi}} \\
&\exp\left(-\frac{1}{2}x(j)^2\beta_j\right)\frac{\lambda^p}{\Gamma(p)}\beta_j^{p-1}\exp(-\lambda\beta_j)d\beta_j = \\
&\left| \begin{array}{ll} \left(x(j)^2+2\lambda\right)\beta_j = t^2 & \beta_j = t^2\left(x(j)^2+2\lambda\right)^{-1} \\ \left(x(j)^2+2\lambda\right)d\beta_j = 2tdt & \end{array} \right| \\
&= \frac{\lambda^p}{\Gamma(p)\sqrt{2\pi}}\int_0^\infty \left(\frac{t^2}{x(j)^2+2\lambda}\right)^{p-\frac{1}{2}}\exp(-\tfrac{1}{2}t^2)\frac{2t}{x(j)^2+2\lambda}dt = \\
&= \frac{\lambda^p}{\Gamma(p)\left(x(j)^2+2\lambda\right)^{p+\frac{1}{2}}}\int_0^\infty \frac{2}{\sqrt{2\pi}}t^{2p}\exp(-\tfrac{1}{2}t^2)dt = \\
&= \frac{\lambda^p\,(2p-1)!!}{\Gamma(p)\left(x(j)^2+2\lambda\right)^{p+\frac{1}{2}}},
\end{aligned} \tag{12}$$

where the last equation is the consequence of the fact that the last integral is the $2p$-th moment of standard normal distribution assuming that p is positive integer. The double factorial is defined as follows

$$(2p-1)!! = 1\cdot 3\cdot\ldots\cdot(2p-1). \tag{13}$$

Since p is a positive integer:

$$\begin{aligned}
E(|x(j)|\,|\lambda) &= 2\int_0^\infty x(j)p(x(j)|\lambda)dx(j) = \\
&= 2\int_0^\infty x(j)\frac{\lambda^p\,(2p-1)!!}{\Gamma(p)\left(x(j)^2+2\lambda\right)^{p+\frac{1}{2}}}dx(j) = \\
&= \frac{2\lambda^p\,(2p-1)!!}{\Gamma(p)}\frac{-1}{(2p-1)\left(x(j)^2+2\lambda\right)^{p-\frac{1}{2}}}\Bigg|_0^{+\infty} = \\
&= \frac{2\lambda^p\,(2p-1)!!}{\Gamma(p)}\frac{1}{(2p-1)\,(2\lambda)^{p-\frac{1}{2}}} = \\
&= \frac{(2p-1)!!}{\Gamma(p)\,(2p-1)}2^{\frac{3}{2}-p}\lambda^{\frac{1}{2}}.
\end{aligned} \tag{14}$$

Therefore the estimate $\hat{\lambda}$ of hyperparameter λ can be calculated based on first absolute sample moment as follows

$$\hat{\lambda} = \left(\frac{\Gamma(p)\,(2p-1)}{(2p-1)!!}2^{p-\frac{3}{2}}\frac{1}{N}\sum_{j=1}^{N}|x(j)|\right)^2. \tag{15}$$

The results above are generalisation of those presented in [8], where the prior distribution of β_j was exponential which is the special case of gamma distribution, for the shape parameter p equal 1.

Therefore the proposed Empirical Bayesian Weighted Averaging (EBWA.1) algorithm can be described as follows, where ε and p (positive integer) are preset parameters:

1. Initialize $\boldsymbol{v}^{(0)} \in R^N$. Set the iteration index $k = 1$.
2. Calculate the hyperparameter $\lambda^{(k)}$ using (15), next $\beta_j^{(k)}$ using (11) for $j = 1, 2, \ldots, N$ and $\alpha_i^{(k)}$ using (6) for $j = 1, 2, \ldots, M$, assuming $\boldsymbol{x} = \boldsymbol{v}^{(k-1)}$.
3. Update the averaged signal for kth iteration $\boldsymbol{v}^{(k)}$ using (7) and $\beta_j^{(k)}$ and $\alpha_i^{(k)}$, assuming $\boldsymbol{v}^{(k)} = \boldsymbol{x}$.
4. If $\frac{\left\|\boldsymbol{v}^{(k)}-\boldsymbol{v}^{(k-1)}\right\|}{\left\|\boldsymbol{v}^{(k)}\right\|} > \varepsilon$ then $k \leftarrow k+1$ and go to 2, else stop.

When the parameter p is positive integer greater than 1, the estimate $\hat{\lambda}$ of hyperparameter λ can be calculated based on third absolute sample moment. Since p is a positive integer:

$$E(|x(j)|^3|\lambda) = 2\int_0^\infty x(j)^3 p(x(j)|\lambda)dx(j) =$$
$$= 2\int_0^\infty x(j)^3 \frac{\lambda^p (2p-1)!!}{\Gamma(p)(x(j)^2+2\lambda)^{p+\frac{1}{2}}}dx(j) =$$
$$= \frac{2\lambda^p (2p-1)!!}{\Gamma(p)}$$
$$\left(\frac{-1}{(2p-3)(x(j)^2+2\lambda)^{p-\frac{3}{2}}} + \frac{2\lambda}{(2p-1)(x(j)^2+2\lambda)^{p-\frac{1}{2}}}\right)\Bigg|_0^{+\infty} =$$
$$= \frac{2\lambda^p (2p-1)!!}{\Gamma(p)}\left(\frac{1}{(2p-3)(2\lambda)^{p-\frac{3}{2}}} - \frac{2\lambda}{(2p-1)(2\lambda)^{p-\frac{1}{2}}}\right) =$$
$$= \frac{(2p-1)!!}{\Gamma(p)}2^{\frac{5}{2}-p}\left(\frac{\lambda^{\frac{3}{2}}}{2p-3} - \frac{\lambda^{\frac{3}{2}}}{(2p-1)}\right) =$$
$$= \frac{(2p-3)!!}{(2p-3)\Gamma(p)}2^{\frac{7}{2}-p}\lambda^{\frac{3}{2}}. \tag{16}$$

Therefore the hyperparameter λ can be calculated based on third absolute sample moment as follows

$$\hat{\lambda} = \left(\frac{(2p-3)\,\Gamma(p)}{2^{\frac{7}{2}-p}(2p-3)!!}\frac{1}{N}\sum_{j=1}^{N}|x(j)|^3\right)^{\frac{2}{3}} \tag{17}$$

In this case the proposed Empirical Bayesian Weighted Averaging (EBWA.3) algorithm can be described as follows, where ε and p (positive integer greater than 1) are preset parameters:

1. Initialize $\boldsymbol{v}^{(0)} \in R^N$. Set the iteration index $k = 1$.
2. Calculate the hyperparameter $\lambda^{(k)}$ using (17), next $\beta_j^{(k)}$ using (11) for $j = 1, 2, \ldots, N$ and $\alpha_i^{(k)}$ using (6) for $j = 1, 2, \ldots, M$, assuming $\boldsymbol{x} = \boldsymbol{v}^{(k-1)}$.
3. Update the averaged signal for kth iteration $\boldsymbol{v}^{(k)}$ using (7) and $\beta_j^{(k)}$ and $\alpha_i^{(k)}$, assuming $\boldsymbol{v}^{(k)} = \boldsymbol{x}$.
4. If $\frac{\|\boldsymbol{v}^{(k)}-\boldsymbol{v}^{(k-1)}\|}{\|\boldsymbol{v}^{(k)}\|} > \varepsilon$ then $k \leftarrow k+1$ and go to 2, else stop.

However, the assumption of the gamma prior for β_j is not necessarily adequate in all cases. When no reliable prior information concerning β_j exists, it is possible to use non-informative Jeffrey's prior [1] given by

$$p(\beta_j) = \beta_j^{-1}, \qquad \beta_j > 0. \tag{18}$$

In this situation the proposed algorithm would not require setting of additional parameters such as p or λ, because

$$E(\beta_j|x(j)) = \int_0^\infty \beta_j p(\beta_j|x(j))d\beta_j =$$
$$= \frac{\int_0^\infty \beta_j p(x(j)|\beta_j)p(\beta_j)d\beta_j}{\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j} =$$
$$= \frac{\int_0^\infty \beta_j \frac{\sqrt{\beta_j}}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x(j)^2\beta_j\right)\beta_j^{-1}\exp(-\lambda\beta_j)d\beta_j}{\int_0^\infty \frac{\sqrt{\beta_j}}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x(j)^2\beta_j\right)\beta_j^{-1}\exp(-\lambda\beta_j)d\beta_j} =$$
$$= \frac{\frac{1}{\sqrt{2\pi}}\left(-\frac{1}{2}x(j)^2\right)^{-1}\beta_j^{\frac{1}{2}}\exp\left(\left(-\frac{1}{2}x(j)^2\right)\beta\right)\Big|_0^{+\infty}}{\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j} + \tag{19}$$
$$+\frac{-\frac{1}{2}\left(-\frac{1}{2}x(j)^2\right)^{-1}\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j}{\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j} =$$
$$= \frac{0 + x(j)^{-2}\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j}{\int_0^\infty p(x(j)|\beta_j)p(\beta_j)d\beta_j} =$$
$$= \frac{1}{x(j)^2}.$$

As can be seen the result above is identical with (11) for $\lambda = 0$ and $p = 0$.
In this case the proposed Bayesian Weighted Averaging (BWA) algorithm can be described as follows, where ε is a preset parameter:

1. Initialize $\boldsymbol{v}^{(0)} \in R^N$. Set the iteration index $k = 1$.
2. Calculate $\beta_j^{(k)}$ using (19) for $j = 1, 2, \ldots, N$ and $\alpha_i^{(k)}$ using (6) for $j = 1, 2, \ldots, M$, assuming $\boldsymbol{x} = \boldsymbol{v}^{(k-1)}$.
3. Update the averaged signal for kth iteration $\boldsymbol{v}^{(k)}$ using (7) and $\beta_j^{(k)}$ and $\alpha_i^{(k)}$, assuming $\boldsymbol{v}^{(k)} = \boldsymbol{x}$.
4. If $\frac{\|\boldsymbol{v}^{(k)}-\boldsymbol{v}^{(k-1)}\|}{\|\boldsymbol{v}^{(k)}\|} > \varepsilon$ then $k \leftarrow k+1$ and go to 2, else stop.

3. Numerical experiments

In all experiments using Weighted Averaging method based on Criterion Function Minimization (WACFM) as well as Bayesian and Empirical Bayesian Weighted Averaging methods (BWA, EBWA.1 and EBWA.3) calculations were initialized as the means of disturbed signal cycles. Iteration were stopped as soon as the L^2 norm for a successive pair of vectors was less than 10^{-6}, respectively $\boldsymbol{w}$ vectors for the WACFM and $\boldsymbol{v}$ vectors for the BWA, EBWA.1 and EBWA.3. For a computed averaged signal the performance of tested methods was evaluated by the maximal absolute difference between the deterministic component and the averaged signal.

The root mean-square error (RMSE) between the deterministic component and the averaged signal was also computed. All experiments were run in the R[1] environment for R version 2.4.0.

The simulated ECG signal cycles were obtained as the same deterministic component with added realizations of random noise. The deterministic component presented in Fig. 1 was obtained by averaging 500 real ECG signal cycles (2000-Hz and 16-bit resolution) with high signal to noise ratio and before averaging these cycles were time-aligned using the cross correlation method. A series of 100 ECG cycles was generated with the same deterministic component and zero-mean white Gaussian noise with four different standard deviations. For the first, second, third and fourth 25 cycles, the noise standard deviations were 10, 50, 100, 200 μV, respectively. These signal cycles were averaged using the following methods: Traditional Arithmetic Averaging (TAA), WACFM with $m = 2$, BWA as well as EBWA.1 and EBWA.3 with various values of parameter p. Subtraction of deterministic component from these averaged signal gives a residual noise.

The RMSE and the maximal value (MAX) of residual noise for all tested methods are presented in Table 1. The best results for each power of noise are bolded.

However, in practice besides of Gaussian types of noise, it can be observed random noise with heavy-tailed distribution [9]. The example of such distribution is Cauchy distribution with probability density function given by

$$f(x) = \frac{1}{\pi s}\left(1 + \left(\frac{x-l}{s}\right)^2\right)^{-1}, \tag{20}$$

where l is the location parameter and s is the scale parameter. The parameters are commonly used instead of expected value and standard deviation because of the absence of first two moments. In next experiment to the ECG signal was added Cauchy distributed random noise with $l = 0$ and $s = 10\mu$V. In Fig. 2 the deterministic component was presented along with example of Cauchy noise.

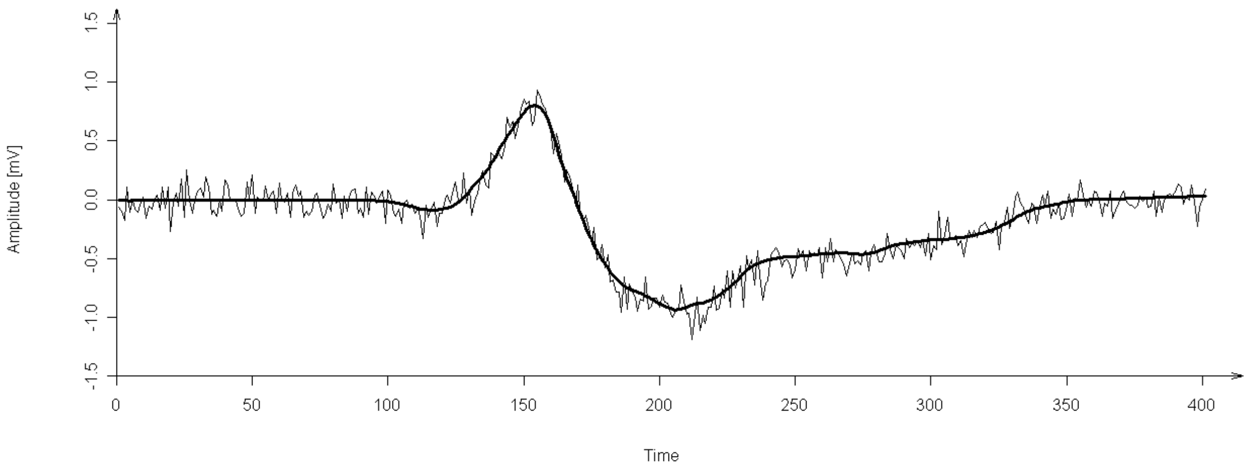

Fig. 1. The simulated ECG signal and this signal with 100 μV standard deviation Gaussian noise

Table 1
RMSE and maximum absolute error for averaged ECG signals with Gaussian noise

Averaging method		MAX [μV]	RMSE [μV]
TAA		39.17285	12.09635
WACFM		5.984498	1.938105
BWA		8.419009	2.345869
EBWA.1	$p = 1$	**5.585312**	**1.925123**
	$p = 2$	5.585997	1.925604
	$p = 3$	5.586081	1.925729
	$p = 4$	5.586110	1.925787
	$p = 5$	5.586123	1.925822
EBWA.3	$p = 2$	5.585873	1.925470
	$p = 3$	5.586123	1.925777
	$p = 4$	5.586178	1.925862
	$p = 5$	5.586202	1.925902

[1]R is a free software environment for statistical computing and graphics (http://www.r-project.org)

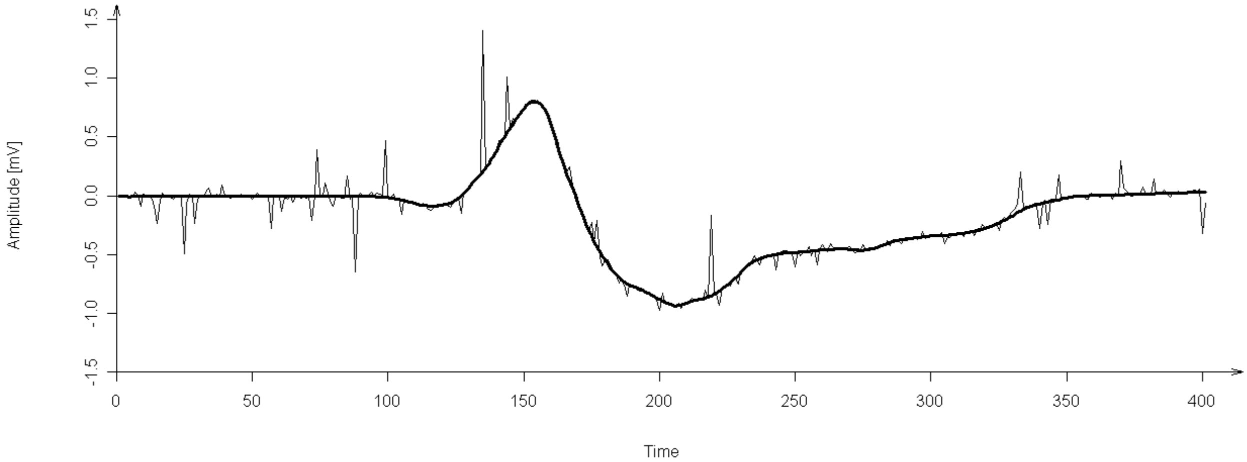

Fig. 2. The simulated ECG signal and this signal with Cauchy noise

Table 2
RMSE and maximum absolute error for averaged ECG signals with Cauchy noise

Averaging method		MAX [μV]	RMSE [μV]
TAA		6907.1413	424.3746
WACFM		69.25222	17.40984
BWA		**56.76064**	**14.71391**
EBWA.1	$p = 1$	60.05436	16.07817
	$p = 2$	60.41708	16.12834
	$p = 3$	60.47729	16.13810
	$p = 4$	60.50249	16.14290
	$p = 5$	60.51657	16.14591
EBWA.3	$p = 2$	60.34284	16.12001
	$p = 3$	60.49377	16.13974
	$p = 4$	60.53017	16.14519
	$p = 5$	60.54688	16.14789

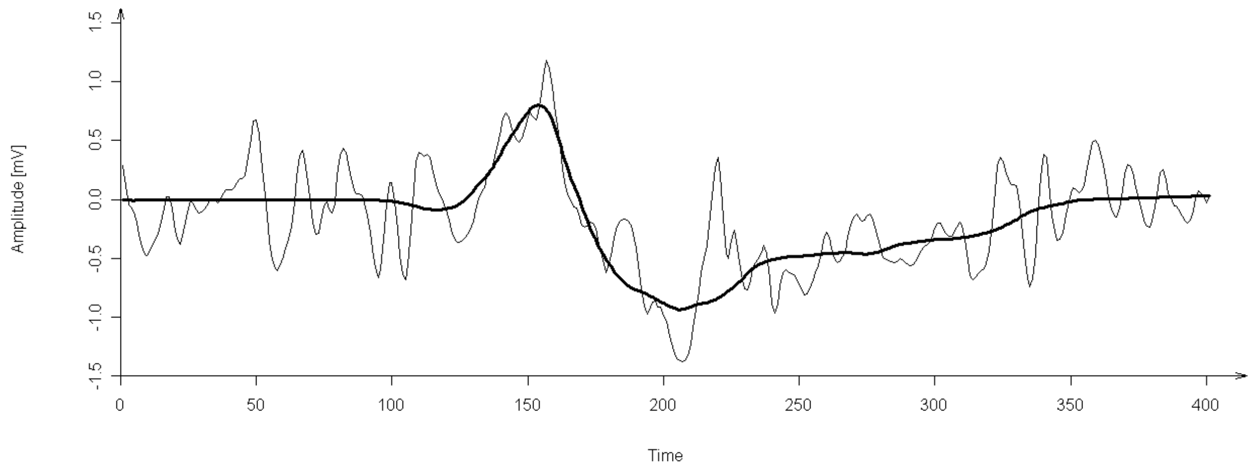

Fig. 3. The simulated ECG signal and this signal with muscle noise

The RMSE and the maximal value (MAX) of residual noise for all tested methods are presented in Table 2 and the best results are bolded. In this experiment the signal cycles were averaged using WACFM with $m = 3$, because for most common value of $m = 2$, the method does not reach stop condition even after 1000 iterations (although in previous cases it did not require more than 20 iterations to stop). It shows that the smallest RMSE were obtained by BWA method and a little bit worse results were obtained by EBWA and WACFM. It can be seen that despite the fact that the assumption of Gaussian distributed random noise is not satisfied, the results of the experiment are much better compared to the ones obtained by traditional arithmetic averaging.

Usually in case of electrocardiographic (ECG) signal, two principal sources of noise can be distinguished: the 'technical' caused by the physical parameters of the recording equipment and the 'physiological' representing the bioelectrical activity of living cells not belonging to the area of diagnostic interest (also called background activity). Both sources produce noise of random occurrence, overlapping the ECG signal in both time and frequency domains [10]. This was motivation to perform next experiment with ECG signal distorted by muscle noise which can be treated as composition of Gaussian and impulsive distortion (well modeled by heavy-tailed distribution). In practical applications signal to noise ratio is often very poor, even below one (which mean that noise level exceeds signal level). In Fig. 3 the deterministic component was presented along with example of such muscle noise where signal to noise ratio is equal one (0 dB).

The RMSE and the maximal value (MAX) of residual noise for all tested methods are presented in Table 3 and the best results are bolded. It shows that the smallest RMSE were obtained by BWA method and a little bit worse results were obtained by EBWA and WACFM. It can be seen that despite the fact that the assumption of Gaussian distributed random noise is not satisfied, the results of the experiment are still acceptable (see Figs. 4–8) and slightly better compared to the ones obtained by traditional arithmetic averaging.

Table 3
RMSE and maximum absolute error for averaged ECG signals with muscle noise

Averaging method		MAX [μV]	RMSE [μV]
TAA		106.47622	35.95089
WACFM		87.64957	33.22810
BWA		79.18179	**26.95332**
EBWA.1	$p=1$	79.04318	30.09347
	$p=2$	77.29849	30.26317
	$p=3$	76.84224	30.28546
	$p=4$	76.63690	30.29371
	$p=5$	76.52048	30.29784
EBWA.3	$p=2$	78.05169	30.25046
	$p=3$	76.64020	30.28531
	$p=4$	76.23566	30.29045
	$p=5$	**76.04061**	30.29180

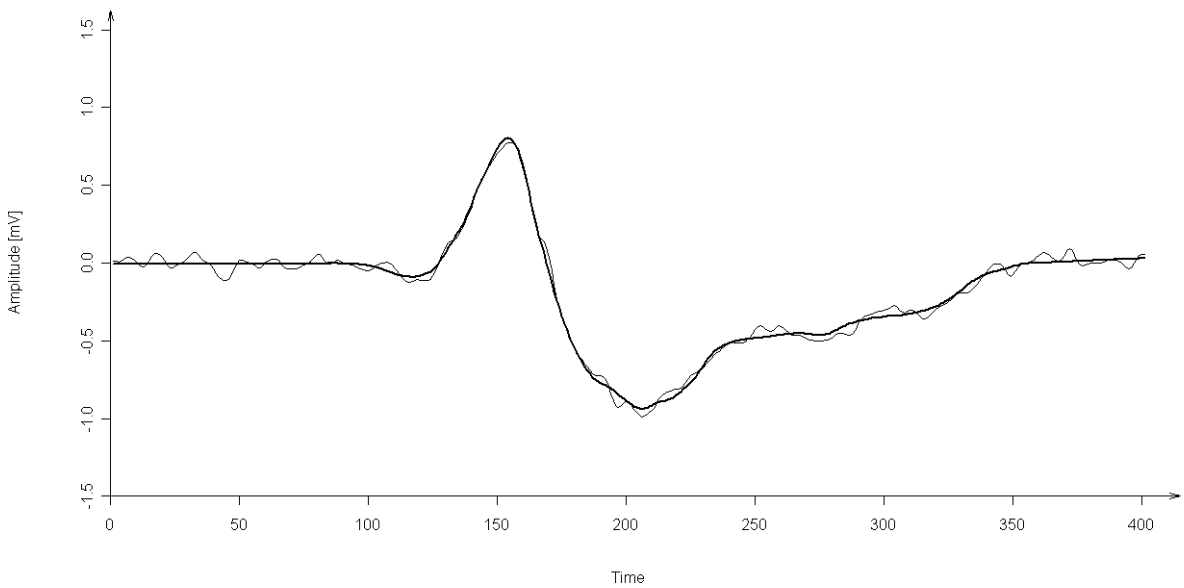

Fig. 4. The simulated ECG signal and the averaged signal using TAA method

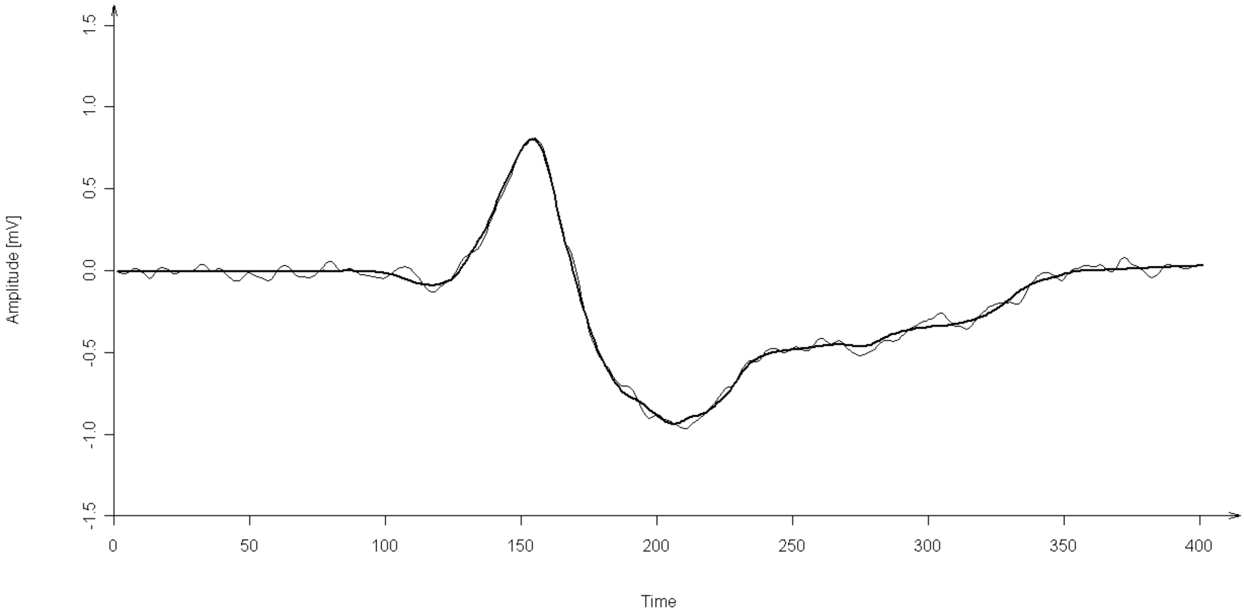

Fig. 5. The simulated ECG signal and the averaged signal using WACFM method

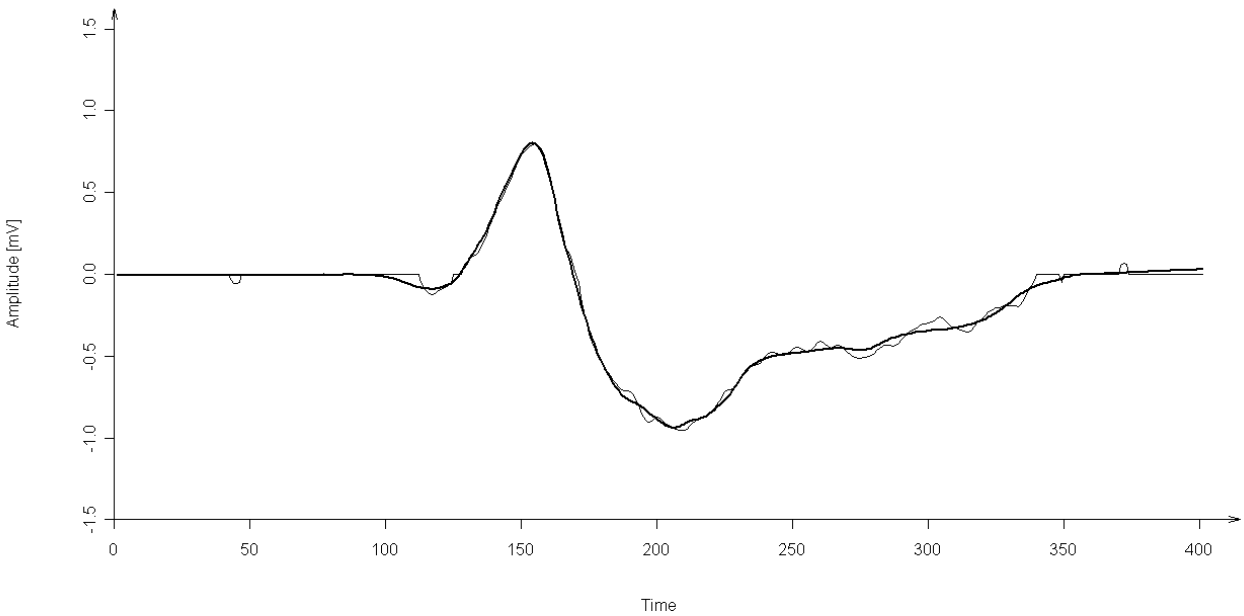

Fig. 6. The simulated ECG signal and the averaged signal using BWA method

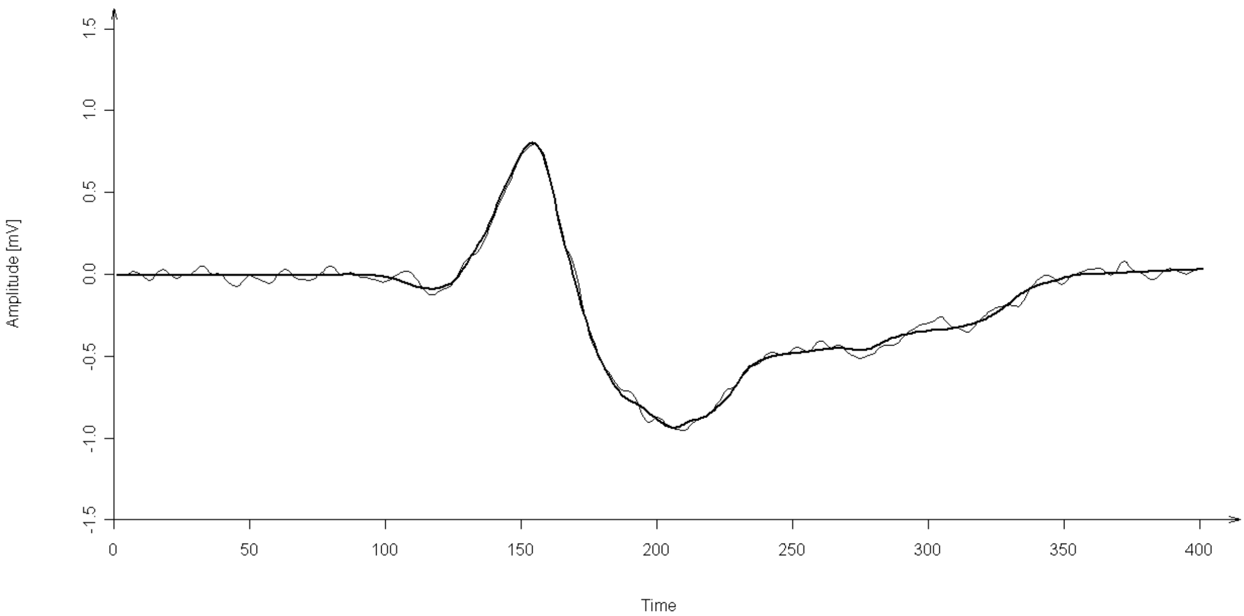

Fig. 7. The simulated ECG signal and the averaged signal using EBWA.1 method with parameter $p = 1$

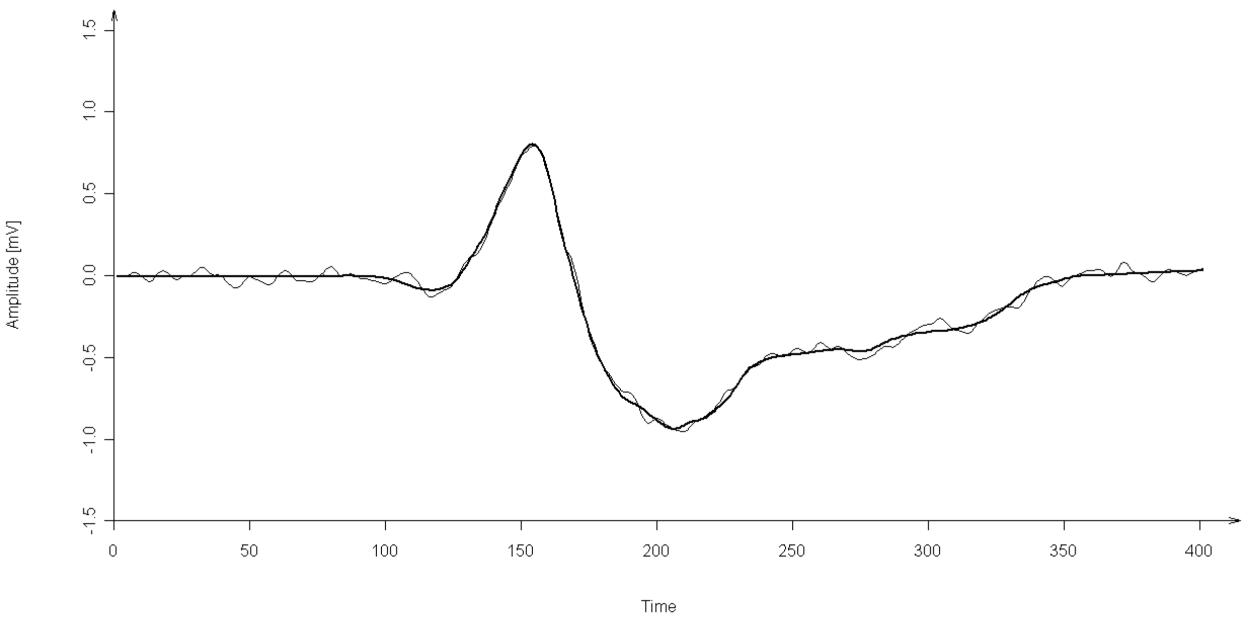

Fig. 8. The simulated ECG signal and the averaged signal using EBWA.3 method with parameter $p = 5$

4. Conclusions

In this work the new approach to weighted averaging of biomedical signal was presented along with the application to averaging ECG signals. Presented methods use the results of Bayesian and empirical Bayesian methodology which leads to improved reduction of noise comparing with alternative methods. The new methods are introduced as Bayesian inference together with expectation-maximization procedure. It is worth noting that the BWA algorithm does not require setting of additional parameters in contrast to for example WACFM which needs value of an exponential parameter m. In the EBWA algorithms the parameter λ which influences performance of the procedure is estimated during iterations from input values by empirical method and the only parameter which must be set manually is parameter p. However in all performed experiments the best results appear for $p = 1$. Another advantage of presented method is fast convergence to the optimal result. In all performed experiments it did not require more than 10 iterations to stop for EBWA and 50 for BWA.

The results of numerical experiments show usefulness of the presented method in the noise reduction in ECG signal competitively to existing algorithms. The short computational study presented in this paper confirms that applying Bayesian inference has a practical and beneficial use to weighted averaging of biomedical signal in particular electrocardiographic signal.

REFERENCES

[1] B. Carlin and T. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall, New York, 1996.

[2] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis*, Chapman & Hall, New York, 2004.

[3] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley & Sons, Inc, New York, 2001.

[4] J. Łęski, "Application of time domain signal averaging and Kalman filtering for ECG noise reduction", *Ph.D Thesis*, Silesian University of Technology, Gliwice, 1989.

[5] J. Łęski, "Robust weighted averaging", *IEEE Transactions on Biomedical Engineering* 49 (8), 796–804 (2002).

[6] M. Figueiredo, "Adaptive sparseness for supervised learning", *IEEE Transaction on Pattern Analysis and Machine Learning* 25 (9), 1150–1159 (2003).

[7] H. White, *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge, 1996.

[8] A. Momot, M. Momot, and J. Łęski, "Empirical Bayesian averaging of biomedical signals", *Proc. XI International Conference MIT 2006*, 176–181 (2006).

[9] R. Adler, R. Feldman, and M. Taqqu, *A Practical Guide to Heavy Tails*, Birkhauser, Boston, 1998.

[10] P. Augustyniak, "Time-frequency modelling and discrimination of noise in the electrocardiogram", *Physiological Measurement* 24, 1–15 (2003).