

# TRADITIONAL AVERAGING, WEIGHTED AVERAGING, AND ERPSUB FOR ERP DENOISING IN EEG DATA

## *A Comparison of the Convergence Properties*

Andriy Ivannikov, Tommi Kärkkäinen, Tapani Ristaniemi  
*Department of Mathematical Information Technology, University of Jyväskylä*  
*P.O. Box 35 (Agora), FIN-40014, Jyväskylä, Finland*

Heikki Lyytinen  
*Department of Psychology, University of Jyväskylä*  
*P.O. Box 35 (Agora), FIN-40014, Jyväskylä, Finland*

**Keywords:** Convergence, EEG, ERP, Denoising, Weighted Averaging, SNR, Source Separation.

**Abstract:** In this article we compare the convergence rates at increase of the number of processed trials of the three methods applied nowadays in electroencephalography research to denoising of event-related potentials: traditional averaging, weighted averaging, and ERPSUB. We derive the weighted averaging procedure by maximizing signal-to-noise ratio in the averaged subject responses and show, thereby, that maximizing signal-to-noise ratio criterion is equivalent to minimizing the originally proposed mean-square error criterion in the sense of the weighted averaging problem solving. Moreover, in order to characterize fully the performance of the selected methods, we compare also noise reduction rates in estimates of event-related potentials provided by methods, while the number of processed trials increases.

## 1 INTRODUCTION

Reliable characterization of event-related potentials (ERPs) is a central task in electroencephalography (EEG) data processing. ERP is a concept used in EEG research to denote brain electromagnetic potentials occurring as responses to the external or mental events, whose quantitative understanding underlies many neuropsychological studies and clinical diagnosis (Huttunen et al., 2007; Luu et al., 2004; Makeig et al., 1999; Näätänen, 1992). However, the signal-to-noise ratio (SNR) is very low in a single measurement (trial) of the brain response following the stimulation event, which makes it impossible to identify ERP characteristics, such as amplitude and latency, reliably. In order to increase SNR and, hence, estimate reliably ERP characteristics, many trials of equal length and synchronized to the same event are measured from different locations on the scalp (channels) and averaged channel-wise (see Sect. 2). Averages of many trials for every channel are assumed to have high SNR and important ERP characteristics can be identified then from the averages with the accuracy depending on the number of trials used for averaging.

Moreover, besides improving the reliability of the estimates of ERP characteristics, it is also important to shorten the experiment time, because subjects under consideration suffer from the long time lasting experiments. They get tired, lose attention and can not adequately perform the experimental tasks anymore. As a consequence, data become less informative from the experimental design point of view. Furthermore, for some groups of probationers (infants or patients) long experiments may be too demanding.

Basically, we need less trials to shorten the time of the experiment. Hence, our attention is focused on methods, which extract useful information from EEG data more effectively than the conventional averaging does. This allows obtaining the desired accuracy of ERP characteristics using fewer trials and, hence, shorter experiment. We consider two methods that were developed to increase SNR in the subject averages as compared to the conventional averaging procedure: weighted averaging (Hoke et al., 1984) and ERPSUB (Ivannikov et al., 2007).

An important assumption underlying the averaging in electroencephalography research is the ergodicity of the noise. However, we should be realistic and

understand that this assumption is violated to some extent in practical applications. This leads us to a situation, when the variance of the noise is different across trials. It then turns out that SNR in the averaged responses can be boosted by weighting the trials inversely to the variance of the noise they contain. The formal derivation of this result was originally obtained in (Hoke et al., 1984) by minimizing the mean-square error criterion. In (Davila and Mobin, 1992) a similar technique has also been derived by maximizing SNR in the average using Rayleigh quotient and solving the generalized eigenvalue problem. Later, in (Łeński, 2002) robust version of weighted averaging was proposed and further developed into computationally more effective algorithm in (Łeński and Gacek, 2004). In this paper we obtain essentially same result as in (Hoke et al., 1984) by maximizing SNR criterion, but using different derivation procedure than that used in (Davila and Mobin, 1992) and show, thereby, that SNR criterion is equivalent to the mean-square error criterion in the sense of the weighted averaging problem solving.

ERPSUB method utilizes the problem specific assumptions for ERP/noise linear subspaces separation in multichannel EEG data and results in more effective denoising of ERPs comparing to the conventional averaging (Ivannikov et al., 2007). Method automatically solves the component classification problem for a priori known dimensionality of ERP subspace. Moreover, it contains also means for estimating the dimensionality of ERP subspace, if prior knowledge is absent, with the accuracy depending on the closeness of the data properties to the values provided by the ideal assumptions.

Since we are interested in decreasing the experimental time (minimizing number of trials necessary for reliable ERP identification), in this paper we concentrate on and compare the convergence rates of ERP estimates provided by selected methods (traditional averaging approach, weighted averaging, and ERPSUB), while the number of processed trials increases. Moreover, in order to give a comprehensive evaluation of the methods' performance, we also compare the noise reduction rates in ERP estimates for the same conditions.

The structure of the work is as follows. First, in Sect. 2, we describe the experimental data and formulate the research area. Then, in Sect. 3, the methods are discussed. Section 4 represents the experimental results. In Sect. 5, conclusions are drawn.

## 2 PRELIMINARIES

In this article we used EEG data that were introduced and studied in (Huttunen et al., 2007) and (Kalyakin et al., 2007). The same data were utilized also for the purposes of testing in (Ivannikov et al., 2007). The data collection experimental design was targeted to elicit mismatch negativity (MMN) component of auditory ERP. In fact, MMN has turned out to be especially useful for the investigation of the brain basis of human auditory cognition (Näätänen, 1992).

In the data collection experiments, the experimental paradigm proposed in (Pihko et al., 1995) was used. It is based on a sequence of standard stimuli consisting of continuously (uninterruptedly) alternated sounds of 600 Hz and 800 Hz, each lasting 100 ms. Two types of deviant stimuli are randomly presented in this sequence with the frequency of 600 Hz and duration of 30 ms or 50 ms. The measured trials contain 300 ms of recordings before the start of the deviant tone and 350 ms after the start of the deviant tone. Measurements were collected with the sampling rate 200 Hz, thus, giving 130 time points for each trial. There were 102 participants (or subjects) involved in the data collection experiment. Measurements were recorded using 12-electrodes scheme resulting in 350 trials collected for each of 102 subjects, each of the two deviants and each of the nine channels of EEG data (i.e., C3, C4, Cz, F3, F4, Fz, Pz, M1, M2) and the two channels of electrooculography (EOG) data (i.e., ER, EL). An additional nose electrode was used as a reference point.

We assume that each recorded trial  $x_i^k(t)$  contains both the weighted sum of the time-locked brain responses  $s^k(t)$  assumed to be deterministic through all trials and the weighted sum of the noise sources  $n_i^k(t)$ , such as spontaneous EEG and artifacts (Vigário, 1997; Jung et al., 2000). Noises  $n_i^k(t)$  are assumed to be uncorrelated with each other and with  $s^k(t)$ . Then, without loss of generality we can assume that  $x_i^k(t)$ ,  $s^k(t)$ , and  $n_i^k(t)$  are zero mean variables, since data always can be centered. Hence, the simplest additive model to describe the phenomenon reads as

$$x_i^k(t) = s^k(t) + n_i^k(t), \quad (1)$$

where  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ , and  $k = 1, \dots, K$ . Here  $N$  denotes the number of measured trials,  $T$  is the number of time points per trial, and  $K$  denotes the number of measured channels. The conventional averaging operation is performed for each channel separately and is described by formula:

$$\bar{x}_N^k(t) = \frac{1}{N} \cdot \sum_{i=1}^N x_i^k(t) = s^k(t) + \bar{n}_N^k(t), \quad (2)$$

where  $s^k(t)$  is the time-locked ERP constituent (signal of interest) and  $\bar{n}_N^k(t)$  is the noise constituent in the average. The resulting average in (2) is assumed to have higher SNR than the single trial does that is confirmed by practical experience and theoretical computations (Näätänen, 1992; Furst and Blau, 1991).

### 3 METHOD DESCRIPTION

#### 3.1 Weighted Averaging

The variance of the sum of  $N$  stochastic variables can be expressed through the formula

$$\sigma^2 \left( \sum_{i=1}^N x_i(t) \right) = \sum_{i=1}^N \sigma^2(x_i(t)) + 2 \sum_{i < j} Cov_{ij}, \quad (3)$$

where  $Cov_{ij} = E[x_i(t)x_j(t)]$  denotes the covariance between the two zero mean stochastic variables or trials  $x_i(t)$  and  $x_j(t)$ , and  $\sigma$  denotes the standard deviation. To simplify the following discussion we omit the channel index  $k$  throughout the paper assuming that all channels are treated in the similar way. Therefore, for the weighted sum/average of trials  $\sum_{i=1}^N a_i x_i(t)$  and taking into account that the covariance of the two perfectly linearly correlated signals equals to the product of their standard deviations, we have

$$\sigma^2 \left( \sum_{i=1}^N a_i x_i(t) \right) = \sum_{i=1}^N a_i^2 \sigma_s^2 + \sum_{i=1}^N a_i^2 \sigma_{n_i}^2 + 2 \sigma_s^2 \sum_{i=1}^{N-1} a_i \sum_{j=i+1}^N a_j, \quad (4)$$

where  $\sigma_s^2$  denotes the variance of the signal and  $\sigma_{n_i}^2$  is the variance of the noise in  $i$ -th trial. Then the portions of the total variance  $\sigma_s^2$  and  $\sigma_n^2$  that are contributed by the time-locked signals and noise sources, correspondingly, to the weighted sum (normal average in case  $a_i = \frac{1}{N}, \forall i = 1, \dots, N$ ) of  $N$  trials read as

$$\sigma_s^2 = \sigma_s^2 \left( \sum_{i=1}^N a_i \right)^2, \quad (5)$$

$$\sigma_n^2 = \sum_{i=1}^N a_i^2 \sigma_{n_i}^2. \quad (6)$$

We define SNR in the weighted sum of  $N$  trials as the variance of ERP constituent in this sum divided by the variance of the noise constituent:

$$SNR_N = \frac{\sigma_s^2}{\sigma_n^2} \quad (7)$$

and try to maximize its value in order to determine the optimal values of  $a_i$ 's. For this purpose, taking the partial derivatives of  $\sigma_s^2$  and  $\sigma_n^2$  with respect to  $a_i$ , we have

$$\frac{\partial \sigma_s^2}{\partial a_i} = 2 \sigma_s^2 \sum_{j=1}^N a_j, \quad \forall 1 \leq i \leq N, \quad (8)$$

$$\frac{\partial \sigma_n^2}{\partial a_i} = 2 a_i \sigma_{n_i}^2, \quad \forall 1 \leq i \leq N. \quad (9)$$

Therefore, the partial derivative of SNR with respect to  $a_i$  is given by

$$\frac{\partial SNR_N}{\partial a_i} = \frac{\frac{\partial \sigma_s^2}{\partial a_i} \sigma_n^2 - \sigma_s^2 \frac{\partial \sigma_n^2}{\partial a_i}}{(\sigma_n^2)^2}, \quad \forall 1 \leq i \leq N. \quad (10)$$

Saddle points in the  $a_i$ 's coordinate space can be found by equating the numerator of equation (10) to zero assuming  $\sigma_n^2 \neq 0$ . Therefore, the problem can be expressed through a system of equations

$$\left( \sigma_s^2 \sum_{j=1}^N a_j \right) \sigma_n^2 - \sigma_s^2 (a_i \sigma_{n_i}^2) = 0, \quad \forall 1 \leq i \leq N. \quad (11)$$

Subtracting any two equations in this system, we obtain

$$a_i \sigma_{n_i}^2 = a_j \sigma_{n_j}^2, \quad \forall 1 \leq i, j \leq N. \quad (12)$$

Plugging  $a_j = \frac{\sigma_{n_i}^2}{\sigma_{n_j}^2} a_i$  back to the system of equations (11), we get a system of identical equations after some manipulations. Moreover, since the values of weighting coefficients  $a_i$ 's were not fixed in this operation, they can be arbitrary within the constraint (12). This means, in turn, that

$$a_i \sigma_{n_i}^2 = a_j \sigma_{n_j}^2 = C, \quad \forall 1 \leq i, j \leq N, \quad (13)$$

where  $C$  can be any constant. Hence, the solution has a form

$$a_i = \frac{C}{\sigma_{n_i}^2}, \quad \forall 1 \leq i \leq N. \quad (14)$$

It is easy to check that this extremum point is the maximum by substituting  $a_i = \frac{C}{\sigma_{n_i}^2} \pm \Delta, \forall 1 \leq i \leq N$  in (11), where  $\Delta > 0$  is an infinitely small shift.

Assuming SNR in a single trial is very low (this follows from the magnitude level of the time-locked signal  $\approx 3-5\mu\text{V}$  compared to the magnitude level of the trial itself  $\approx 50-100\mu\text{V}$ ), we can disregard the variance contributed by the time-locked signal to the trial and approximate

$$\sigma_{n_i}^2 \approx \sigma_{x_i}^2. \quad (15)$$

Thus, we can approximately compute the coefficients  $a_i$ 's by arbitrarily fixing  $C$  constant first.

Note that in (Hoke et al., 1984) the minimization of the mean-square error leads to a single unique solution, whereas in our case the maximization of SNR yields an infinite set of solutions due to the arbitrary choice of  $C$  in (14). This result can be explained by the obvious reasoning that only the ratio between  $a_i$ 's is emphasized by SNR criterion (the weighted sum can be multiplied by any number keeping SNR on a same level), whereas the solution based on minimizing mean-square error criterion is associated with the original level of ERP signal and with the highest SNR as well. Hence, in order to correct the level of ERP signal to original in the weighted average with weighting coefficients fixed as in (14), where  $C$  is arbitrary, we need to multiply  $\sum_{i=1}^N a_i x_i(t)$  by a correction factor  $\alpha$  that eliminates uncertainty introduced by arbitrariness of  $C$ . Apparently  $\alpha$  depends on  $C$  and plays role of a constraint imposed on  $C$  and  $a_i$ 's that specifies only single set of  $a_i$ 's preserving the original level of ERP signal in the weighted average. From (5)  $\alpha$  is obviously expressed through the formula

$$\alpha = \frac{1}{\sqrt{\sigma_s^2 / \sigma_s^2}} = \frac{1}{\sum_{i=1}^N a_i}. \quad (16)$$

After embedding the correction factor  $\alpha$  into (14) the final solution for the weighting coefficients becomes

$$a_i = \sigma_{n_i}^{-2} / \sum_{j=1}^N \sigma_{n_j}^{-2}, \forall 1 \leq i \leq N \quad (17)$$

that coincides with the results from (Hoke et al., 1984). These values of the weighting coefficients are unique in the sense that they are connected to the original level of ERP signal and, thus, do not require multiplication by the correction factor  $\alpha$ , which equals to 1 in this case.

### 3.2 ERPSUB

In the contemporary research EEG data is often considered in the scope of the linear instantaneous noiseless mixing model, which is also assumed in this paper:

$$X_i = A \cdot Y_i, \forall i = 1, \dots, N, \quad (18)$$

where  $X_i$  is a matrix of size  $K \times T$ , which contains measurements from  $K$  channels and one trial of length  $T$  time points,  $Y_i$  is a matrix of size  $K \times T$ , which contains the realizations of  $K$  sources of length  $T$  time points, and  $A$  stands for the mixing matrix. It is assumed that every row in  $X_i$  has zero mean for all  $i$ , i.e. the data are centered. In addition we assume that the mixing matrix  $A$  does not change in time. Practically it means that for one subject during one experiment

with the static conditions matrix  $A$  stays the same for all trials within the experiment. Therefore, we are allowed to form a data matrix by concatenating matrices  $X_i$  channel-wise:

$$X = A \cdot Y, \quad (19)$$

where  $X = [X_1 \ X_2 \ \dots \ X_N]$  is the matrix of concatenated measurements of size  $K \times TN$  and  $Y = [Y_1 \ Y_2 \ \dots \ Y_N]$  is the matrix of concatenated realizations of the sources of the same size. Matrix equivalent of (2) can now be written as

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i = A \frac{1}{N} \sum_{i=1}^N Y_i = A \bar{Y}. \quad (20)$$

Furthermore, in the framework of the model (18) it is assumed that all  $K$  measurements in every multi-dimensional trial  $X_i$  are linearly independent and the number of sources does not exceed the number of channels. These assumptions are introduced to ensure that measurements form the basis for the linear space of the same dimension as sources do. This, in turn, guarantees the existence of the pure signal and noise subspaces in theory. Both assumptions are practically addressed by reasonable selection of  $K$  and  $T$  parameters. Moreover, we assume that subspaces of ERP signals and noise are statistically independent. The imposed assumptions, except the one concerning the linear independence of measurements, are rather strict and can not be completely justified in practical applications. However, they are necessary on the stage of the method development. In real situations one is instructed to reinterpret the results of the method according to the types and extent of the assumptions' violations.

The main idea of ERPSUB is to use the relevant information stored in data along all time, trial, and channel dimensions, while separating ERP/noise subspaces. In contrary, most of the Independent Component Analysis (ICA) methods also applied in EEG data processing to ERP/noise sources separation exploit the information kept along the time and channel dimensions only, whereas the trial dimension is ignored (Hyvärinen et al., 2001; Jung et al., 2000; Vigário, 1997). Traditional averaging is one-channel technique, and it exploits the information hidden in trial dimension only for ERP denoising. Weighted averaging is also one-channel procedure, but it utilizes the information taken from trial and time dimensions for the purposes of ERP denoising. ERPSUB exploits the fact that after the averaging the variance of data should decrease along the directions in the noise subspace, while the variance along the signal directions should stay on the original level in ideal conditions. This means that after whitening, which should make

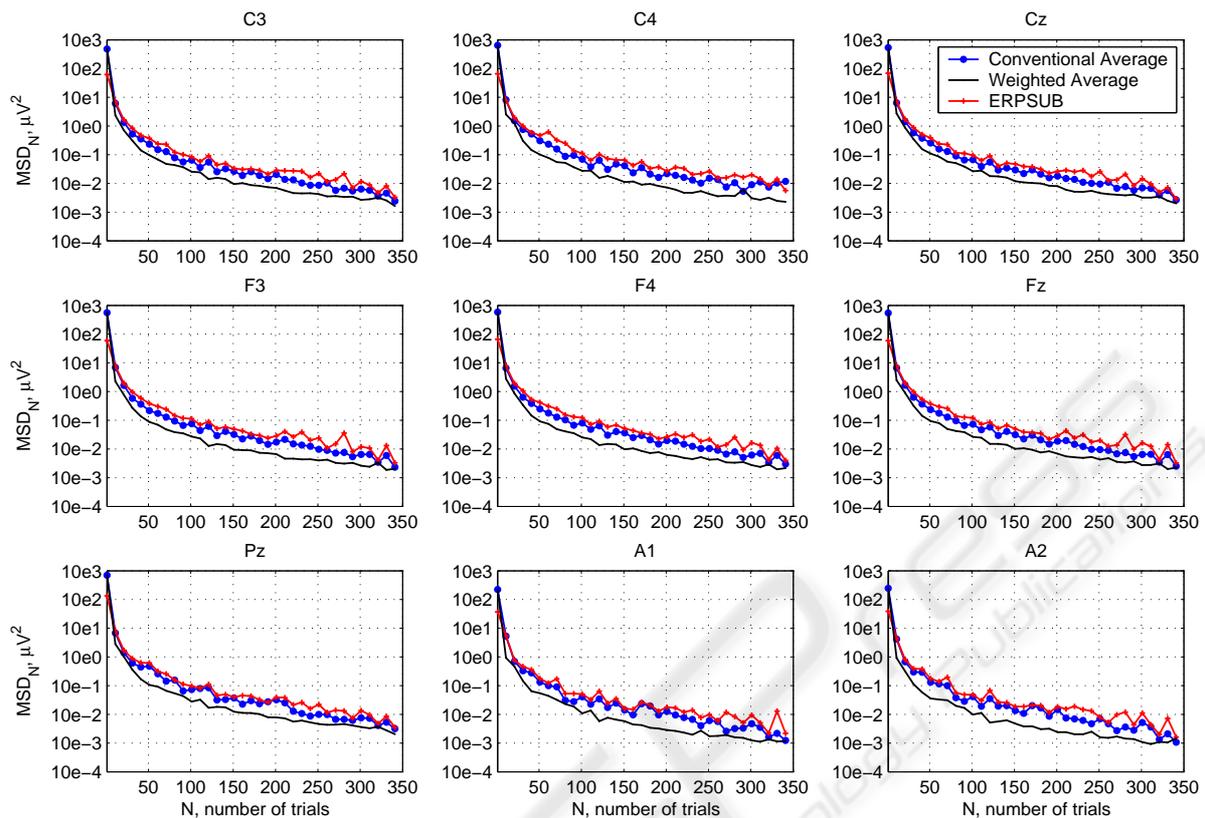


Figure 1: The averaged over 102 subjects MSD tracks provided by traditional averaging, weighted averaging and ERPSUB (nine EEG channels, 30ms deviant, logarithmic scale).

subspaces orthogonal and standardize the data to similar variances along all directions, and averaging ERP components should have the largest variances in contrary to the noise components, and, hence, subspaces can be extracted by standard linear Principal Component Analysis (PCA) algorithm (Hyvärinen et al., 2001; Oja, 1992). In practice, however, the variance of data most likely will reduce along all directions after the averaging, because subspaces are overlapped, and additive noise is always present, and, thus, pure signal/noise subspaces do not exist. In this case the results are interpreted in terms of SNR: higher SNR is obtained in data projected to the directions describing larger data variations after whitening and averaging. Thus, practically, we intend to separate the subspace of dimension  $N_{\text{ERP}}$  having maximal possible SNR from the subspace of dimension  $K - N_{\text{ERP}}$  with the minimal possible SNR. As one can see, ERPSUB is based on a sequence of linear transformations applied in a problem-specific manner to multidimensional EEG data and results in effective denoising of ERP signals (Ivannikov et al., 2007).

ERPSUB:

1. Whiten the centered concatenated data:

$$Z = D^{-1/2}W^T X, \quad (21)$$

where matrices  $D$  and  $W$  are taken from the eigenvalue decomposition  $\hat{\Sigma} = WDW^T$  of the estimated covariance matrix  $\hat{\Sigma} = XX^T/(TN - 1)$ .

2. Average the whitened data:

$$\bar{Z} = D^{-1/2}W^T \bar{X}. \quad (22)$$

3. Apply the standard linear PCA to the averaged whitened data

$$\bar{Y}'_{\text{ERP}} = \Delta_{N_{\text{ERP}}} \bar{W}^T D^{-1/2}W^T \bar{X}, \quad (23)$$

where matrix  $\bar{W}^T$  is obtained from the eigenvalue decomposition  $\bar{Z}\bar{Z}^T/(T - 1) = \bar{W}\bar{D}\bar{W}^T$  and  $\Delta_{N_{\text{ERP}}}$  is the diagonal projection matrix having ones on  $N_{\text{ERP}}$  first diagonal elements corresponding to the components contributing energy to ERP (maximal SNR) subspace and zeros otherwise. Here,  $N_{\text{ERP}}$  is the amount of assumed ERP sources present in EEG measurements. In practice, when pure signal/noise subspaces do not exist,  $N_{\text{ERP}}$  has different meaning interpreted in terms of SNR. In this

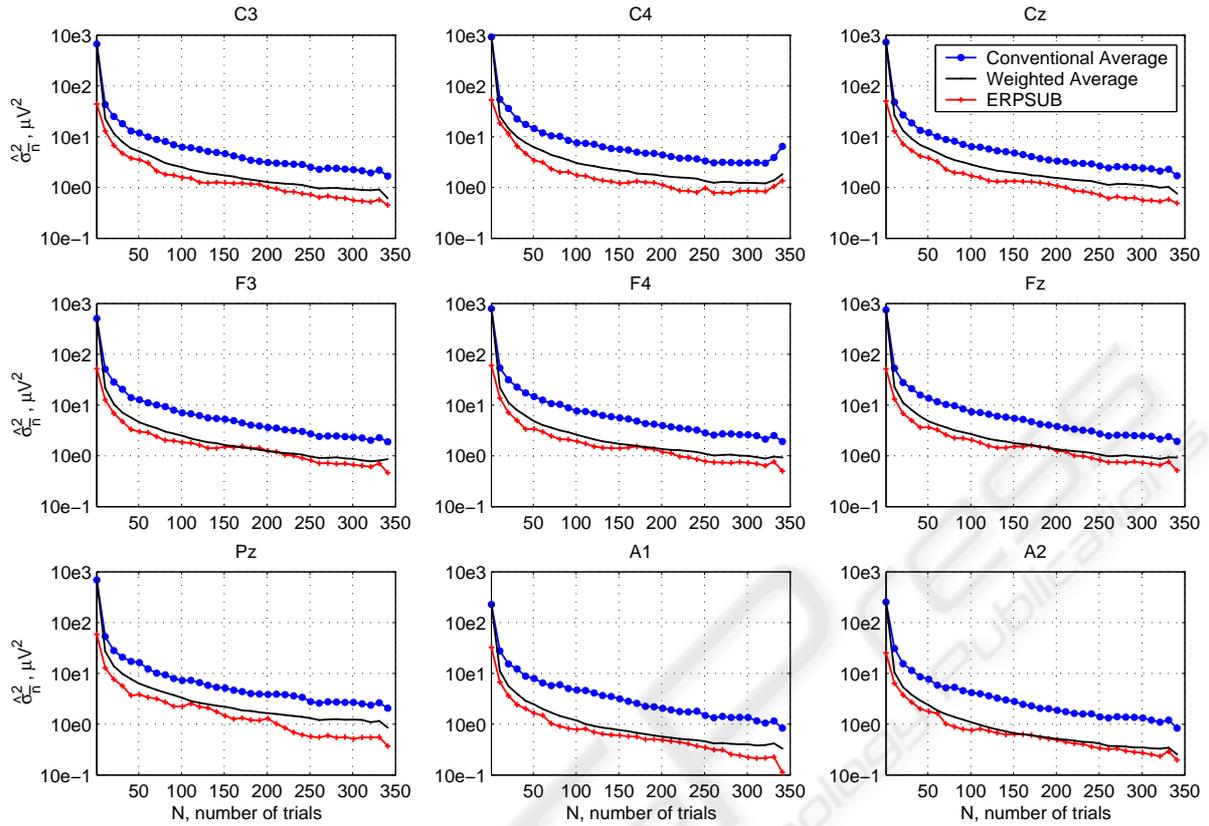


Figure 2: The averaged over 102 subjects tracks of the remaining noise variance in ERP estimates provided by traditional averaging, weighted averaging and ERPSUB (nine EEG channels, 30ms deviant, logarithmic scale).

case  $N_{ERP}$  is the amount of the components having largest SNR, which in our opinion describe ERP and noise variations in channels in proportions providing suitable SNR and tolerable ERP energy loss. Hence,  $\bar{Y}'_{ERP}$  is a matrix of the averaged components, where all components from noise (minimal SNR) subspace are zeroed. Note that ERP components have the largest corresponding eigenvalues and, thus, the component classification problem is solved automatically for fixed  $N_{ERP}$ . In addition, if the difference between eigenvalues corresponding to ERP and noise components is clearly observed, one can estimate  $N_{ERP}$  value providing optimal separation of the components into subspaces in the sense of SNR and ERP energy loss. Moreover, each  $K$ -dimensional trial  $X_i$  can be decomposed into the components using the same transformation as in (23):

$$Y'_{ERP_i} = \Delta_{N_{ERP}} \bar{W}^T D^{-1/2} W^T X_i, \quad (24)$$

4. The matrix  $\bar{Y}'_{ERP}$  containing only averaged components related to ERP subspace is then transformed back to the original data space (channels) to result in the subject average with the reduced

noise:

$$\bar{X}_{ERP} = W D^{1/2} \bar{W} Y'_{ERP}. \quad (25)$$

A similar relation applies also to a single trial denoising:

$$X_{ERP_i} = W D^{1/2} \bar{W} Y'_{ERP_i}. \quad (26)$$

## 4 EXPERIMENTAL RESULTS

It was noticed during the simulations and is theoretically predictable that the weighted averaging method is highly sensitive to the trials having small portions of variance concentrated on short time intervals. Generally, such trials do not carry much of the information and are usually recorded at the saturation state of the amplifier, when parts of the trials are truncated resulting in peaks alternating with flat periods. Saturation state occurs, when signal exceeds the dynamical range of the amplifier. The weighting coefficients  $a_i$ 's assigned for such trials are very large following the algorithm. As a consequence, when trials are weighted, peaks in truncated trials become very strong against a background of other trials' amplitude resulting in a

high frequency noise in the averaged signal. To annul the harmful consequences introduced by the truncated trials we performed the trial rejection procedure for our data before doing the computations. The successful upper limit of the trial's variance for the trial removal was  $30\mu V^2$ , which finally rejected all truncated trials in our database.

Apparently, for our problem the converged ERP estimate (subject average) is indicated by only insignificant change introduced by the consequent trial. We measure the amount of change between the two subsequent ERP estimates in one channel for method LABEL by MSD score:

$$\text{MSD}_N^{\text{LABEL}} = \frac{1}{T} \sum_{t=1}^T (\bar{x}_N^{\text{LABEL}}(t) - \bar{x}_{N-1}^{\text{LABEL}}(t))^2,$$

where  $\bar{x}_N^{\text{LABEL}}(t)$  denotes ERP estimate obtained after application of a particular method LABEL involving  $N$  trials. Thus, for example, for ERPSUB  $\bar{x}_N^{\text{ERPSUB}}(t)$  equals to a row in the matrix of averaged filtered channels  $\bar{X}_{\text{ERP}}$  corresponding to the considered channel; for weighted averaging  $\bar{x}_N^{\text{WA}}(t) = \sum_{i=1}^N \hat{a}_i x_i(t)$ , where  $\hat{a}_i$ 's are computed as in (17) substituting approximation from (15) for  $\sigma_{n_i}^2$  for all  $i = 1, \dots, N$ . To compare the convergence rates of ERP estimates provided by methods under consideration at increase of the number of processed trials, we computed averaged over 102 subjects MSD values for  $N = 1, \dots, 350$  (MSD tracks) for each method (see Fig. 1). We did this for the nine EEG channels and for 30ms deviant only, where ERP appeared to be the strongest. The value of  $N_{\text{ERP}}$  parameter of ERPSUB method was set to 3, that is, a good choice of maximal SNR subspace dimension for our data, because signal loss is insignificant and noise reduction is sufficiently high resulting in essential SNR increase (Ivannikov et al., 2007). According to the obtained results, the weighted averaging procedure outperforms both the traditional averaging scheme and ERPSUB algorithm, because MSD provided by weighted averaging, in general, decreases faster than for other methods at increase of the number of processed trials. The superiority of the weighted averaging here is probably a consequence of the core idea underlying the method. Weighted averaging is designed in a way that trials are 'equalized' in the sense of the variance. This should make the convergence of the ERP estimate smoother and faster. Although application of ERPSUB should result in higher noise reduction rate than the conventional average provides (Ivannikov et al., 2007), ERPSUB has shown the lowest convergence rate of MSD to zero. Most likely this happens, because new-coming trial influences the denoising of all previous trials by changing the projection axes. Since the shapes of all

filtered trials are affected, when new trial is added to processing, the difference between the two adjacent ERP estimates becomes more significant.

Therefore, in order to have a complete and fair comparative picture of the methods' performance, we also computed averaged over 102 subjects remaining noise variances in ERP estimates obtained under the same conditions as used in the first test (see Fig. 2). We used the following estimate of the noise variance in the averaged brain responses taken from (van de Velde, 2000):

$$\hat{\sigma}_{\bar{n}}^2 \text{LABEL} = \text{var}\left\{\frac{1}{N} \sum_{i=1}^N (-1)^i x_i^{\text{LABEL}}(t)\right\}, \quad (27)$$

where  $x_i^{\text{LABEL}}(t)$  is the modified trial  $x_i(t)$  obtained after application of method LABEL, and  $\hat{\sigma}_{\bar{n}}^2 \text{LABEL}$  is the estimate of the remaining noise variance in ERP estimate obtained after application of method LABEL involving  $N$  trials. For instance,  $x_i^{\text{WA}}(t) = \hat{a}_i x_i(t)$ , where  $\hat{a}_i$  is computed as in (17) replacing  $\sigma_{n_i}^2$  with the approximation from (15) for all  $i = 1, \dots, N$ ; and  $x_i^{\text{ERPSUB}}(t)$  equals to a row in the matrix of filtered trial  $X_{\text{ERP}i}$  corresponding to the considered channel. In this test the performance order of the methods appeared to be different. ERPSUB has shown now the highest effectiveness in the sense of the noise reduction rate, since the remaining noise variance in ERP estimate provided by ERPSUB, in general, decreased faster than for other methods at increase of the number of processed trials. This outstanding performance can be explained here by the algorithmic nature of ERPSUB, which simultaneously operates through all time, trial, and channel dimensions that allows more efficient extraction of the information discriminating ERP and noise from data. The conventional averaging has shown the lowest noise reduction rate in ERP estimate following the results of the test.

## 5 CONCLUSIONS

In this article we compared the performance of the three methods used nowadays in EEG research for ERP denoising: conventional averaging, weighted averaging and ERPSUB. For this purpose we carried out two tests investigating the convergence and the noise reduction rates in ERP estimates provided by the selected methods at increase of the number of processed trials. The convergence rate of ERP estimate appeared to be the highest for the weighted averaging technique and the lowest for ERPSUB. However, ERPSUB has shown stronger noise reduction power than the traditional and weighted averaging methods have. The noise reduction rate in ERP estimate provided by the

traditional averaging was the poorest among tested methods.

The paper touches practical issues the neuropsychology researchers are faced with during EEG/ERP data processing and analyzing. Namely, it points out the bottlenecks of the traditional averaging technique used for the time-locked brain responses denoising. The roots of these bottlenecks are connected to the violation of the assumptions underlying the averaging in real applications and insufficiently powerful 'decoding' of the relevant information 'encrypted' in the data. The weighted averaging method addresses the bottlenecks, which arise due to the violation of the assumptions underlying traditional averaging. We have shown that this strategy for improving the performance and the reliability of the traditional averaging technique can be derived based on different criteria and, in particular, SNR and mean-square error as it has been shown in (Hoke et al., 1984). ERPSUB tries to eliminate the second type of the bottlenecks, which come from the disability of the traditional averaging to effectively extract from data the overall available information related to ERP and noise discrimination. ERPSUB 'decrypts' more effectively this information hidden in data, since it operates through all time, trial, and channel dimensions.

## REFERENCES

- Davila, C. E. and Mobin, M. S. (1992). Weighted averaging of evoked potentials. *IEEE Transactions on Biomedical Engineering*, 39(4):338–345.
- Furst, M. and Blau, A. (1991). Optimal *a posteriori* time domain filter for average evoked potentials. *IEEE Transactions on Biomedical Engineering*, 38(9):827–833.
- Hoke, M., Ross, B., Wickesberg, R., and Lütkenhöner, B. (1984). Weighted averaging - theory and application to electric response audiometry. *Electroenceph. Clin. Neurophysiol.*, 57:484–489.
- Huttunen, T., Halonen, A., Kaartinen, J., and Lyytinen, H. (2007). Does mismatch negativity show differences in reading disabled children as compared to normal children and children with attention deficit? *Developmental Neuropsychology*, 31(3):453–470.
- Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent component analysis*. John Wiley & Sons, Inc.
- Ivannikov, A., Kärkkäinen, T., Ristaniemi, T., and Lyytinen, H. (2007). Extraction of ERP from EEG data. In *9th International Symposium on Signal Processing and its Applications*.
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., and Sejnowski, T. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37:163–178.
- Kalyakin, I., González, N., Joutsensalo, J., Huttunen, T., Kaartinen, J., and Lyytinen, H. (2007). Optimal digital filtering versus difference waves on the mismatch negativity in an uninterrupted sound paradigm. *Developmental Neuropsychology*, 31(3):429–452.
- Łęski, J. (2002). Robust weighted averaging. *IEEE Transactions on Biomedical Engineering*, 49(8):796–804.
- Łęski, J. and Gacek, A. (2004). Computationally effective algorithm for robust weighted averaging. *IEEE Transactions on Biomedical Engineering*, 51(7):1280–1284.
- Luu, P., Tucker, D., and Makeig, S. (2004). Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, 115(8):1821–35.
- Makeig, S., Westerfield, M., Townsend, J., Jung, T.-P., Courchesne, E., and Sejnowski, T. (1999). Functionally independent components of the early event-related potentials in a visual spatial attention task. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354:1135–44.
- Näätänen, R. (1992). *Attention and brain function*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Oja, E. (1992). Principal components, minor components, and linear neural networks. *Neural Networks*, 5:927–935.
- Pihko, E., Leppäsaari, T., and Lyytinen, H. (1995). Brain reacts to occasional changes in duration of elements in continues sound. *NeuroReport*, 6:1215–18.
- van de Velde, M. (2000). *Signal validation in electroencephalography research*. PhD thesis, Eindhoven University of Technology.
- Vigário, R. (1997). Extraction of ocular artifacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, 103:395–404.