ADAPTIVE NOISE CANCELLATION FOR MULTI-SENSORY SIGNALS

SERGIY A. VOROBYOV, ANDRZEJ CICHOCKI

Laboratory for Advanced Brain Signal Processing, Brain Science Institute

The Institute of Physical and Chemical Research (RIKEN)
2-1 Hirosawa, Wako-shi, Saitama 351-0198 JAPAN
Email: {svor,cia}@bsp.brain.riken.go.jp

YEVGENIY V. BODYANSKIY

Control Systems Research Laboratory

Kharkiv State Technical University of Radioelectronics
14 Lenin Ave., Kharkiv 61166 UKRAINE
Email: bodya@kture.kharkov.ua

This paper describes a fast adaptive algorithm for noise cancellation using multi-sensory signal recordings of the same noisy source. It is shown that the performance of the new procedure for noise cancellation for multi-sensory signals is improved when compared to previously proposed methods. A short overview of the previously proposed methods is given. Optimality of the algorithm is discussed and numerical simulation is included to show the validity and effectiveness of the algorithm.

Keywords: Noise cancellation; Multi-sensory signal; Adaptive filtering; Optimization; Learning.

1. Introduction

Noise cancellation is a special case of optimal filtering which can be applied when some information about the reference noise signal is available. The noise cancellation technique has many applications, e.g. speech processing, echo cancellation and enhancement, antenna array processing, biomedical signal and image processing and so on [1-4].

The standard methods of noise cancellation use only one primary signal [1]. However, in many applications, especially in biomedical signal processing, we are able to measure several primary signals. Often this possibility can help to improve the performance of noise cancellation procedure.

The standard approach is to make use of several noisy signals, by recording from

the same source. This consists in using a number of noise cancellation systems in parallel with one primary input to each system [1]. The estimated signal is obtained by selecting the best one in the sense of some criterion from the multichannel output signal. This approach can be realized by automatic selection of the best primary signal [5]. Another approach consists in averaging of outputs of all noise cancellation systems. In this paper we will show that the approach, based on a linear combination of outputs of all noise cancellation systems [6] with appropriate adjustment of weight parameters, is always not worse and typically better than an approach based on the automatic selection of the best primary signal. It is also easy to see that the approach based on linear combination of outputs of all noise cancellation systems includes the averaging method as a particular case. For this approach we use two adaptive filters for the primary signals and reference noise, respectively.

The paper is organized as follows. In Section 2 we formulate the problem. Section 3 explains the scheme for noise cancellation for multi-sensory signals and contains the derivation of our algorithms and comments about a convergence of the full procedure. Section 4 gives a proof of optimality of the algorithm for the weight parameter adjustment. Finally, Section 5 presents some simulation results to show validity and performance of the proposed method.

2. Problem formulation

The standard model for noise cancellation with one noise input and one signal input [1] is formulated as follows. We observe the source signal corrupted by additive noise:

$$d(k) \quad = \quad s(k) + \nu(k), \tag{1}$$

where $s(k)$ is an unknown primary source signal and $\nu(k)$ is an undesired interference or noise signal. It is assumed that only one noisy signal $d(k)$ and reference noise $\nu_R(k)$ are available. Moreover, it is assumed that between the reference noise $\nu_R(k)$ and the non-available interference signal $\nu(k)$, there exists an unknown linear dynamic relationship described by some filter $H(z^{-1})$. The task is to identify or design an appropriate transversal filter $W(z^{-1})$, which estimates filter $H(z^{-1})$ in such a way that we optimally estimate the interference $\nu(k)$ and subtract it from the signal $d(k)$.

However, in practice, we often have several observations of the same signal. Symbolically we can write

$$d_i(k) = G_i(z^{-1})s(k) + \nu_i(k), \quad (i = 1, 2, ..., l), \tag{2}$$

where $G_i(z^{-1})$ are the transfer functions of some unknown filters, $i = 1, 2, ..., l$ and l is the number of observation channels. In this paper we assume that $G_i(z^{-1}) \equiv g_i$, where g_i is an unknown scaling memoryless coefficient and $\nu_i(k)$ is the additive interference for each transmission channel. Without loss of generality, we also assume that we have only one reference noise signal $\nu_R(k)$. Otherwise, we should only

slightly modify the transversal filter $W(z^{-1})$ for processing not only one reference noise signal $\nu_R(k)$ but a number of reference noise signals $\nu_{R1}(k), \nu_{R2}(k), ..., \nu_{Rl}(k)$ [1]. Thus, in this paper we consider the following simplified model (see Fig. 1):

$$d_i(k) = g_i s(k) + \nu_i(k), \quad (i = 1, 2, ...l), \tag{3}$$

or in vector form

$$\mathbf{d}(k) = \mathbf{g}(k)s(k) + \overline{\nu}(k),$$

where $\mathbf{d}(k) = (d_1(k), d_2(k), ..., d_l(k))^\top$, $\mathbf{g}(k) = (g_1(k), g_2(k), ..., g_l(k))^\top$ and $\overline{\nu}(k) = (\nu_1(k), \nu_2(k), ..., \nu_l(k))^\top$. The model (3) is realistic for a number of application. The most important are biomedical applications to noise cancellation for brain signals such as electroencephalograms and magnetoencephalograms (EEG/MEG).

3. Models and algorithms for interference and noise cancellation for multi-sensory signals

In this section we consider a new scheme for noise cancellation for multi-sensory signals shown in Fig. 1 and associated learning algorithms for an adaptive filter $W(z^{-1})$ and a Linear Combiner (LC). The unknown part of this scheme is described by (3). We have observations of l noisy signals $d_i(k)$ and one reference noise signal $\nu_R(k)$ that is uncorrelated with the source signal $s(k)$, but is related to interferences $\nu_1(k), \nu_2(k), ..., \nu_l(k)$. This relation is represented symbolically by filters $H_1(z^{-1}), H_2(z^{-1}), ..., H_l(z^{-1})$. Filters $H_1, H_2, ..., H_l$ may not necessarily be linear, but nonlinear in general. However, previously proposed standard methods approximate these nonlinear filters by linear ones. We also assume for simplicity that $H_1(z^{-1}), H_2(z^{-1}), ..., H_l(z^{-1})$ are FIR filters with orders $n_1, n_2, ..., n_l$. The results can be generalized for the nonlinear case using neural network models [7,8]. According to our assumptions the transversal filter $W(z^{-1})$ should be an FIR filter with adaptive parameters.

Estimation of the original source signal $\hat{s}(k)$ is found by using an LC defined as follows:

$$\hat{s}(k) = \sum_{i=1}^{l} v_i d_i(k) - \hat{\nu}(k) = \mathbf{v}^\top \mathbf{d}(k) - \hat{\nu}(k) \tag{4}$$

with the natural constraint

$$\sum_{i=1}^{l} v_i = \mathbf{v}^\top \mathbf{e} = 1, \tag{5}$$

where $\hat{\nu}(k)$ is estimation of interference signal, $\mathbf{e} = (1, 1, ..., 1)^\top$ is $(l \times 1)$ vector of all ones and $\mathbf{v} = (v_1, v_2, ..., v_l)^\top$ is the $(l \times 1)$ weight parameter vector of the LC.

We should take into account that the one-dimensional corrupted signal $d(k)$ and estimation $\hat{s}(k)$ must be unbiased. Constraint (5) is nothing but an unbiasedness condition.
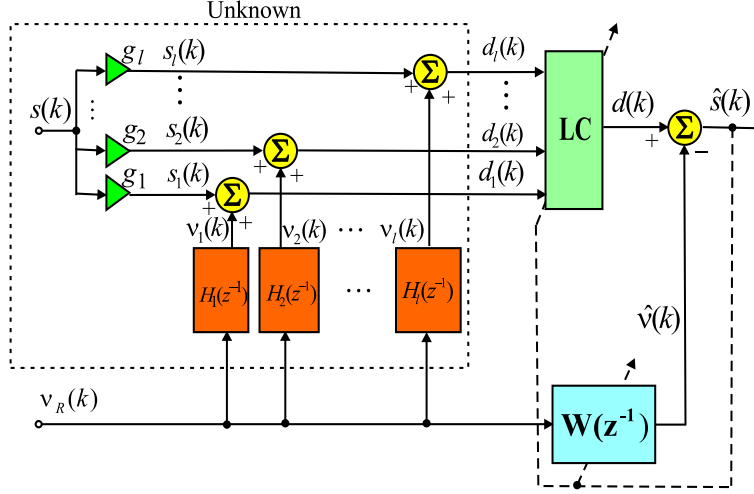
Figure 1: Proposed basic interference cancellation system for multi-sensory signals.

The aim is to estimate the coefficients of the filter $W(z^{-1}) = \sum_{p=1}^{m} w_p z^{-p}$ and parameters of the LC $\mathbf{v} = (v_1, v_2, ..., v_l)^\top$, which give an optimal estimation of the desired signal $s(k)$ at the output of the noise cancellation system shown in Fig. 1.

3.1. *Simultaneous learning of LC and transversal filter* $W(z^{-1})$

In the noise cancellation problem for learning of the LC we minimize the output signal power. Generally speaking, what we need is to maximize the Signal to Noise Ratio (SNR) of $\hat{s}(k)$. However, this objective is not achievable explicitly. Minimizing the output signal power under the assumption that $\nu(k)$ and $\hat{\nu}(k)$ are uncorrelated with $s(k)$ is equivalent to minimization of the Mean Square Error (MSE) $E\{e^2(k)\} = E\{(s(k) - \hat{s}(k))^2\}$. More precisely, minimization of $E\{\hat{s}^2(k)\} = E\{(s(k) + \nu(k) - \hat{\nu}(k))^2\} = E\{s^2(k)\} + E\{(\hat{\nu}(k) - \nu(k))^2\} = const + E\{(\hat{\nu}(k) - \nu(k))^2\}$ is equivalent to minimization of $E\{(s(k) - \hat{s}(k))^2\} = E\{(\hat{\nu}(k) - \nu(k))^2\}$. In this way we achieve maximization of SNR, at the same time.

Taking into account the equation for the LC (4) we can introduce the following cost function:

$$J = \frac{1}{2} \sum_{j=1}^{N} \hat{s}^2(j) = \frac{1}{2} \sum_{j=1}^{N} (\mathbf{v}^\top \mathbf{d}(j) - \hat{\nu}(j))^2, \tag{6}$$

where N is the number of observations. Minimization of the above cost function under constraint (5) leads to the Lagrangian

$$L = \frac{1}{2} \sum_{j=1}^{N} (\mathbf{v}^\top \mathbf{d}(j) - \hat{\nu}(j))^2 + \lambda(\mathbf{v}^\top \mathbf{e} - 1), \tag{7}$$

where λ is a non-negative Lagrange multiplier.

Optimizing the Lagrangian (7) we can find the saddle point that is the solution of the Kuhn-Tucker equations

$$
\begin{aligned}
\nabla_{\mathbf{v}} L &= \sum_{j=1}^{N} \mathbf{d}^{\top}(j) \mathbf{v} \mathbf{d}(j) - \sum_{j=1}^{N} \mathbf{d}(j) \hat{\nu}(j) + \lambda \mathbf{e} = 0, \\
\frac{\partial L}{\partial \lambda} &= \mathbf{v}^{\top} \mathbf{e} - 1 = 0.
\end{aligned} \tag{8}
$$

Let us consider new notations for inverse matrix $\mathbf{P}^{-1} = \left(\sum_{j=1}^{N} \mathbf{d}(j) \mathbf{d}^{\top}(j) \right)^{-1}$ and least square estimation $\tilde{\mathbf{v}} = \left(\sum_{j=1}^{N} \mathbf{d}(j) \mathbf{d}^{\top}(j) \right)^{-1} \times \left(\sum_{j=1}^{N} \mathbf{d}(j) \hat{\nu}(j) \right)$. The solution of the system (8) can be given by:

$$
\begin{aligned}
\lambda &= \frac{\mathbf{e}^{\top} \tilde{\mathbf{v}} - 1}{\mathbf{e}^{\top} \mathbf{P}^{-1} \mathbf{e}}, \\
\mathbf{v} &= \tilde{\mathbf{v}} - \mathbf{P}^{-1} \frac{\mathbf{e}^{\top} \tilde{\mathbf{v}} - 1}{\mathbf{e}^{\top} \mathbf{P}^{-1} \mathbf{e}} \mathbf{e}.
\end{aligned} \tag{9}
$$

Using the fact that covariance matrix $\mathbf{P}$ can be represented in the form $\mathbf{P}(k) = \sum_{j=1}^{k-1} \mathbf{d}(j) \mathbf{d}^{\top}(j) + \mathbf{d}(k) \mathbf{d}^{\top}(k) = \mathbf{P}(k-1) + \mathbf{d}(k) \mathbf{d}^{\top}(k)$, we can apply the Sherman-Morrison inversion lemma and finally write the following recursive fast procedure for tuning the vector $\mathbf{v}(k)$:

$$
\begin{aligned}
\mathbf{P}^{-1}(k) &= \mathbf{P}^{-1}(k-1) - \frac{\mathbf{P}^{-1}(k-1) \mathbf{d}(k) \mathbf{d}^{\top}(k) \mathbf{P}^{-1}(k-1)}{1 + \mathbf{d}^{\top}(k) \mathbf{P}^{-1}(k-1) \mathbf{d}(k)}, \\
\tilde{\mathbf{v}}(k) &= \tilde{\mathbf{v}}(k-1) + \frac{\mathbf{P}^{-1}(k-1) \mathbf{d}(k)}{1 + \mathbf{d}^{\top}(k) \mathbf{P}^{-1}(k-1) \mathbf{d}(k)} \hat{s}(k), \\
\mathbf{v}(k) &= \tilde{\mathbf{v}}(k) - \mathbf{P}^{-1}(k) \frac{\mathbf{e}^{\top} \tilde{\mathbf{v}}(k) - 1}{\mathbf{e}^{\top} \mathbf{P}^{-1}(k) \mathbf{e}} \mathbf{e}.
\end{aligned} \tag{10}
$$

In (10) least squares estimation $\tilde{\mathbf{v}}(k)$ is used as an intermediate for calculation of the parameter vector $\mathbf{v}(k)$ of the LC. This procedure is fully adaptive, very fast and does not require *a priori* information about signals except for above assumptions, which are natural and not very strict.

There is no explicit training set of input-output examples for learning of the transversal filter $W(z^{-1})$. Hence, the objective for learning of the filter $W(z^{-1})$ becomes the minimization of the output signal $\hat{s}(k) = e(k) = d(k) - \hat{\nu}(k)$ power. This is equivalent to minimizing the MSE between $\nu(k)$ or $\nu_1(k), \quad \nu_2(k), ..., \nu_l(k)$ and $\hat{\nu}(k)$ under assumption that $s(k)$ is uncorrelated with $\nu(k)$ or with all $\nu_1(k), \nu_2(k), ..., \nu_l(k)$. Thus, we can use $d(k) = s(k) + \nu(k)$ as the "desired output" and $\hat{\nu}(k)$ as the actual output for learning of an FIR filter $W(z^{-1})$ that can be formally described as

$$
\hat{\nu}(k) = \sum_{i=1}^{m} w_i(k) \nu_R(k-i) = \mathbf{w}^{\top}(k) \overline{\nu}_R(k), \tag{11}
$$

where $\mathbf{w}(k) = (w_1(k), w_2(k), ..., w_m(k))^\top$ is $(m \times 1)$ vector of unknown coefficients of the FIR transversal filter $W(z^{-1})$, $\overline{\nu}_R(k) = (\nu_R(k-1), \nu_R(k-2), ..., \nu_R(k-m))^\top$ and m is the order of the filter. The order m must be not less than the maximal order $\max_{1 \leq i \leq l} n_i$ of the original filters $H_1(z^{-1}), H_2(z^{-1}), ..., H_l(z^{-1})$ which are unknown. In practice, m can be chosen sufficiently large that desired accuracy will be achieved.

Minimizing the output signal power $E\{\hat{s}^2(k)\}$ with the weighting factor $\rho_k(n) = \alpha^{k-n}$, $n = 1, 2, ..., k$, we obtain the standard Recursive Least Squares (RLS) algorithm [9], which for our application can be written as

$$
\begin{aligned}
\mathbf{w}(k) &= \mathbf{w}(k-1) + \frac{\mathbf{Q}^{-1}(k-1)\overline{\nu}_R(k)}{\alpha + \overline{\nu}_R^\top(k)\mathbf{Q}^{-1}(k-1)\overline{\nu}_R(k)}\hat{s}(k), \\
\mathbf{Q}^{-1}(k) &= \frac{1}{\alpha}\left(\mathbf{Q}^{-1}(k-1) - \frac{\mathbf{Q}^{-1}(k-1)\overline{\nu}_R(k)\overline{\nu}_R^\top(k)\mathbf{Q}^{-1}(k-1)}{\alpha + \overline{\nu}_R^\top(k)\mathbf{Q}^{-1}(k-1)\overline{\nu}_R(k)}\right),
\end{aligned} \quad (12)
$$

where $\mathbf{Q}(k) = \sum_{k=1}^{N} \overline{\nu}_R(k)\overline{\nu}_R^\top(k)$ and α is known as the forgetting factor.

Let us note that other standard algorithms for learning of the transversal filter $W(z^{-1})$ can be used. Least Mean Square (LMS) algorithm [1,10,11] is the most popular. Stochastic Approximation (SA) algorithms [12,13] or generalization of SA [14] are also widely used. However, for us the speed of convergence is very import. This follows from a comparison between the algorithms for learning of the transversal filter (12) and the LC (10). We can easily see that both are based on an RLS procedure. The important question is the convergence of both algorithms running simultaneously. This is the reason why we implement both algorithms as RLS procedures with the same speed of convergence. However, the theoretical problem of convergence of both filters running simultaneously is not trivial. This problem is similar to the problem of convergence of two adaptive filters in tandem [15]. In this paper we do not discuss the details of the convergence analysis.

3.2. *Alternative approach with block of FIR filters and optimal LC*

Whereas the standard averaging approach use only one filter $W(z^{-1})$ to approximate the interference $\nu(k)$ with a reasonable accuracy [1], for the proposed approach possibly better results can be achieved by using a transversal filter $W_i(z^{-1})$ for each channel to estimate the interference signal $\hat{\nu}_i(k)$. Each filter $W_i(z^{-1})$ is learned, based on minimization of the output signal power $E\{\hat{s}_i^2(k)\}$ in corresponding channel. The model of the LC can now be defined as

$$
\hat{s}(k) = \sum_{i=1}^{l} v_i \hat{s}_i(k) = \mathbf{v}^\top \hat{\mathbf{s}}(k), \quad (13)
$$

where $\hat{\mathbf{s}}(k) = (\hat{s}_1(k), \hat{s}_2(k), ..., \hat{s}_l(k))^\top = (d_1(k) - \hat{\nu}_1(k), d_2(k) - \hat{\nu}_2(k), ..., d_l(k) - \hat{\nu}_l(k))^\top$ is a $(l \times 1)$ vector. The learning algorithm for the LC can be derived in the

same manner as algorithm (10). Here we use the following covariance matrix

$$\mathbf{R} = \sum_{j=1}^{N} \hat{\mathbf{s}}(j)\hat{\mathbf{s}}^\top(j) \tag{14}$$

instead of the matrix $\mathbf{P}$. The same property as for the matrix $\mathbf{P}$ is valid for the covariance matrix $\mathbf{R}$, i.e. $\mathbf{R}(k) = \sum_{j=1}^{k} \hat{\mathbf{s}}(j)\hat{\mathbf{s}}^\top(j) = \sum_{j=1}^{k-1} \hat{\mathbf{s}}(j)\hat{\mathbf{s}}^\top(j) + \hat{\mathbf{s}}(k)\hat{\mathbf{s}}^\top(k)$ $= \mathbf{R}(k-1) + \hat{\mathbf{s}}(k)\hat{\mathbf{s}}^\top(k)$. Hence, the inversion lemma can be applied for $\mathbf{R}$ as well as for $\mathbf{P}$. The derivation will be given in the next section. Here we only write the final result for the recursive version of the procedure based on covariance matrix $\mathbf{R}$

$$\begin{aligned}
\mathbf{R}^{-1}(k) &= \mathbf{R}^{-1}(k-1) - \frac{\mathbf{R}^{-1}(k-1)\hat{\mathbf{s}}(k)\hat{\mathbf{s}}^\top(k)\mathbf{R}^{-1}(k-1)}{1 + \hat{\mathbf{s}}^\top(k)\mathbf{R}^{-1}(k-1)\hat{\mathbf{s}}(k)}, \\
\mathbf{v}(k) &= \frac{\mathbf{R}^{-1}(k)\mathbf{e}}{\mathbf{e}^\top\mathbf{R}^{-1}(k)\mathbf{e}}.
\end{aligned} \tag{15}$$

Let us note that computational complexity of the full procedure, including adjustment of each transversal filter, is higher than the complexity of the scheme shown in Fig. 1.

4. Optimality of algorithms (10) and (15)

The optimality of the proposed algorithms can be formulated in the form of the following theorem.

Theorem: The minimal value of the cost function (6) $E\{\hat{s}^2(k)\}$, for output signal of the noise cancellation system shown in Fig. 1, that can be achieved using the algorithm (10) for learning of the LC will be always not larger than one achieved for the best channel.

Remark: For the same transversal filter $W(z^{-1})$, the channel in which the best result of noise cancellation is achieved (best channel) corresponds to the channel with the best primary signal [5]. Hence, the approach based on automatic selection of the best primary signal is identical to finding the single channel $d_{i*}(k)$ for which we obtain $\min_{1 \le i \le l} E\{\hat{s}_i^2(k)\}$ after noise cancellation. More rigorously, the statement of the theorem is: **for a given transversal filter $W(z^{-1})$ the inequality** $E\{\hat{s}^2(k)\} \le \min_{1 \le i \le l} E\{\hat{s}_i^2(k)\}$ **is always valid**.

Proof: Let us rewrite the procedure (10) in equivalent form. LC is defined now by the equation (13) which is equivalent to equation (4) because $\sum_{i=1}^{l} v_i\hat{s}_i(k) = \sum_{i=1}^{l} v_i \times (d_i(k) - \hat{\nu}(k)) = \sum_{i=1}^{l} v_i d_i(k) - \sum_{i=1}^{l} v_i\hat{\nu}(k)$, where the second term $\sum_{i=1}^{l} v_i\hat{\nu}(k) = \hat{\nu}(k)$, $\hat{\nu}(k)$ does not depend on i and $\sum_{i=1}^{l} v_i = 1$ by definition (5). Then the Lagrangian (7) can be rewritten as

$$L = \sum_{j=1}^{N} \mathbf{v}^\top\hat{\mathbf{s}}(j)\hat{\mathbf{s}}^\top(j)\mathbf{v} + \lambda(\mathbf{v}^\top\mathbf{e} - 1) = \mathbf{v}^\top\mathbf{R}\mathbf{v} + \lambda(\mathbf{v}^\top\mathbf{e} - 1). \tag{16}$$

It is easy to find the saddle point of the Lagrangian (16):

$$L^* = (\mathbf{e}^\top\mathbf{R}^{-1}\mathbf{e})^{-1} \tag{17}$$

and the solution for weight parameter vector of the LC

$$\mathbf{v} = \frac{\mathbf{R}^{-1}\mathbf{e}}{\mathbf{e}^\top \mathbf{R}^{-1}\mathbf{e}}. \tag{18}$$

Finally, let us note that (18) is equivalent to the second equation of (9). The diagonal elements of the covariance matrix $\mathbf{R}$ are $E\{\hat{s}_i^2(k)\}, \quad i = 1, 2, ..., l$. Thus, the diagonal elements of the covariance matrix $\mathbf{R}$ are simply the variances of the corresponding local estimations of the unknown signal $s(k)$.

Now, we can begin the proof of the theorem. Let us consider two arbitrary $(l \times 1)$ vectors: $\mathbf{x}$ and $\mathbf{z}$ and write the obvious relationship:

$$(\mathbf{x}^\top \mathbf{z})^2 = \left(\mathbf{x}^\top \mathbf{R}^{\frac{1}{2}}\mathbf{R}^{-\frac{1}{2}}\mathbf{z}\right)^2 = \left(\left(\mathbf{R}^{\frac{1}{2}}\mathbf{x}\right)^\top \left(\mathbf{R}^{-\frac{1}{2}}\mathbf{z}\right)\right)^2. \tag{19}$$

Using the Cauchy-Schwarz inequality we can write

$$\left(\left(\mathbf{R}^{\frac{1}{2}}(k)\mathbf{x}\right)^\top \left(\mathbf{R}^{-\frac{1}{2}}(k)\mathbf{z}\right)\right)^2 \leq \|\mathbf{R}^{\frac{1}{2}}(k)\mathbf{x}\|^2 \|\mathbf{R}^{-\frac{1}{2}}(k)\mathbf{z}\|^2. \tag{20}$$

For the right-hand part of inequality (20) the following equation

$$\|\mathbf{R}^{\frac{1}{2}}(k)\mathbf{x}\|^2 \|\mathbf{R}^{-\frac{1}{2}}(k)\mathbf{z}\|^2 = (\mathbf{x}^\top \mathbf{R}(k)\mathbf{x})(\mathbf{z}^\top \mathbf{R}^{-1}(k)\mathbf{z}) \tag{21}$$

is valid. Hence, using (19) and (21) and substituting them into (20) we obtain

$$(\mathbf{x}^\top \mathbf{z})^2 \leq (\mathbf{x}^\top \mathbf{R}(k)\mathbf{x})(\mathbf{z}^\top \mathbf{R}^{-1}(k)\mathbf{z}). \tag{22}$$

Letting $\mathbf{e}_i$ denote the $l \times 1$ vector of all zeros except in the ith place where there is a 1, and taking $\mathbf{x} = \mathbf{e}$ and $\mathbf{z} = \mathbf{e}_i$, we can rewrite (22) in the form

$$(\mathbf{e}^\top \mathbf{e}_i)^2 \leq (\mathbf{e}_i^\top \mathbf{R}(k)\mathbf{e}_i)(\mathbf{e}^\top \mathbf{R}^{-1}(k)\mathbf{e}). \tag{23}$$

Here $(\mathbf{e}^\top \mathbf{e}_i)^2 = 1$ and $\mathbf{e}_i^\top \mathbf{R}(k)\mathbf{e}_i = \mathbf{R}_{ii}(k)$ is i-th diagonal element of the covariance matrix $\mathbf{R}(k)$ that is nothing but the variance of $\hat{s}_i(k)$. Thus, we can see that the following inequality

$$1 \leq \mathbf{R}_{ii}(k)(\mathbf{e}^\top \mathbf{R}^{-1}(k)\mathbf{e}) \tag{24}$$

is always valid. Explicitly we have

$$E\{\hat{s}_i^2(k)\} \geq (\mathbf{e}^\top \mathbf{R}^{-1}(k)\mathbf{e})^{-1} = L^* = E\{\hat{s}^2(k)\}. \tag{25}$$

The inequality (25) is valid for all $i = 1, 2, ..., l$ because 1 can be written in any place of vector $\mathbf{e}_i$. Hence, (25) is also true for $\min_{1 \leq i \leq l} E\{\hat{s}_i^2(k)\}$ corresponded to the best channel and the best primary signal.

Consequence: Using algorithm (10) or (15) for learning of the LC (see Fig. 1) we will always achieve better than or equivalent performance of noise cancellation upon comparison to the method based on automatic selection of the best primary signal. This follows from the theorem and the fact that minimization of the output signal power is equivalent to the minimization of the MSE $E\{e^2(k)\} = E\{(s(k) - \hat{s}(k))^2\}$, which simultaneously leads to the maximization of the SNR.
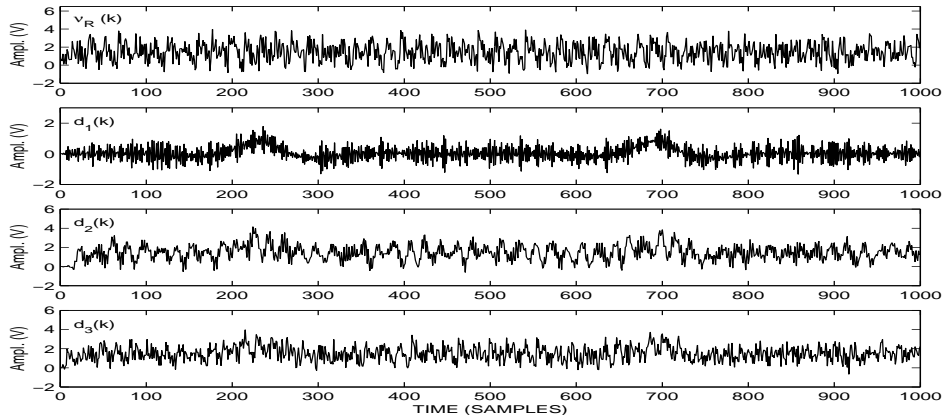
Figure 2: The reference noise signal $\nu_R(k)$ and three observed signals $d_1(k)$, $d_2(k)$, $d_3(k)$ contaminated by interference signals.

5. Simulations

The following simulations have been done in order to evaluate the performance of the proposed method for noise cancellation for multi-sensory signals. We also make a comparison of the proposed method with other methods. These methods are: a) the method based on synchronous signal averaging of the primary signals, which is widely used in practice and will be considered here as the averaging method; b) the method based on selection of the best primary signal [5], which will be considered here as the Best Primary Signal Selection (BPSS) method.

The reference noise signal $\nu_R(k)$ and three noisy observed signals $d_1(k)$, $d_2(k)$, $d_3(k)$ are shown in Fig. 2. The signal to estimate is a simulated Evoked Potential (EP) embedded in interference signals which are different for each measurement channel. Reference noise is modelled as the sum of the signal generated from uniform distribution in the interval $[1, 3]$ and the sawtooth signal with amplitude 1 and frequency 0.145573 rad/sec. Interference signals are generated as follows. For the first channel a highpass filter of order 38 with the cutoff frequency 0.7 and a Kaiser window is used. Interference signal in the second channel is generated using a multiband filter of order 50 with the cutoff frequency vector $[0.2, 0.5, 0.8]$ and a Hamming window. The gain in the first band is equal to 1. For the third channel a bandstop filter of order 26 with the lower cutoff frequency 0.5 and upper cutoff frequencies 0.6, Chebyshev window and stopband attenuation of 10 dB is used.

For each channel we use individual transversal RLS adaptive filters. Characteristics of these filters are the following. For the transversal filters in the first and second channels the length of the FIR filters is equal to 32, the forgetting factor is equal to 1.0, the initial value of filter taps is 0 and the initial input variance estimate is 0.1. Similarly, for the third channel the transversal FIR filter length is equal to 27, the forgetting factor is equal to 0.999 and the initial conditions are the same
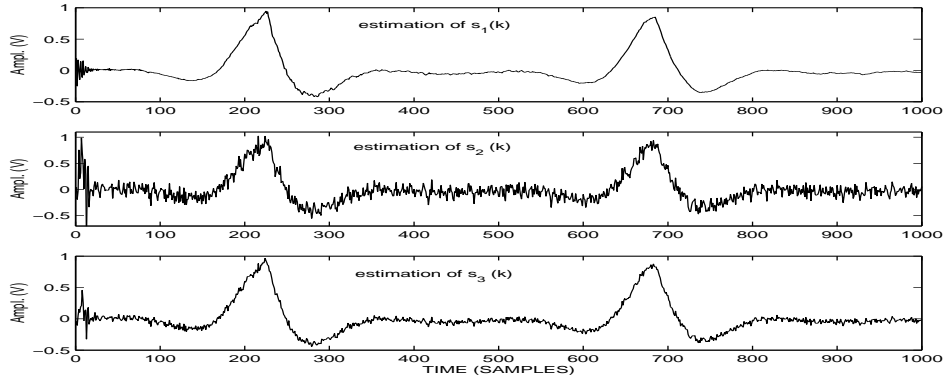
Figure 3: The results of noise cancellation for each individual channel.

as for the filters for the first and second channels. The algorithm (15) is used for learning of the LC. Setting the initial values of the weight parameters as $v_i = \frac{1}{3}$, $i = 1, 2, 3$ corresponds to the averaging method.

The results of noise cancellation for each individual channel are shown in Fig. 3. It is easy to see that the best result is achieved for the first channel, which corresponds at the same time to the best primary signal. However, the results of noise cancellation for the method based on selection of the best primary signal depend also on a good choice of order of the transversal filter. Moreover, this method does not work for time-varying systems, when the best primary channel can change with a time.

Fig. 4 shows the estimated signals using the three methods and Fig. 5 shows the Normalized Mean Square Errors (NMSEs), $NMSE = \frac{E\{(\hat{s}(k)-s(k))^2\}}{E\{\nu^2(k)\}}$, for each method depending on the time. The worst result was obtained by the averaging method. This is due to the few number of averaged signals. For practical applications, especially in biomedical signal processing, where the number of channels is big enough, averaging methods give better results and are the most popular. However, we can see that even for a few number of channels the best result is obtained by our proposed method. This is due to optimal adjustment of weight parameters of the LC.

It is important to note that the proposed method will also work for time-varying systems due to the recursive estimation of weight parameters of the LC. The output of the noise cancellation system tends to estimate signal $\hat{s}(t)$ but not to zero. Hence, the algorithm does not tend to a stable solution. Changes lead to variations of both transversal filters coefficients and weight parameters of the LC from their optimal values. However, according to above theorem, weight parameters of the LC tend again to optimal values after changes. Let us also note that often the reference signal $\nu_R(k)$ is not available, however for a number of real problems it can be approximated as standard Gaussian noise.
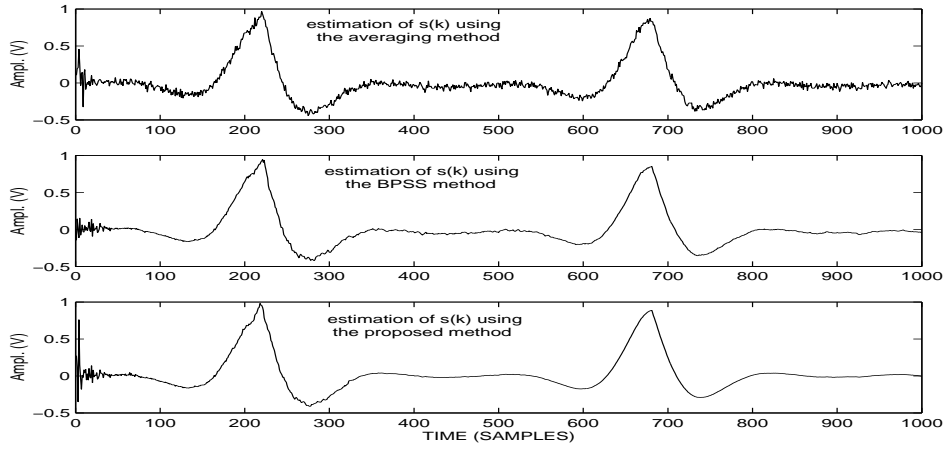
Figure 4: The results of noise cancellation for the averaging method, for the method based on selection of the best primary signal and for the proposed method.
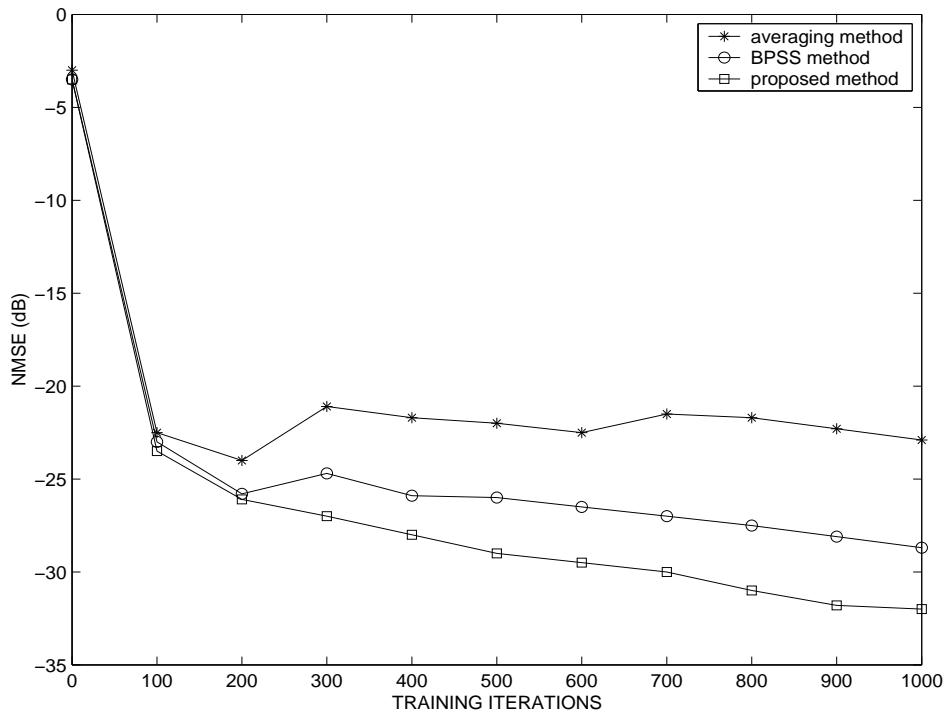

Figure 5: NMSE for the averaging method, for the method based on selection of the best primary signal and for the proposed method.

6. Conclusions

In this topical review, simple and fast algorithms for noise cancellation for multi-sensory signals have been proposed. The theorem proved in Section 4 allows us to assert that using the proposed noise cancellation system for a multi-sensory signal we will achieve better then or equivalent performance of noise cancellation in comparison to the method based on automatic selection of the best primary signal and the averaging method. Simulation results confirm theoretical consequences and demonstrate the effectiveness of the proposed algorithm.

Acknowledgement

References

[1] B. Widrow and E. Walach, *Adaptive Inverse Control*, Prentice-Hall, Inc., S.S.Series, N.J. (1996).

[2] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 3rd Edition (1996).

[3] M. Feder, A.V. Oppenheim, E. Weistein, *Maximum likelihood noise cancellation using the EM algorithm*, IEEE Trans. on Acoustics, Speech and Signal Processing, **37** (1989) 204-216.

[4] S.A. Billings and C.F. Fung, *Recurrent radial basis function network for adaptive noise cancellation*, Neural Networks, **8** (1995) 273-290.

[5] S.M. Kuo and J. Chen, *Multiple-microphone acoustic echo cancellation system with the partial adaptive process*, Digital Signal Processing: A Review Journal, **3** (1993) 54-63.

[6] E. Bataillou and H. Rix, *A new method for noise cancellation,* in book *Signal Processing VII. Theories and Applications*, eds. M. Holt, C. Cowan, P. Grant, W.Sandham, European Association for Signal Processing (1994) 1046-1049.

[7] I. Cha, and S.A. Kassam, *Interference cancellation using radial basis function networks*, Signal Processing, **47** (1995) 247-268.

[8] A. Cichocki, S.A. Vorobyov, T. Rutkowski, *Nonlinear interference cancellation using neural networks*, Proc. Int. Symposium on Nonlinear Theory and Its Applications, Hawaii, (Dec. 1999) 875-878.

[9] L. Ljung, *System Identification Theory for the User*, Prentice-Hall, Englewood Cliffs, New York (1986).

[10] S. Kaczmarz, *Approximate solution of systems of linear equation*, International Journal of Control, **57** (1993) 1269-1271.

[11] N.S. Raibman and V.M. Chadeev, *Design of the models for industrial processes*, Energiya, Moscow (1975). *(in Russian)*

[12] H.J. Kushner and G.G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer, New York (1997).

[13] G.C. Goodwin, P.J.Ramadge, P.E. Caines, *A globally convergent adaptive predictor*, Automatica, **17** (1981) 135-140.

[14] S.A. Vorobyov and Ye.V. Bodyanskiy, *On one non-parametric algorithm for smoothing parameter control in adaptive filtration*, Engineering Simulation, **16** (1999) 341-350.

[15] K.C. Ho, *A study of two adaptive filters in tandem*, IEEE Transactions on Signal Processing, **48** (2000) 1626-1636.