Robust Weighted Averaging

Jacek M. Łęski, *Member, IEEE*

***Abstract*—Signal averaging is often used to extract a useful signal embedded in noise. This method is especially useful for biomedical signals, where the spectra of the signal and noise significantly overlap. In this case, traditional filtering techniques introduce unacceptable signal distortion. In averaging methods, constancy of the noise power is usually assumed, but in reality noise features a variable power. In this case, it is more appropriate to use a weighted averaging. The main problem in this method is the estimation of the noise power in order to obtain the weight values. Additionally, biomedical signals often contain outliers. This requires robust averaging methods. This paper shows that signal averaging can be formulated as a problem of minimization of a criterion function. Based on this formulation new weighted averaging methods are introduced, including weighted averaging based on criterion function minimization (WACFM) and robust ε-insensitive WACFM. Performances of these new methods are experimentally compared with the traditional averaging and other weighted averaging methods using electrocardiographic signal with the muscle noise, impulsive noise, and time-misalignment of cycles. Finally, an application to the late potentials extraction is shown.**

***Index Terms*—ε-insensitivity, noise reduction, robust methods, weighted averaging.**

I. Introduction

NOISE reduction is a first step in every biomedical signal processing system. Accuracy of all later operations performed on the signal, i.e., detections, measurements, classifications, depends on the quality of noise-reduction algorithms. Therefore, the improvement of noise-reduction methods is of practical importance. Often the spectra of the biomedical signals and noise overlap for a wide range of frequency [1]. For example, the spectra of electrocardiographic (ECG) signal significantly overlap the spectra of muscle noise and motion artifacts [2]. Similarly, the spectra of evoked potentials (EP) associated with visual and auditory stimulation overlap the spectra of the activity of other brain cells (background electroencephalogram—EEG) [3], [4] . One more common example is a noninvasive late potentials (LP) detection in high amplified ECG signal.

Biomedical signals listed above are periodical (or quasi-periodical) and therefore an averaging in time domain may be used. Signal averaging aggregate information from individual cycles of the periodical signal. Usually, as aggregation operation the arithmetic mean is used. In this case, the method is named ensemble averaging. This is the simplest method of noise reduction of periodical biomedical signal without the signal distortion. It leads to the noise-reduction factor equal to $\sqrt{N}$, where N is the number of averaged signals. The assumption is that the noise is stationary, zero mean, and not correlated with the signal [5].

A preprocessing step prior to the averaging should determine for each cycle, the so-called fiducial point required by the time-alignment [6], and a similarity measure to dominant morphology necessary to make a decision about an inclusion or exclusion cycle from further processing (classification) [2]. Errors in time-alignment and classification of signal cycles cause additional distortion of the averaged signal.

In fact, most types of noise are not stationary, i.e., the noise power measured by the noise variance, features some variability. In this case, it is better to use a weighted mean as the aggregation operation. In literature, there is a number of approaches to weighted averaging, including methods based on the minimum energy principle (MEP) [7], [8], maximizing the signal-to-noise ratio (SNR) using Rayleigh quotient and generalized eigenvalue problem (GEP) [3], estimating weights adaptationally [9], [10] based on Kalman filter (KF) theory [5], [11], and with highly quantized weights [12].

Let us assume that each signal cycle is the sum of deterministic (useful) signal and a random noise

$$x_i(k) = s(k) + n_i(k) \tag{1}$$

where i is the cycle number, $i = 1, 2, \ldots, N$; k is the discrete time index, $k = 1, 2, \ldots, p$. The deterministic component $s(k)$ is the same in all cycles. The noise $n_i(k)$ is a random component with zero mean and variance $\sigma_i^2(k)$. Usually, in $\sigma_i^2(k)$ the index k is omitted. This is equivalent to an assumption that in each cycle the noise variance is constant.

The weighted average is given by

$$v(k) = \sum_{i=1}^{N} w_i x_i(k) \tag{2}$$

where w_i is a weight for the ith signal cycle and $v(k)$ is the averaged signal. It can be easily proved that optimal weights (minimizing the mean square error) are function of the noise variance of all signal cycles and for Gaussian noise are given by [5] and [10]

$$w_i = \left[\sum_{j=1}^{N} \frac{\sigma_i^2}{\sigma_j^2} \right]^{-1}. \tag{3}$$

If the noise variance is constant for all cycles $\sigma_i^2 = \sigma^2$, then the optimal weights are equal to $w_i = 1/N$ for $i = 1, 2, \ldots, N$. In this case, we obtain the traditional ensemble averaging with arithmetic mean as the aggregation

Manuscript received August, 27 2001; revised February 19, 2002. This work was supported in part by a National Grant KBN 4T11E00622.

The author is with the Division of Biomedical Electronics, Institute of Electronics, Silesian University of Technology, Akademicka 16, 44-101 Gliwice, Poland. He is also with the Institute of Medical Technology and Equipment, Roosevelt St. 118A, 41-800 Zabrze, Poland (e-mail: jl@boss.iele.polsl.gliwice.pl).

Publisher Item Identifier 10.1109/TBME.2002.800757.

operation. When the noise has a non-Gaussian distribution, the estimate (2) is not optimal, but it is still the best of all linear estimators [5]. The weights in (3) are optimal if the variances of noise in cycles σ_i^2 are known. It is impossible to measure them directly. So, for example, in [11] the noise variance is estimated using the variance of innovation process of KF, in [10] adaptive process compute the optimal weights without direct estimation of the noise variance. In this paper, a new weighted averaging method based on criterion function minimization is proposed. In this case, direct estimation of noise variance is also not necessary.

One of the greatest disadvantage of traditional and weighted methods is their sensitivity to the presence of outliers caused by, i.e., spike artifacts, included cycles with nondominant morphology, bursts of noise and baseline shifts. In this case, computed averaged signal $v(k)$ can be very different from the true signal $s(k)$. If the signal is corrupted by outliers, then the assumed (for simplicity) models approximate only the reality. For example, if we assume that the distribution of noise is Gaussian, then using weighted mean should not cause a bias. In this case, L_2 (Euclidean) norm is used as dissimilarity measure of a disturbed signal from its true value. Usually outliers exist in biomedical signals. This requires robust averaging methods to be used. According to Huber [13], a robust method should have the following properties: 1) it should have a reasonably good accuracy at the assumed model; 2) small deviations from the model assumptions should impair the performance only by a small amount, and 3) larger deviations from the model assumptions should not cause a catastrophe.

In literature, there is a number of robust averaging methods [14]. By choosing a median as the aggregation operation, it is possible to reduce the effect of outliers. But in general, median averaging provides less noise reduction. By using a trimmed mean it is possible to reduce the effect of outliers and achieve a better noise reduction than obtained by the median [2]. In the above averaging methods, again constant noise variance is assumed.

The goal of this paper is to establish a connection between weighted signal averaging and robust statistics using Vapnik's ε-insensitive estimator. This paper presents a new robust weighted averaging method based on a new weighted ε-insensitive estimator. The weighted median averaging can be obtained as a special case of the introduced method.

The remaining part of this paper is organized as follows: Section II presents an averaging as a criterion function minimization. A novel robust weighted averaging is described in Section III. Section IV presents simulation results of ECG cycles averaging with the muscle noise, impulsive noise and time-misalignment. Application to the late potentials extraction is also shown. Finally, conclusions are drawn in Section V.

II. Averaging as a Minimization of Criterion Function

Let us start with some vector notations: $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]^T$ is the ith signal cycle, $\mathbf{v} = [v_1, v_2, \ldots, v_p]^T$ is the averaged signal, $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T$ is the weight vector. The set of all possible weight vectors for N signal cycles is defined as

$$\Im_w = \left\{ \mathbf{w} \in \Re_N \,\middle|\, \underset{1 \le i \le N}{\forall} \; w_i \in [0, 1], \; \sum_{i=1}^{N} w_i = 1 \right\} \quad (4)$$

where $\Re_N$ denotes a space of all real N-dimensional vectors. The property that sum of weights is equal to one, leads to an unbiased estimate of the deterministic component $\mathbf{s}$.

Now, we define the following scalar criterion function:

$$I_m(\mathbf{w}, \mathbf{v}) = \sum_{i=1}^{N} (w_i)^m \, \varrho(\mathbf{x}_i - \mathbf{v}) \quad (5)$$

where $\varrho(\cdot)$ is a measure of dissimilarity for vector argument, $m \in (1, \infty)$ is a weighting exponent parameter.

Scalar criterion function (5) can be interpreted as a measure of total dissimilarity between $\mathbf{v}$ and signal cycles $\mathbf{x}_i$ weighted by $(w_i)^m$, where $m > 1$. We search for $\mathbf{v}^\star$ that has minimal dissimilarity to all $\mathbf{x}_i$ using optimal weights $\mathbf{w}^\star$. The problem of searching for an optimal weight vector $\mathbf{w}^\star$ and an optimal averaged signal $\mathbf{v}^\star$, can be written in the form

$$I_m(\mathbf{w}^\star, \mathbf{v}^\star) = \min_{(\mathbf{w}, \mathbf{v}) \in (\Im_w \times \Re_p)} I_m(\mathbf{w}, \mathbf{v}). \quad (6)$$

Now, we obtain the necessary condition for minimization of (5) with respect to the weights vector. If $\mathbf{v} \in \Re_p$ is fixed, then the Lagrangian of (5) with constraint from (4) is

$$L(\mathbf{w}, \lambda) = \sum_{i=1}^{N} (w_i)^m \varrho(\mathbf{x}_i - \mathbf{v}) - \lambda \left[\sum_{i=1}^{N} w_i - 1 \right] \quad (7)$$

where λ is the Lagrange multiplier. Setting the Lagrangian's gradient to zero, we obtain

$$\frac{\partial L(\mathbf{w}, \lambda)}{\partial \lambda} = \sum_{i=1}^{N} w_i - 1 = 0 \quad (8)$$

and

$$\underset{1 \le j \le N}{\forall} \quad \frac{\partial L(\mathbf{w}, \lambda)}{\partial w_j} = m(w_j)^{m-1} \varrho(\mathbf{x}_j - \mathbf{v}) - \lambda = 0. \quad (9)$$

From (9) we get

$$w_j = \left(\frac{\lambda}{m} \right)^{1/(m-1)} [\varrho(\mathbf{x}_j - \mathbf{v})]^{1/(1-m)}. \quad (10)$$

From (8) and (10), we obtain

$$\left(\frac{\lambda}{m} \right)^{1/(m-1)} \sum_{j=1}^{N} [\varrho(\mathbf{x}_j - \mathbf{v})]^{1/(1-m)} = 1. \quad (11)$$

Combining (10) and (11) yields

$$\underset{1 \le i \le N}{\forall} \; w_i = [\varrho(\mathbf{x}_i - \mathbf{v})]^{1/(1-m)} \Bigg/ \left[\sum_{j=1}^{N} [\varrho(\mathbf{x}_j - \mathbf{v})]^{1/(1-m)} \right]. \quad (12)$$

It is postulated, that for all vectors $\mathbf{y}, \mathbf{z} \in \Re_p$ the dissimilarity measure ϱ satisfies the following properties:

1) $\varrho(\mathbf{0}) = 0$;
2) $\varrho(\mathbf{y}) = \varrho(-\mathbf{y})$—symmetry;
3) $\forall_{1 \le j \le p}\, y_j \le z_j \Longrightarrow \varrho(\mathbf{y}) \le \varrho(\mathbf{z})$—monotonicity.

There are many possible measures $\varrho(\cdot)$, which satisfy properties 1)–3). In order to obtain the necessary condition for minimization of (5) with respect to the averaged signal, we assume a simple and the most frequently used quadratic function: $\varrho(\mathbf{e}) = \mathbf{e}^T\mathbf{e} = \|\mathbf{e}\|_2^2$. If we assume that $\mathbf{w} \in \Im_w$ is fixed, the quadratic function as the dissimilarity measure is used, and the criterion function's (5) gradient with respect to an averaged signal $\mathbf{v}$ is set to zero, then we obtain

$$\frac{\partial I_m(\mathbf{w}, \mathbf{v})}{\partial \mathbf{v}} = -2\sum_{i=1}^{N}(w_i)^m(\mathbf{x}_i - \mathbf{v}) = \mathbf{0}. \tag{13}$$

From (13), we get

$$\mathbf{v} = \left[\sum_{i=1}^{N}(w_i)^m \mathbf{x}_i\right] \Bigg/ \left[\sum_{i=1}^{N}(w_i)^m\right]. \tag{14}$$

Using the quadratic function as the dissimilarity measure in (12) yields

$$\underset{1\le i\le N}{\forall} w_i = [\|\mathbf{x}_i - \mathbf{v}\|_2]^{2/(1-m)} \Bigg/ \left[\sum_{j=1}^{N}[\|\mathbf{x}_j - \mathbf{v}\|_2]^{2/(1-m)}\right]. \tag{15}$$

The optimal solution $(\mathbf{w}^\star, \mathbf{v}^\star) \in (\Im_w \times \Re_p)$ is a fixed point of (14) and (15), and it is obtained from the Picard algorithm. This algorithm can be called weighted averaging based on criterion function minimization (WACFM), and can be described as follows:

1° Fix $m \in (1, \infty)$. Initialize $\mathbf{v}^{(0)} \in \Re_p$. Set the iteration index $k = 1$;
2° Calculate the weight vector $\mathbf{w}^{(k)}$ for the kth iteration using (15);
3° Update the averaged signal for the kth iteration $\mathbf{v}^{(k)}$ using (14) and $\mathbf{w}^{(k)}$;
4° If $\|\mathbf{w}^{(k)} - \mathbf{w}^{(k-1)}\|_2 > \xi$ then $k \leftarrow k+1$ and go to 2°.

ξ is a preset parameter and the upper index (k) denotes an iteration number.

Equation (12) can be written as

$$\underset{1\le i\le N}{\forall} \quad w_i = \left[\sum_{j=1}^{N}\left(\frac{\varrho(\mathbf{x}_i - \mathbf{v})}{\varrho(\mathbf{x}_j - \mathbf{v})}\right)^{1/(m-1)}\right]^{-1} \tag{16}$$

and it is easy to observe that

$$\lim_{m\to 1^+} w_i = \begin{cases} 1, & \varrho(\mathbf{x}_i - \mathbf{v}) \le \varrho(\mathbf{x}_k - \mathbf{v}) \\ & \text{for all } k \ne i; \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

In other words, 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 tends to infinity, then weights tend to $w_i = 1/N$ for all i. In this case, the traditional averaging is obtained. Generally, a larger m results in a smaller influence of the dissimilarity measures $\varrho(\mathbf{x}_i - \mathbf{v})$.

If the quadratic function as the dissimilarity measure is used, and the averaged signal has a low noise power, then $\|\mathbf{x}_i - \mathbf{v}\|_2^2 \cong N\,\sigma_i^2$. Taking this into account and using $m = 2$ in (16), we obtain previous results (3). Using $m = 2$ in (14), we obtain a different formula from the traditional one, shown in (2). This formula is obtained for $m = 1$, but in this case, we obtain trivial optimal weights (17). Using exponent $m = 2$ in (14) results in greater decrease of medium weights [about 0.5, because $(w_i)^2 - w_i$ has a maximum in 0.5], which corresponds to signal cycles with greater uncertainty about inclusion to averaging process.

The traditional solution can be obtained for $m = 2$, by adding one step to the previously described algorithm.

5) Calculate averaged signal using $\mathbf{w}^\star$ in (2).

This algorithm can be called modified WACFM (MWACFM).

III. ε-Insensitive Averaging

The combination of the previously obtained necessary condition for weights (12) with the criterion function (5) yields

$$I_m(\mathbf{v}) = 1 \Bigg/ \left[\sum_{i=1}^{N}\left(\frac{1}{\varrho(\mathbf{x}_i - \mathbf{v})}\right)^{1/(m-1)}\right]^{m-1}. \tag{18}$$

The expression $(1/(\varrho(\mathbf{x}_i - \mathbf{v}))^{1/(m-1)}$ is a measure of similarity between $\mathbf{x}_i$ and $\mathbf{v}$. Thus, denominator of (18) is a measure of the total similarity between $\mathbf{x}_i$s and $\mathbf{v}$. Finally, $I_m(\mathbf{v})$ is a measure of the total dissimilarity between $\mathbf{x}_i$s and $\mathbf{v}$. Our goal is to minimize this dissimilarity.

Averaging methods described in Section II use a quadratic function as a dissimilarity measure. This measure is used for mathematical reasons, that is, for simplicity and a low computational burden. However, this approach is sensitive to outliers. If outliers are present in a signal, then the averaged signal $\mathbf{v}$ is attracted toward them. In the literature, there are many outliers robust functions $\varrho(\cdot)$ [13].

One of the simplest function of this kind is the absolute value: $\varrho(\mathbf{e}) = \sum_{j=1}^{p}|e_j| = \|\mathbf{e}\|_1$. Using this function and assuming in (5) that $w_i = 1/N$ lead to the following criterion function: $(1/N)\sum_{i=1}^{N}\sum_{j=1}^{p}|x_{ij} - v_j|$. The minimization of this criterion can be decomposed to minimization of p subcriterions: $(1/N)\sum_{i=1}^{N}|x_{ij} - v_j|$. It is well known (cf. [15]) that minimization of the above subcriterions leads to the median of samples, i.e., $v_j = \text{median}(x_{1j}, x_{2j}, \ldots, x_{Nj})$. If weights are not equal to $1/N$, then we obtain an estimator that can be called a weighted median.

In the remainder of this section, a more general $\varrho(\cdot)$ function will be used. This function uses the Vapnik ε-insensitive function [16]

$$\rceil\mathbf{e}\lceil_\varepsilon = \sum_{j=1}^{p}\,, \rceil e_j\lceil_\varepsilon \tag{19}$$

where

$$\rceil e_j\lceil_\varepsilon = \begin{cases} 0, & |e_j| \le \varepsilon \\ |e_j| - \varepsilon, & |e_j| > \varepsilon \end{cases} \tag{20}$$

and ε denotes the insensitivity parameter. Adding to the signal a value less than ε does not change the dissimilarity measure. The well-known absolute error function is a special case of (19), (20) for $\varepsilon = 0$. The main goal of this section is to apply a weighted version of the ε-insensitive function to robust signal averaging.

Using the ε-insensitive function the criterion function (5) takes the form

$$I_{m\varepsilon}(\mathbf{w}, \mathbf{v}) = \sum_{i=1}^{N} (w_i)^m \rceil \mathbf{x}_i - \mathbf{v} \lceil_{\varepsilon}. \tag{21}$$

Combination (21) and (19) yields

$$I_{m\varepsilon}(\mathbf{w}, \mathbf{v}) = \sum_{i=1}^{N} (w_i)^m \sum_{j=1}^{p} \rceil x_{ij} - v_j \lceil_{\epsilon} = \sum_{j=1}^{p} g_j(v_j) \tag{22}$$

where

$$g_j(v_j) = \sum_{i=1}^{N} (w_i)^m \rceil x_{ij} - v_j \lceil_{\varepsilon} \tag{23}$$

and can be called the weighted ε-insensitive estimator. For $\varepsilon = 0$, we obtain the weighted median estimator.

Our problem of minimization criterion (21) with respect to the averaged signal $\mathbf{v}$ can be decomposed to p minimization problems of (23) for $j = 1, \ldots, p$. In a general case, not for all data, x_{ij} the inequalities: $x_{ij} - v_j \le \varepsilon$, $v_j - x_{ij} \le \varepsilon$ are satisfied. If we introduce slack variables $\xi_{ij}^{+}, \xi_{ij}^{-} \ge 0$, then for all data x_{ij} we can write

$$\begin{cases} v_j - x_{ij} \le \varepsilon + \xi_{ij}^{+} \\ x_{ij} - v_j \le \varepsilon + \xi_{ij}^{-}. \end{cases} \tag{24}$$

Now, the criterion (23) can be written in the form

$$g_j(v_j) = \sum_{i=1}^{N} (w_i)^m (\xi_{ij}^{+} + \xi_{ij}^{-}) \tag{25}$$

and is minimized subject to constraints (24) and $\xi_{ij}^{+} \ge 0$, $\xi_{ij}^{-} \ge 0$. The Lagrangian function of (25) with the above constraints is

$$\begin{aligned} G_j(v_j) = &\sum_{i=1}^{N} (w_i)^m (\xi_{ij}^{+} + \xi_{ij}^{-}) - \sum_{i=1}^{N} (\mu_{ij}^{+} \xi_{ij}^{+} + \mu_{ij}^{-} \xi_{ij}^{-}) - \sum_{i=1}^{N} \lambda_{ij}^{+} \\ &\cdot (\varepsilon + \xi_{ij}^{+} - v_j + x_{ij}) - \sum_{i=1}^{N} \lambda_{ij}^{-} (\varepsilon + \xi_{ij}^{-} + v_j - x_{ij}) \end{aligned} \tag{26}$$

where $\lambda_{ij}^{+}, \lambda_{ij}^{-}, \mu_{ij}^{+}, \mu_{ij}^{-} \ge 0$ are the Lagrange multipliers. The objective is to minimize this Lagrangian with respect to v_j, ξ_{ij}^{+}, ξ_{ij}^{-}. It must also be maximized with respect to the Lagrange multipliers. The following optimality conditions (the saddle point of the Lagrangian) are obtained by differentiating (26) with respect to v_j, ξ_{ij}^{+}, ξ_{ij}^{-} and setting the results equal to zero

$$\begin{cases} \dfrac{\partial G_j(v_j)}{\partial v_j} = \displaystyle\sum_{i=1}^{N} (\lambda_{ij}^{+} - \lambda_{ij}^{-}) = 0 \\ \dfrac{\partial G_j(v_j)}{\partial \xi_{ij}^{+}} = (w_i)^m - \lambda_{ij}^{+} - \mu_{ij}^{+} = 0 \\ \dfrac{\partial G_j(v_j)}{\partial \xi_{ij}^{-}} = (w_i)^m - \lambda_{ij}^{-} - \mu_{ij}^{-} = 0. \end{cases} \tag{27}$$

The last two conditions (27) and the requirements $\mu_{ij}^{+}, \mu_{ij}^{-} \ge 0$ imply $\lambda_{ij}^{+}, \lambda_{ij}^{-} \in [0, (w_i)^m]$. By putting conditions (27) in the Lagrangian (26) we get

$$G_j(v_j) = -\sum_{i=1}^{N} (\lambda_{ij}^{+} - \lambda_{ij}^{-}) x_{ij} - \varepsilon \sum_{i=1}^{N} (\lambda_{ij}^{+} + \lambda_{ij}^{-}). \tag{28}$$

Maximization of (28) with respect to $\lambda_{ij}^{+}, \lambda_{ij}^{-}$, subject to constraints

$$\begin{cases} \displaystyle\sum_{i=1}^{N} (\lambda_{ij}^{+} - \lambda_{ij}^{-}) = 0 \\ \lambda_{ij}^{+}, \lambda_{ij}^{-} \in [0, (w_i)^m] \end{cases} \tag{29}$$

is the so called Wolfe dual formulation (problem) [16]. It is well known in the optimization theory that at the saddle point, for each Lagrange multiplier, the Karush–Kühn–Tucker conditions must be satisfied

$$\begin{cases} \lambda_{ij}^{+} (\varepsilon + \xi_{ij}^{+} - v_j + x_{ij}) = 0 \\ \lambda_{ij}^{-} (\varepsilon + \xi_{ij}^{-} + v_j - x_{ij}) = 0 \\ ((w_i)^m - \lambda_{ij}^{+}) \xi_{ij}^{+} = 0 \\ ((w_i)^m - \lambda_{ij}^{-}) \xi_{ij}^{-} = 0. \end{cases} \tag{30}$$

The last two conditions (30) show that $\lambda_{ij}^{+} \in (0, (w_i)^m) \Longrightarrow \xi_{ij}^{+} = 0$ and $\lambda_{ij}^{-} \in (0, (w_i)^m) \Longrightarrow \xi_{ij}^{-} = 0$. In this case, the first two conditions (30) yield

$$\begin{cases} v_j = x_{ij} + \varepsilon, & \text{for } \lambda_{ij}^{+} \in (0, (w_i)^m) \\ v_j = x_{ij} - \varepsilon, & \text{for } \lambda_{ij}^{-} \in (0, (w_i)^m). \end{cases} \tag{31}$$

Thus, we may determine the averaged signal samples v_j from (31) by taking any x_{ij} for which the Lagrange multipliers are within the open interval $(0, (w_i)^m)$. From numerical point of view, it is better to take the mean value of v_j obtained for all data which satisfy the conditions (31)

$$\underset{1 \le j \le p}{\forall} \quad v_j = \frac{\left[\displaystyle\sum_{\{i | \lambda_{ij}^{+} \in \Lambda_j^{+}\}} (x_{ij} + \varepsilon) + \sum_{\{i | \lambda_{ij}^{-} \in \Lambda_j^{-}\}} (x_{ij} - \varepsilon) \right]}{\operatorname{card}(\Lambda_j^{+} \cup \Lambda_j^{-})} \tag{32}$$

where card(A) denotes cardinality of a set A, $\lambda^{(\pm)}$ denotes λ^+ or λ^- and

$$\Lambda_j^{(\pm)} = \left\{ \lambda_{ij}^{(\pm)} \in (0, (w_i)^m) \,\middle|\, \begin{array}{l} \min\limits_{\{\lambda_{ij}^+\},\{\lambda_{ij}^-\}} \sum\limits_{i=1}^{N} (\lambda_{ij}^+ - \lambda_{ij}^-) x_{ij} + \varepsilon \sum\limits_{i=1}^{N} (\lambda_{ij}^+ + \lambda_{ij}^-), \\ \text{subject to} \begin{cases} \sum\limits_{i=1}^{N} \lambda_{ij}^+ = \sum\limits_{i=1}^{N} \lambda_{ij}^-, \\ \lambda_{ij}^+, \lambda_{ij}^- \in [0, (w_i)^m]. \end{cases} \end{array} \right\}. \quad (33)$$

On the basis of (12), (19), (32) and (33) we obtain an algorithm that can be called ε-insensitive WACFM (ε WACFM).

1) Fix $m \in (1, \infty)$ and $\varepsilon \geq 0$. Initialize $\mathbf{v}^{(0)} \in \Re_p$, Set the iteration index $k = 1$.
2) Calculate the weight $\mathbf{w}^{(k)}$ for the kth iteration using (12), (19), and (32).
3) Update the averaged signal for the kth iteration $\mathbf{v}^{(k)}$ using (32), (33), and $\mathbf{w}^{(k)}$.
4) If $\|\mathbf{w}^{(k)} - \mathbf{w}^{(k-1)}\|_2 > \xi$, then $k \leftarrow k + 1$ and go to 2°.

IV. Numerical Experiments and Discussion

In all experiments using WACFM, MWACFM, and εWACFM calculations were initialized as the mean of disturbed signal cycles. Iterations were stopped as soon as the L_2 norm for a successive pair of $\mathbf{w}$ vectors was less than 10^{-5} for the WACFM, MWACFM, and 10^{-2} for the εWACFM. 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 MATLAB environment. The linear optimization with constraints was performed using the MATLAB "linprog" procedure.

A. Signals With Gaussian Noise

The simulated ECG signal cycles were obtained as the same deterministic component with added different 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 a high signal-to-noise ratio. Before averaging these cycles were time-aligned using the cross-correlation method.

The purpose of this experiment was to investigate the proposed in this paper averaging methods in presence of Gaussian noise, where optimal weights can be determined by (3). According to Huber's definition (see Section I), we have: case 1) for the MWACFM method; and case 2) for the εWACFM method. A series of 100 ECG cycles was generated with the same deterministic component and zero-mean white Gaussian noise

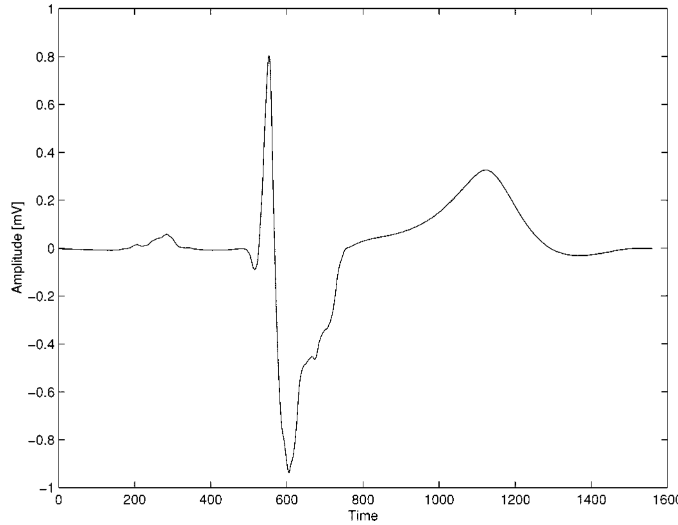

Fig. 1. Original ECG signal—deterministic component.

TABLE I
RMSE and Maximum Error for Averaged ECG Signals With Gaussian Noise

Method	Optimal	Optimal with σ_i^2 calculation	MWACFM $m=2$	WACFM $m=1.5$	WACFM $m=2$
RMSE [μV]	1.9297	1.9179	1.9242	1.9569	1.9323
MAX [μV]	5.3146	5.1241	5.1621	5.8359	5.7341
Method	**WACFM $m=4$**	**WACFM $m=100$**	**εWACFM $\varepsilon=0\mu V$**	**εWACFM $\varepsilon=0.5\mu V$**	**εWACFM $\varepsilon=1\mu V$**
RMSE [μV]	1.9275	1.9257	1.9254	1.9230	1.9437
MAX [μV]	5.5589	5.2034	5.1621	5.1552	5.3502

with four different standard deviations. For the first, second, third and fourth 25 cycles, the noise standard deviations were 10, 50, 100, and 200 μV, respectively. These signal cycles were averaged using the following methods: optimal weights from (3), optimal weights from (3) with σ_i^2 estimated directly from the noise, the WACFM with m equal to 1.5, 2, 4, and 100, the MWACFM with $m = 2$, and the εWACFM with ε equal to 0, 0.5, and 1 μV. Subtraction of the deterministic component from these averaged signals gives a residual noise.

The RMSE and the maximal value (MAX) of residual noise for all tested methods are presented in Table I. It shows that the smallest RMSE and MAX error were obtained by an optimal method with σ_i^2 estimated directly from noise. A little bit worse results were obtained by the optimal method based on an assumption that the variances of generated noise were known. The above averaging methods can not be applied to real signals processing, because the noise variance is not known. These methods were used as reference, because they offer in Gaussian case the best possible noise reduction.

The best noise reduction for the remaining methods is obtained by the MWACFM method, with RMSE and MAX errors close to reference methods. A little worse results were obtained by the WACFM with $m = 4$ and the εWACFM with $\varepsilon = 0.5$. Increasing the value of parameter m results in an increase of RMSE and MAX. For the εWACFM method, selection of the

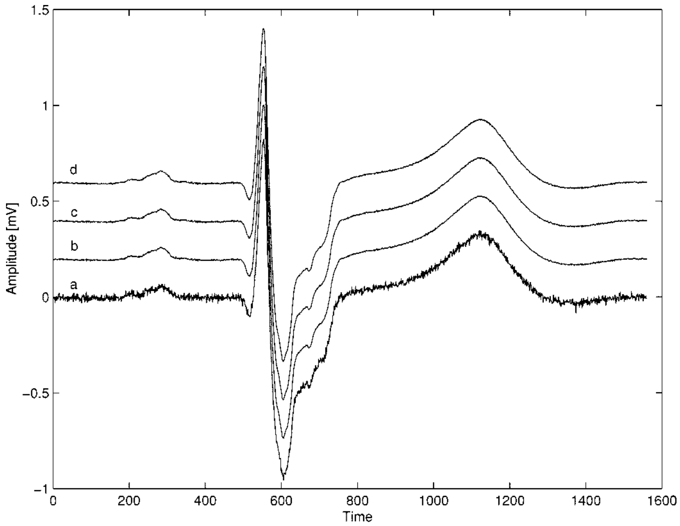

Fig. 2. Result of ECG signal averaging: (a) traditional, (b) optimal, (c) MWACFM with $m = 2$, (d) εWACFM with $\varepsilon = 0$. Signals (b), (c), and (d) are shifted by 0.2, 0.4, and 0.6 mV, respectively, for better presentation.

insensitivity parameter $\varepsilon = 0.5$ μV implies the best noise reduction. From Table I we also see that WACFM and ε WACFM are not very sensitive to the choice of m and ε parameters. The reasonable choice includes $m \in (1.5, 100)$ and ε equal to a few analog-to-digital (A/D) converter quantization levels.

The results of ECG signal cycles averaging are shown in Fig. 2. Signal (a) was obtained using the traditional (equally weights) method, (b) using optimal method with σ_i^2 estimated directly from noise, (c) using the MWACFM with $m = 2$, and (d) using the εWACFM with $\varepsilon = 0$. Note that for better presentation signals (b), (c), and (d) are shifted by 0.2, 0.4, and 0.6 mV, respectively.

B. Signals With Gaussian and Impulsive Noise

The purpose of this experiment was to investigate the proposed averaging methods in case of Gaussian and impulsive noise. According to Huber's definition we have the case 2) or 3) for all methods. Again, a series of 100 ECG cycles were generated with the same deterministic component (presented in Fig. 1), Gaussian noise with levels as previously described, and additionally impulsive noise. The impulsive noise was generated as a Bernoulli–Gaussian sequence [17]: $\mu(k) = r(k)q(k)$. In this model, $q(k)$ is a Bernoulli sequence, i.e., a random sequence of zeros and ones, with parameter λ

$$\text{Prob}[q(k)] = \begin{cases} \lambda, & q(k) = 1 \\ 1 - \lambda, & q(k) = 0. \end{cases} \tag{34}$$

The $r(k)$ is a zero-mean Gaussian white noise with variance σ_r^2. Sequences $r(k)$ and $q(k)$ are statistically independent. In this experiment, the following values of parameters were used: $\lambda = 0.2$ and $\sigma_r^2 = 1$ mV. An example of ECG cycle with the above defined impulsive noise and 20 μV standard deviation white Gaussian noise is presented in Fig. 3.

The RMSE and MAX values of residual noise obtained by the tested methods are presented in Table II. It shows that the smallest RMSE and MAX were obtained by the εWACFM method with $\varepsilon = 1$ μV. Increase of the insensitivity parameter ε

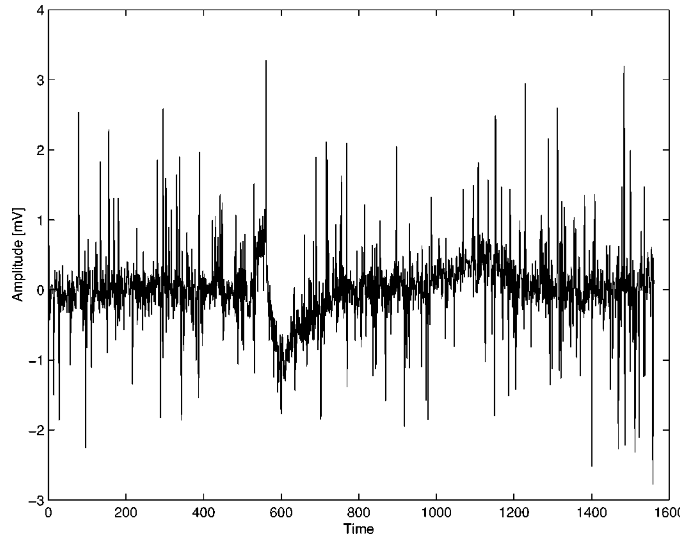

Fig. 3. Example of ECG signal cycle with Gaussian and impulsive noise.

TABLE II
RMSE FOR AVERAGED ECG SIGNALS WITH GAUSSIAN AND IMPULSIVE NOISE

Method	WACFM $m=2$	MWACFM $m=2$	εWACFM $\varepsilon=0\mu V$	εWACFM $\varepsilon=0.5\mu V$
RMSE $[\mu V]$	46.9738	46.0129	5.0170	4.9779
MAX $[\mu V]$	129.4084	119.8513	19.3639	19.2317
Method	εWACFM $\varepsilon=1\mu V$	εWACFM $\varepsilon=2\mu V$	εWACFM $\varepsilon=5\mu V$	εWACFM $\varepsilon=10\mu V$
RMSE $[\mu V]$	4.9037	4.9186	5.1856	5.1941
MAX $[\mu V]$	19.0691	19.1083	16.4611	29.5916

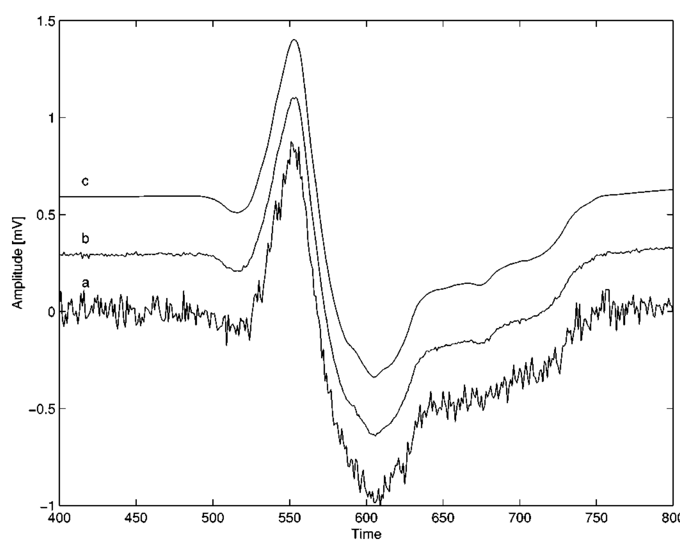

Fig. 4. Result of ECG signal averaging: (a) WACFM with $m = 2$, (b) εWACFM with $\varepsilon = 1$ μV, (c) deterministic component. Note that signals are enlarged (samples from 400 to 800 from previous figure) and signals (b) and (c) are shifted for better presentation.

implies a decrease of RMSE. Much worse results were obtained by the WACFM with $m = 2$ and the MWACFM with $m = 2$. The "optimal" method with σ_i^2 estimated directly from the noise leads to RMSE $= 52.7726$ μV and MAX $= 131.4337$ μV. The above results are expected because the method is optimal

TABLE III
RMSE FOR AVERAGED ECG SIGNALS WITH MISALIGNMENT ERROR AND GAUSSIAN NOISE

Method	Optimal	MWACFM $m=2$	WACFM $m=2$
RMSE $[\mu V]$	52.0157	19.0432	18.9448
MAX $[\mu V]$	94.8025	51.6805	51.3060
Method	εWACFM $\varepsilon=0\mu V$	εWACFM $\varepsilon=0.5\mu V$	εWACFM $\varepsilon=1\mu V$
RMSE $[\mu V]$	18.4708	18.2750	18.0884
MAX $[\mu V]$	46.6203	46.5094	46.4012

for the Gaussian noise only. From Table II we again see that the εWACFM is not very sensitive to the choice of the parameter ε.

The results of averaging of ECG signal cycles are presented in Fig. 4. Signal (a) was obtained using the MWACFM method, (b) using the εWACFM with $\varepsilon = 1$ μV. In Fig. 4, the deterministic component (c) is also presented for comparison. Note that signals (b) and (c) are shifted by 0.3 and 0.6 mV, respectively, for better presentation.

C. Time-Misaligned Signals With Gaussian Noise

The purpose of this experiment was to investigate the proposed averaging methods in the case of time-misaligned signal cycles corrupted by the Gaussian noise. According to Huber's definition, we have case 2) or 3) for all methods. The misalignment is caused by the quantization and noise of averaged cycles. This results in a low-pass filtering of the averaged signal cycles. The low-pass filter characteristic is described by the distribution of the fiducial point jitter.

A series of 100 ECG cycles were generated with the same deterministic component (presented in Fig. 1). Each signal cycle was delayed with a jitter following the Gaussian probability density function with a zero-mean and the standard deviation equal to six sampling periods. To make the signal delaying operation possible, each realization of Gaussian jitter was rounded to the nearest integer. The Gaussian noise with levels as in Section IV-A were added to these time-misaligned deterministic components.

These signals were averaged using the following methods: optimal with σ_i^2 estimated directly from the noise, the WACFM with $m = 2$, the MWACFM, and the εWACFM with $m = 2$, $\varepsilon = 0, 0.5, 1$ μV. The RMSE and MAX values of the residual noise for the tested methods are presented in Table III. The results show clearly the superiority of the minimization-of-criterion-function methods. The best result is obtained for εWACFM with $\varepsilon = 1$ μV. Again we see that εWACFM is not very sensitive to the choice of the parameter ε.

The results of time-misaligned signal cycles averaging are presented in Fig. 5: the solid line represents the deterministic component with a zero-delay, the dotted line represents averaged signal by an optimal method, and the dashed line represents the averaged signal by εWACFM with $\varepsilon = 1$ μV. It is easy to explain the results obtained in this subsection. Signal cycles with small similarity to the averaged signal $\mathbf{v}$ have small weights, and the low-pass filtering effect is reduced.

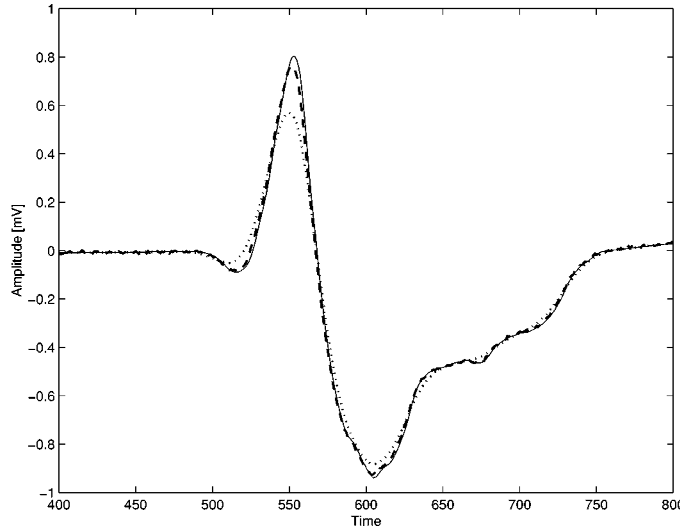

Fig. 5. Result of misaligned ECG signal averaging: (solid line) deterministic component, (dotted line) optimal method, (dashed line) εWACFM with $\varepsilon = 1$ μV.

TABLE IV
COMPARISON OF WEIGHTED AVERAGING METHODS

Test	Method	MEP	GEP	AFA	KF
1	RMSE	12.1953	12.0963	12.0886	1.9606
	MAX	41.6712	39.1728	40.0867	5.7948
2	RMSE	46.7231	46.6172	47.7721	46.1474
	MAX	119.6918	120.1288	124.4633	126.9785
3	RMSE	49.8926	49.8575	52.6917	64.3769
	MAX	98.5567	273.3925	110.6834	221.6549
4	Time	0.0733	0.4200	0.5641	0.0777

D. Comparison to Other Weighted Averaging Methods

The purpose of this subsection was to make a comparative study of the proposed methods performance to other weighted averaging methods. The following methods were tested: method based on the MEP [7], method leading to GEP [3], adaptive filter averaging (AFA) [10] and a method based on the KF theory [5], [11]. For this comparison the signals generated in Sections IV-A–IV-C signed as "test1," "test2," and "test3" were used. Additionally, as "test4" the time of computing the average of signals from Section IV-A was performed. All computations were run on Pentium IV 1.5-GHz processor running Windows NT4 and MATLAB environment.

The results are presented in Table IV. It shows that for Gaussian noise a little bit worse results comparing with the proposed methods are obtained by KF method only. The results obtained for impulsive noise are similar to WACFM and MWACFM methods and significantly worse comparing with the εWACFM method. In this case, the best result is obtained by KF method. The results obtained for the time-misaligned signals are much worse comparing with methods proposed in this work. In this case, the best result for other methods is obtained by GEP method.

The last row in Table IV reflects the time required to compute the average of signals from Section IV-A. For methods proposed in this work the following results were obtained:

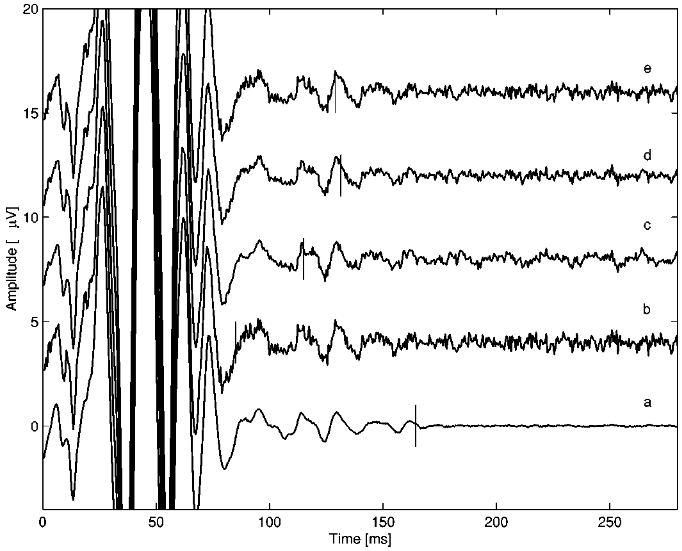

Fig. 6. Real ECG signal averaging for late potentials detection. (a) The traditional averaging of 1200 cycles. (b) The traditional averaging of 100 cycles, (c) MWACFM, (d) εWACFM with $\varepsilon = 1\ \mu$V, and (e) based on KF theory. Signals (b)–(e) are shifted by 4, 8, 12, and 16 μV, respectively, for better presentation. The endpoints are marked by vertical lines.

WACFM, 0.0533 s; MWACFM, 0.0733 s; and εWACFM, 62.3599 s. We can formulate the following remarks: 1) the WACFM method is the fastest; 2) a little bit slower are MWACFM, MEP, and KF methods; and 3) the great disadvantage of the εWACFM method is its computational burden; that is about 1100 fold greater comparing with (M)WACFM methods. However, in [18], a method to solve the linear or the quadratic programming problem by solving a linear inequalities system with computation time about 1000 times faster comparing with the traditional programming methods is introduced. Thus, application of this method helps to decrease the computational burden to twofold greater comparing with the AFA method.

E. ECG Late Potentials Extraction

The performance of the introduced averaging method is exemplified by means of real ECG recording. The ECG signal was recorded using low-noise amplifiers with the bandwidth 0.05–500 Hz, noise level 1.5 μV. The signal was digitized at a sampling rate 2000 Hz and quantized by 18-bit A/D converter (resolution 0.381 μV). The signal cycles alignment was done by the cross-correlation method. The reference signal was obtained by the traditional (equally weights) averaging of 1500 beats. For the testing purpose 50 beats were used. To prevent a nonstationarity effect every thirtieth beat was included to the test set. Signals from the test set were averaged using the following methods: the traditional, MWACFM, ε WACFM with $\varepsilon = 0.1$ μV, and based on KF.

The bidirectionally high-pass filtered averages are shown in Fig. 6. Signal (a) was obtained using the traditional averaging of 1500 beats, (b) using the traditional averaging of beats from the test set, (c) using the MWACFM method, (d) εWACFM with $\varepsilon = 1\ \mu$V, and (e) was obtained using the averaging based on KF. Note that the signals (b)–(e) are shifted by 4, 8, 12, and 16 μV, respectively, for better presentation. A fourth-order Butterworth type filter was applied with the cutoff frequency at 40 Hz. In Fig. 6, the endpoints are marked by vertical lines. These endpoints were determined as a time for which modulus of the mean in 5-ms interval exceeds a threshold. The threshold was defined by the mean noise level plus three times the standard deviation of the noise.

The accuracy of the late potentials detection in the time domain depends on the accuracy of the endpoint determination, which depends on the quality of noise reduction. Fig. 6 shows that the εWACFM method leads to the endpoint location which is the nearest to the endpoint determined on the reference signal. The farthest endpoint is located on the traditionally averaged signal. Locations of the endpoints obtained for signals averaged by MWACFM and KF methods are slightly moved into the QRS complex. We also see that the endpoint location error is a monotonically decreasing function of noise-reduction ability of the used averaging methods.

V. CONCLUSION

Biomedical signals contain noise and outliers and require robust processing methods. This paper establishes the connection between weighted averaging of signal and robust statistics. The introduced ε-insensitive weighted averaging method is based on weighted version of Vapnik's ε-insensitive function as a dissimilarity measure. A new method is introduced as a constrained minimization problem of the criterion function. The necessary conditions for obtaining minimum of the criterion function are shown. The averaging with weighted median can be obtained as a special case of the method introduced in this paper. A comparative study of the above methods are also included. These numerical examples show that the proposed method with quadratic function as dissimilarity measure reduces the Gaussian noise comparably to the optimal method. For weakly satisfied model assumption, in the case of impulsive noise, the ε-insensitive weighted averaging leads to the best results. This is caused by the robustness of the proposed method (εWACFM). In this case, deviations from the model assumption (for example noise distribution, noise and signal uncorrelation, etc.), should impair slightly the performance of the method. A comparative study of the performance of the proposed and weighted averaging methods known from literature is also presented. Finally, an application to the late potentials extraction shows usefulness of the proposed method to real biomedical signals processing.

ACKNOWLEDGMENT

The author would like to thank the anonymous referees for their valuable suggestions.

REFERENCES

[1] W. J. Tompkins, Ed., *Biomedical Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[2] O. Pahlm and L. Sörnmo, "Data processing of exercise ECG's," *IEEE Trans. Biomed. Eng.*, vol. BME-34, pp. 158–165, Feb. 1987.

[3] C. E. Davila and M. S. Mobin, "Weighted averaging of evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 338–345, Apr. 1992.

[4] L. Gupta, D. L. Molfese, R. Tammana, and P. G. Simos, "Nonlinear alignment and averaging for estimating the evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 43, pp. 338–345, Apr. 1996.

[5] J. Łęski, "Application of time domain signal averaging and Kalman filtering for ECG noise reduction," Ph.D. dissertation, Silesian Univ. Technol., Gliwice, Poland, 1989.

[6] R. Jané, H. Rix, P. Caminal, and P. Laguna, "Alignment methods for averaging of high-resolution cardiac signals: A comparative study of performance," *IEEE Trans. Biomed. Eng.*, vol. 38, pp. 571–579, June 1991.

[7] Z. Fan and T. Wang, "A weighted averaging method for evoked potential based on the minimum energy principle," in *Proc. IEEE EMBS Conf.*, vol. 13, 1991, pp. 411–412.

[8] ——, "Weighted averaging method for evoked potential: Determination of weighted coefficients," in *Proc. IEEE EMBS Conf.*, vol. 14, 1992, pp. 2473–2474.

[9] E. Bataillou, E. Thierry, and H. Rix, "Weighted averaging with weight estimation," in *Proc. IEEE Comp. Cardiol.*, Valenzia, Venezia, 1991, pp. 37–40.

[10] E. Bataillou, E. Thierry, H. Rix, and O. Meste, "Weighted averaging using adaptive estimation of the weights," *Signal Processing*, vol. 44, pp. 51–66, 1995.

[11] J. Łęski, "New concept of signal averaging in time domain," in *Proc. IEEE EMBS Conf.*, vol. 13, 1991, pp. 367–368.

[12] J. Łęski and N. Henzel, "Biomedical signal averaging with highly quantized weights," in *Proc. BMES/EMBS Conf.*, Atlanta, GA, 1999, p. 1013.

[13] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[14] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, K. Michler, and S. Luo, "Comparing stress ECG enhancement algorithms," *IEEE Eng. Med. Biol. Mag*, vol. 15, no. 3, pp. 37–44, May/June 1996.

[15] G. R. Arce, "A general weighted median filter structure admitting negative weights," *IEEE Trans. Signal Processing*, vol. 46, pp. 3195–3205, Dec. 1998.

[16] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[17] J. M. Mendel, *Optimal Seismic Deconvolution. An Estimation-Based Approach*. New York: Academic , 1983.

[18] J. Łęski, "Neuro-fuzzy modeling with ε-insensitive learning," in *Methods of Artificial Intelligence in Mechanics and Mechanical Engineering*. Gliwice, Poland: SUT Press, 2001, pp. 133–138.

Jacek M. Łęski (M'98) was born in Gliwice, Poland. He received the M.Sc., Ph.D., and D.Sc. degrees in electronics from Silesian University of Technology, Gliwice, Poland, in 1987, 1989, and 1995, respectively.

He is a Professor of Biomedical Information Processing at Silesian University of Technology and at present is, Head of its Division of Biomedical Electronics. He also is a Professor at Institute of Medical Technology and Equipment, Zabrze, Poland. His research interests are biomedical digital signal processing, fuzzy and neuro-fuzzy modeling, pattern recognition, and learning theory. He has coauthored three monographs, nine chapters in monographs, and more than 120 international journal and conference papers in these subjects.

Dr. Łęski is a Member the IEEE Engineering in Medicine and Biology Society, the IEEE Signal Processing Society, the IEEE System, Man, and Cybernetics Society, the Polish Biomedical Engineering Society, and the Polish Society of Theoretical and Applied Electrotechnics.