Robust moving averages, with Hopfield neural network implementation, for monitoring evoked potential signals

N. Laskaris a,*, S. Fotopoulos b, P. Papathanasopoulos c, A. Bezerianos a

a Department of Medical Physics, School of Medicine, University of Patras, 26500 Patras, Greece
b Electronics Laboratory, Department of Physics, University of Patras, 26500 Patras, Greece
c Neurology Clinic, School of Medicine, University of Patras, 26500 Patras, Greece

Accepted for publication: 6 December 1996

Abstract

This technical note describes a robust version of moving averages, that enables reliable monitoring of the evoked potential (EP) signals. A cluster analysis (CA) procedure is introduced to robustify the signal averaging (SA). It is implemented via a Hopfield neural network (HNN), which performs selection of the trials forming a cluster around the current state of the EP signal. The core of this cluster serves as an estimate of the instantaneous EP. The effectiveness of the method, indicated by application to real data, and its computation efficiency, due to the use of simple matrix operations, makes it very promising for clinical observations. © 1997 Elsevier Science Ireland Ltd.

Keywords: Evoked potential; Moving averages; Clustering; Hopfield neural network; Monitoring

1. Introduction

The study of EP requires extraction of the actual brain’s response from accompanying noise. Under the assumption of constant EP signal, this is accomplished via the recording of a great number of signals and the computation of the ensemble average. When we record an EP signal which is, or might be, changeable, the method of moving averages can be used (Soustiel et al., 1993). A window of fixed depth \( w \) slides along the stack of the recorded trials and signal averaging is performed on the continually updated set of trials. The calculated averages can be displayed as a sequence of waveforms for easy visual detection of changes (Fig. 1) (Rossini et al., 1982; Chiappa, 1990).

The main shortcoming of the method is the trade-off between using a large window for sufficient noise reduction and a small window for preservation of EP rapid changes.

The problem lies in the operation inside the sliding window. The performed signal averaging is sensitive to artifacts whenever a small number of trials are to be processed. Even if an on-line artifact rejection routine is activated the low amplitude artifacts can escape, deteriorating the estimate of the EP (Gevins, 1987). Therefore, a large \( w \) is needed, but in this way EP changes are smoothed out and may become undetectable. Moreover, the conventional operation provides waveforms with no index for their quality (e.g. signal-to-noise ratio) that could be used for adjusting the depth \( w \).

In this work, we introduce a robust averaging (RA) technique based on the concept of cluster analysis. The operation inside the window is replaced with a search of finding \( k \) among the \( w \) trials that form the most compact cluster. An estimation of the EP signal is provided through the core of this cluster and measurements of the signal and noise power for the cluster are given. In this way effective elimination of artifacts, revealing of the current EP and monitoring of the signal-to-noise ratio is achieved simultaneously.

The search inside the window is implemented via a Hopfield neural network, exploiting the ability of this family of networks to solve hard combinatorial optimization problem and sharing their possibility of a VLSI realization.

Special attention is given to the visualization of the results that could provide complementary information about the evolution of the brain’s response (Janssen et al., 1991; Machado et al., 1991) and help to design the control parameters \( w \) and \( k \).

First we introduce the idea of selective averaging and describe the clustering procedure as a combinatorial opti-
mization problem, then design the Hopfield neural network to solve it and give the algorithm of the proposed method. Finally, we demonstrate the new method by applying it to visual EP signals.

2. Method

The ith trial is considered as a p-dimensional multivariate observation/vector with components the signal value at time instants \( t = 1, \ldots, p \):

\[
x_i = \{x(1) \times(2) \ldots x(p)\}^T = s_i + n_i
\]

(1)

The trial \( X_i \) is the superposition of the actual EP signal \( S_i \) and the additive noise \( N_i \).

A window slides, or moves with constant step, along the stack of the trials operating on the \( w \) included vectors and produces the estimate \( Y_i \) of the current state of the signal (Fig. 1).

\[
y_i = O(\{x_i^w\})
\]

(2)

where \( \{x_i^w\}/i = 1, \ldots, w = [x_{-w/2} \ldots x_{-1} \ldots x_{w-1} \ldots x_{w/2}] \)

Here we replace the conventional operation of signal averaging with that of a selective averaging which is based on cluster analysis.

2.1. Cluster analysis for selective averaging

Cluster analysis is a powerful method for determining the existence of subgroups and unmasking outliers in a given sample of vectors (Manly, 1994). Inside the window, the goal is to find those \( k \) subsets and unmask the outliers. Which also serves as an index of the noise power (Raz et al., 1986): the objective is to find those \( k \) clusters among the set of \( \{x_i^w\}^w \), that form the collection \( \{Z_j^w\}^w \) with the smallest dispersion (Tou and Gonzalez, 1974):

\[
J_{\{x_i^w\}} = \sum_{k=1}^{w} \sum_{j=1}^{k} \left| x_j^w - z_{ave} \right|^2
\]

(3)

where \( z_{ave} = \frac{1}{k} \sum_{j=1}^{k} x_j^w \) and \( \left| x_j^w - z_{ave} \right|^2 = \frac{1}{k} \sum_{i=1}^{k} d_i^2 \)

which also serves as an index of the noise power (Raz et al., 1986) for the certain sample of trials:

\[
J_{\{x_i^w\}} = \sum_{k=1}^{w} \sum_{j=1}^{k} \left| x_j^w - z_{ave} \right|^2 = \sum_{i=1}^{k} \sigma_i^2(t) = p \times \text{noise power}
\]

(4)

It can be proved that \( J \) can be expressed in the form of a summation of pairwise distances:

\[
J_{\{x_i^w\}} = \frac{1}{2M(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| x_j^w - x_j^w \right|^2
\]

(5)

By defining an indicator set \( \{u_i^w\}^w \)

\[
\{u_i^w\}^w = [u_i^w/i = 1, \ldots, w] \text{ with } \sum_{i=1}^{w} u_i^w = k
\]

(6)

the problem can be formulated as a minimization of an objective function with arguments the set \( \{u_i^w\}^w \)

\[
J^*_{\{u_i^w\}^w} = \sum_{i=1}^{k} \sum_{j=1}^{w} u_i^w \left| x_j^w - x_i^w \right|^2
\]

or equivalently

\[
J^*_{\{u_i^w\}^w} = \sum_{i=1}^{k} \sum_{j=1}^{w} u_i^w \left| x_j^w \right|^2 - x_i^w \left| x_i^w - x_j^w \right|^2
\]

(7)

(8)

Since an exhaustive test of all possible \( k \) out of \( w \) combinations is computationally prohibitive, especially for large \( w \) and \( k \), we propose a neural network implementation for the selection of \( \{Z_j^w\}^w \). For convenience the superscripts \( l,m \) are dropped in the following paragraphs, since we focus on a certain \( \{X_i\}^w \).

2.2. Hopfield NNet implementation of the clustering procedure

Hopfield neural networks (NNet) are single layer fully-connected networks (see Fig. 1). Each node (neuron) \( i \) receives signals \( V_i \) from all the others and in addition an external bias \( I_i \) producing its output \( V_i \) according to the formula:

\[
V_i = g(c) = g(\sum_{j=1}^{N} T_{ij} V_j + I_i),
\]

with \( g(c) = 0.5 (1 + \tanh(\lambda c)) = 0.5 (1 + e^{2\lambda c} - e^{-2\lambda c}) / e^{2\lambda c} + e^{-2\lambda c}) \)

(9)

It has been shown (Tawk and Hopfield, 1986) that when the weights \( T_{ij} \) on the link from \( i \) to \( j \) node are symmetric, \( T_{ii} \) are zeros and the gain parameter \( \lambda \) is sufficiently high, the network, starting from a random initial state, iterates, converging always to a stable state which is a minimum of the computational energy of the system:

\[
E = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} V_i V_j - \sum_{i=1}^{N} I_i V_i
\]

(10)
When Hopfield neural networks are used for function optimization, the objective function to be minimized and any constraints to be satisfied are written as energy functions in a form of Eq. (10) and additively combined to an overall function $E^*$. The comparison between $E^*$ and $E$ leads to the design, i.e. definition of links and biases, of the network that can solve the problem. For a detailed discussion see Tagliarini et al. (1991).

In our approach we use $w$ neurons with each one corresponding to a certain trial. A possible selection of $\{Z_j\}$ is encoded in this network by including the $i$th trial in the cluster $\{Z_j\}$ when the neuron is ‘on’ i.e. $V_i = 1$. This is sketched in Fig. 1.

The ‘energy’ $E^{obj}$ of a possible selection is the cluster compactness: from Eq. (8),

$$E^{obj} = J^2(\{V_j\}) = \sum_{i=1}^{\infty} V_i V_j D(i,j)$$

and in the form of Eq. (10):

$$E^{obj} = -\frac{1}{2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} T_{ij}^{obj} V_i V_j - \sum_{i=1}^{\infty} t_i^{obj} V_i$$

where $D(i,j) = |x_i - x_j|^2$ pairwise distance

and in the form of Eq. (10):

$$E^{obj} = -\frac{1}{2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} T_{ij}^{obj} V_i V_j - \sum_{i=1}^{\infty} t_i^{obj} V_i$$

with $T_{ij}^{obj} = T_{ji}^{obj} = -2 D(i,j)$

$$= \begin{cases} 0 & \text{if } i=j \\ -2 \frac{1}{2} |x_i - x_j|^2 & \text{if } i \neq j \end{cases} \quad \wedge \quad t_i^{obj} = 0$$

Such a selection should satisfy the hard constraint that $k$ neurons should be ‘on’. The ‘energy’ $E^{con}$ is

$$E^{con}_1 = (\sum_{i=1}^{\infty} V_i - k)^2 = \sum_{i=1}^{\infty} V_i V_j - 2k \sum_{i=1}^{\infty} V_i + k^2$$

$$E^{con}_2 = -\frac{1}{2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (-2) V_i V_j - \sum_{i=1}^{\infty} (2k) V_i$$

In addition, a second constraint is imposed, to force the continuous valued outputs of the neurons to be ‘digital’, with corresponding ‘energy’:

$$E^{con}_2 = \sum_{i=1}^{\infty} V_i (1 - V_i) = \sum_{i=1}^{\infty} V_i - \sum_{i=1}^{\infty} V_i^2$$

Moreover, it counterbalances the non-zero self-feedback links of the first term in the previous constraint.

The total ‘energy’ of the constraints is:

$$E^{con} = E^{con}_1 + E^{con}_2 = -\frac{1}{2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (-2) V_i V_j - \sum_{i=1}^{\infty} (2k - 1) V_i$$

and in the form of Eq. (10):

$$E^{con} = -\frac{1}{2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} T_{ij}^{con} V_i V_j - \sum_{i=1}^{\infty} t_i^{con} V_i$$

with $T_{ij}^{con} = T_{ji}^{con} = \begin{cases} 0 & \text{if } i=j \\ -2 & \text{if } i \neq j \end{cases} \quad \wedge \quad t_i^{con} = 2k - 1$

The overall ‘energy’ function, coming from the combination of the objective and the constraint energy, is:

$$E^* = E^{obj} + \alpha E^{con} =$$

$$-\frac{1}{2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (-2[D(i,j) + \alpha] V_i V_j - \sum_{i=1}^{\infty} \alpha(2k - 1) V_i$$

and in the form of Eq. (10):
$E^* = -\frac{1}{2} \sum_{i=1}^{w} \sum_{j \neq i} T_{ij}^* V_i V_j - \sum_{i=1}^{w} I_i^* V_i$

with $T_{ij}^* = T_{ij}^{obj} + \alpha T_{ij}^{con}$ and $I_i^* = \alpha I_i^{con}$

(18)

where $\alpha$ is a sufficiently large positive scaling factor, easy to set experimentally.

The $E$ (see Eq. 17) consists of two distinct terms: the first one deals with the ‘clustering’ while the second deals with the hard constraints. The scaling factor $\alpha$ controls the relative importance of these two terms. The conventional technique for the definition of $\alpha$ is a ‘trial and error’ procedure, the constant $\alpha$ are obtained averages. On the basis of these measurements typical values for $\alpha$ can be proposed; e.g. $\alpha = 15$ for visual EP signals and $\alpha = 120$ for brain-stem auditory EP signals. On the contrary, $k$ is strongly dependent on the subject under examination. Consequently, only a range of values can be proposed: $k = 0.6w - 0.9w$.

The averaging of the $k$ trials in $\{Z_{i}^{w}\}$ increases the signal-to-noise ratio, approximately $k$ times (Krieger et al., 1993). So the value of $k$ is crucial. It should be high, but

2.3. The algorithm of robust averaging

The procedure of the introduced RA is summarized in the sequel, where proposed values for the constants are given. When the window $w$ centers at $x_l$:

(1) Form the distance matrix of the set $\{X_i\}$:

$D^{w} = \begin{cases} 0 & D_{i,j}^{w} = \Vert X_i - X_j \Vert^2 \end{cases}$

[hint: $D_{i,j}^{w} = D_{i-j}^{w} = k(i + j + k)$]

(19)

$T^* = -2 D^{w} + \alpha T^{con}$ and $I^* = \alpha I^{con}$

(20)

(2) Compute the weights and biases matrices for the Hopfield neural network (Eq. (12, 16, 18)):

where $\alpha = 5 \times E(\delta(i,j)) = 5 \times (w(w-1))^{-1}$

(21)

(3) Simulating the Hopfield network: follow the steps to compute the activation level of neuron $i$ at time $t$, $V(t)$.

Step (i): at $t = 0$, set input $c_i(0)$ to a slightly-perturbed value around $c = 1/\lambda(tanh^2(2k/nw - 1))$ and compute $V(0)$ through Eq. (9), with $\lambda = 2$.

Step (ii): at $t > 0$, visit at random the neuron $m$, compute the input $c_i(t)$ through $c_i(t) = T_{m,n} V_T + I_n^*$

(22)

Step (iii): repeat Step (ii) until equilibrium or a predefined number of iterations is completed. Then the pattern of activations represents the optimized solution for the indicator set $\{a\}^{w}$.

(4) Assign as output of the window, the core of the formed cluster:

$y_i = O(|x_i^{w}|) = \frac{\sum_{i=1}^{w} V_i x_i}{\sum_{i=1}^{w} V_i}$

(23)

Additionally, the noise power (np) can be estimated (Eqs. (4, 5, 8)) as:

$np_i = np(\alpha'^{w}) = \frac{1}{2pk(k-1)} V D^{w} V^T$ (24)

and the signal power (sp) as (Raz et al., 1988):

$sp_i = sp(\alpha'^{w}) = \frac{1}{p} \Vert y_i \Vert^2 - \frac{1}{k} np_i$ (25)

2.4. Selection of the control parameters: visualizing the results of cluster analysis

The last two equations play an important role in the adjustment of the control parameters $w$ and $k$, since they provide reliable measurements for the quality of the obtained averages. On the basis of these measurements typical values for $w$ can be proposed; e.g. $w = 15$ for visual EP signals and $w = 120$ for brain-stem auditory EP signals. On the contrary, $k$ is strongly dependent on the subject under examination. Consequently, only a range of values can be proposed: $k = 0.6w - 0.9w$.
without violating the condition for the formation of a compact cluster.

In the sequel, we propose a proper ‘projection’ of the original $p$-dimensional vectors onto a 1D space that can enable a visualization useful for having a sense about the distribution of the vectors $\{X_i\}_{i=1}^w$ in and around the cluster $\{Z_i\}_{i=1}^w$.

Using the distances of the vectors from the core of the formed cluster, we can construct 1D charts of the single trials. These charts would be informative about the cluster compactness and the presence of outliers and therefore useful for guiding the design of $k$ (and $w$) and also for monitoring the evolution of noise power.

To avoid computation of new distances, aiming to keep the computational cost low, we adopted a slightly different procedure with very similar results. We utilize elements from the $D_{i,w}$ to perform the mapping of the vectors to scalars:

$$
X_{i,w} \rightarrow d_{i,w} = \begin{cases} 
\frac{1}{(k-1)}D_{i,w}V^T & \text{if } V_i = 1 \\
\frac{1}{k}D_{i,w}V^T & \text{if } V_i = 0 
\end{cases} 
$$

(26)

The scalar attached to each vector is its average distance to the $\{Z_i\}_{i=1}^w$. It expresses the membership of the vector to the formed cluster. This mapping was found to be quite insensitive to the selection of $k$.

Therefore, the proposed algorithm can be run once with a small $k$ and this visualization can give a sense about the upper value of $k$. When, even with this value, the obtained SNR ($= k \times \text{Eq. (25)}/\text{Eq. (24)}$) is not acceptable, the $w$ has to (slightly) increase and the procedure to be repeated.
3. Experimental results

Pattern shifted visual EP signals were used for the demonstration of the method. The stimulus was a checkerboard field with 52° checksize and 90% contrast, reversed with 1.6 rev/s. The signals were recorded, in the Oz electrode position, according to the guidelines of an IFCN Committee (Celesia et al., 1993), with a Nicolet Compact IV, which uses a rejection routine based on voltage threshold. After 1–100 Hz band-pass filtering (-3 dB down, 12 dB/octave) and digitization (at 512 Hz) the signals were transferred to a PC for further processing.

The merit of clustering is demonstrated in Fig. 2. The output of the moving window when the operation is signal averaging (w = 15) and robust averaging (w = 15, k = 9) is shown in parallel for certain instances l. The signal-to-noise ratio (SNR) of the \( \{ X_i \}^{15} \) (\( = E_q(25)/E_q(24) \)) with \( k = w \) and \( V = \{ 1,1,1 \} \) and the corresponding \( \{ Z_i \}^{15} \) is shown (Fig. 2 below) for every l, where it becomes evident that the clustering procedure offers significant increase in SNR.

In Fig. 3 the visualization of cluster analysis, given for certain l, is demonstrated for two different subjects. The charts follow different patterns for the subject presenting limited artifact contamination (left), when compared with the ones from a subject presenting high artifact contamination. Both stacks of charts present a displacement to the right, as l increases, which is indicative of an increasing trend for the noise power.

4. Discussion

The proposed enhanced version of moving averages can facilitate clinical observations concerning the study of the brain’s response evolution. An important characteristic of this version is that a quality assurance procedure is embodied, via the signal-to-noise ratio measurements of the obtained average waveforms. In this way, wrong choice of the control parameters can be avoided. An incorrect choice of \( w \) has similar effects as in the conventional method, while an incorrect choice of \( k \) affects the robustness of the proposed method.

The computation efficiency is the main advantage of the proposed method, since simple matrix operations are employed and the main part is implemented by a Hopfield neural network that, when it is simulated, converges in a few iterations and if realized via VLSI hardware converges in a few nanoseconds (Kechriotis and Manolakos, 1996). Therefore, experimentation with different values of \( k \) (and \( w \)) and selection of the values offering the best performance in terms of signal-to-noise ratio, are feasible.

References