ANALYSIS OF REPEATED MEASUREMENTS IN CURVES (BRAIN POTENTIALS)

Kongming Wang, Henri Begleiter, Bernice Porjesz, SUNY
Kongming Wang, Box 1203, SUNY HSC, 450 Clarkson Avenue, Brooklyn, NY 11203

Key Words: Brain potentials, Multiresolution analysis.

Abstract:

Huge data of repeated measurements in waveforms are collected in neurodynamic research. We are particularly interested in the problem of analyzing event-related potentials (ERP) recorded from humans. In this short communication, we concentrate on extracting evoked electrical signals of the human brains from the recorded mixture of evoked electrical signals to stimuli, background electroencephalogram (EEG), artifacts (e.g., eye movements), and noise (instrumental and human error). This is one of the challenging topics in ERP analysis because it is the basis of some important theoretical and clinical research (e.g., source localization, effects of alcohol on human brains). Our method is based on multiresolution analysis.

A level dependent wavelet shrinkage is applied to remove the noise and the high frequency background EEG. Then a linear fitting across trials is applied to remove the low frequency EEG and to make sure that the extracted ERPs are slowly changing from trial to trial.

1. Introduction

Event-related potentials (ERP) are electrical signals of human brain, caused by external events such as sound and light, and recorded by placing electrodes on head. Figure 1 shows a diagram of recording visually evoked brain potentials. In this experiment, subjects are presented with up to 350 visual stimuli on a PC screen with a uniform inter-stimulus interval of 1.6 seconds. There are 280 non-target stimuli in the shape of an outline of a square, 35 target stimuli in the shape of an X, and 35 novel stimuli each a different colored polygon or other geometrical figure. These stimuli are presented to subjects in random order. Subjects are instructed to respond only to target stimuli by pressing a button. We often use 64 electrodes for recording (nose channel as reference, two electrodes for monitoring eye blinks, and 61 electrodes for recording brain potentials). A trial simply means the 1.6 seconds signal recorded after a stimulus is presented. In reality the recorded signals are a mixture of evoked activity and noise (background electroencephalogram (EEG) and human error in placing electrodes) rather than pure evoked activity. Therefore, extracting ERPs from the recorded mixture is a fundamental step in the study of ERPs.

Let $s_{ijk}(t)$ be the recorded signal from i^{th} channel, j^{th} subject, k^{th} trial, and at time t. The most frequently used model for ERP [3, Chapter 5] is

$$s_{ijk}(t) = p_{ijk}(t) + n_{ijk}(t) \tag{1}$$

where p_{ijk} is the unobserved evoked potential and n_{ijk} is the noise process, a time series of mean zero. This model is a first and probably crude approximation, but useful results have been obtained with its use. In addition, many methods of extracting signals from noisy data sets are based on this model.

Figure 2 plots some EEG segments (brain potentials when resting) and repeated ERP trials with target stimuli. One can see from Figure 2 that the signal (ERP) to noise (EEG) ration is very low, normally from 0.1 to 1.

The conventional method of averaging the recorded mixture over a number of identical trials assumes that the evoked potentials are time-locked. That is, $p_{ijk}(t) = p_{ij}(t)$ for repeated trials $k = 1, \cdots$. With K trials using identical stimulus, the ERP $p_{ij}(t)$ is estimated by

$$\hat{p}_{ij}(t) = \frac{1}{K}\sum_{k=1}^{K} s_{ijk}(t).$$

There are shortcomings with this estimation. In reality the evoked potentials are not completely time-locked and the averaging suppresses the individual characteristics of single trials p_{ijk}. These individual characteristics such as the P3 (the positive component around 300 msec after stimuli) latencies and amplitudes might be more interesting since they

This research is supported by N.I.H. grants AA 05524 and AA 02686 to HB.

Reprinted from the 1998 Proceedings of the Section on Statistics in Epidemiology of the American Statistical Association

give information on the degree of attention of the person and reveal the habituation process [10]. Even in the case of time-locked ERPs, one might have only a small number of trials (usually 20 - 35 for target and novel stimuli) and the signal to noise ratio (SNR) would not be improved significantly.

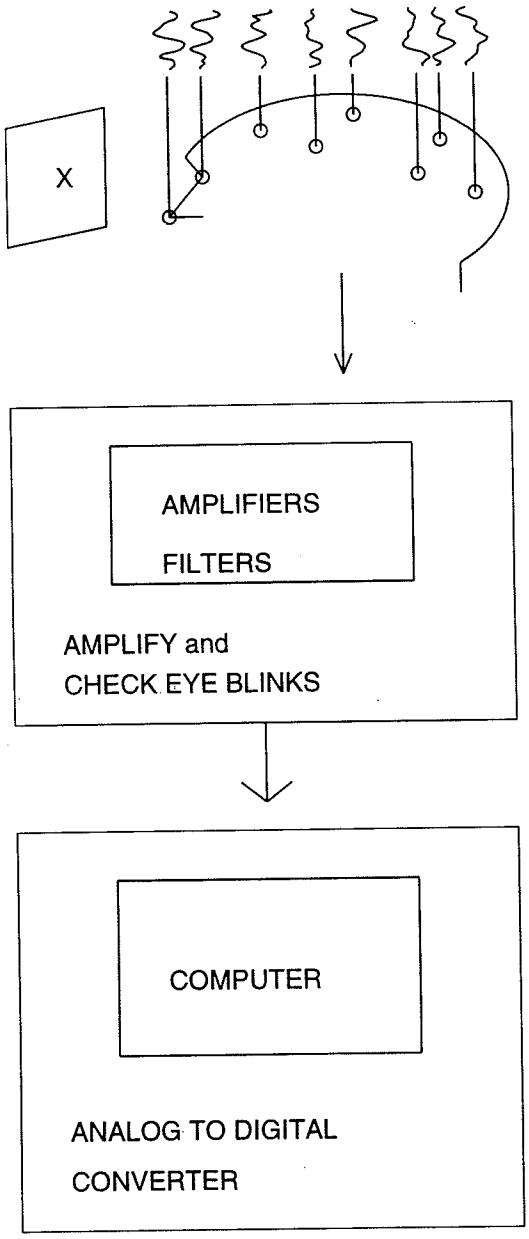

Figure 1. Diagram of recording visually evoked potentials.

Some methods have been proposed to overcome these shortcomings. A multivariate composite estimator (MCE) [14] is proposed for estimating p_{ij}. MCEs are nonparametric estimators resulting from fitting the data at hand by weighted least-square procedures, where the weight matrix depends on the covariance matrix of EEG and where the high correlation of the signals from different electrodes are taken into account. The non-stationary problem of EEG is taken into account because of the involvement of the covariance matrix of the prestimulus signals. This method results in better SNR for estimating p_{ij}, but it does not estimate the single trial ERPs p_{ijk}.

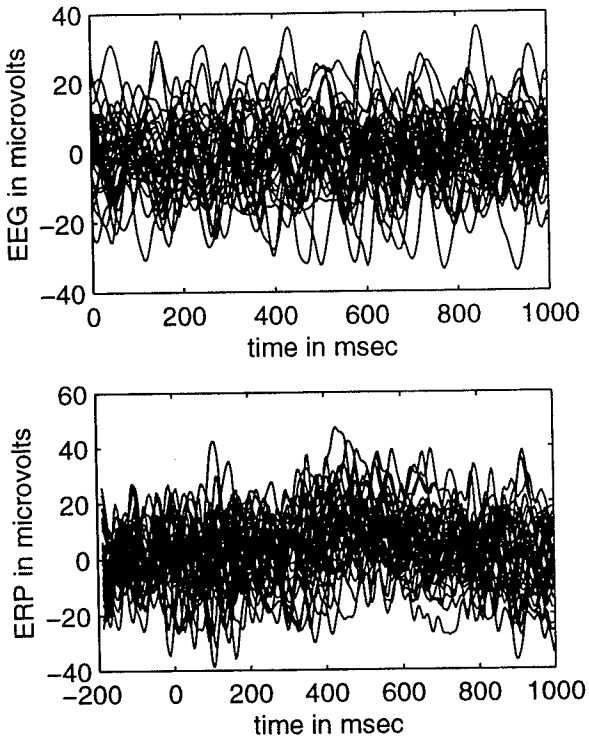

Figure 2. Plot of the EEG and ERP of a subject: the 35 segments of EEG (top: each of 1 second long) and the 35 repeated trials of ERP with target stimuli (bottom: stimuli are presented at time 0), recorded from a specified electrode (PZ).

A robust moving average procedure [7] is introduced for improving the signal to noise ratio (SNR) and for identifying outliers of single trials. A window of fixed depth slides along the stake of single trials. At each step, a cluster analysis is applied to the trials falling in the window to pick up a fixed number of trials which has the smallest dispersion among all such subgroups. The average of the selected trials is the estimated ERP at this step. The SNR is not improved very much.

In [4], ERPs of single trials are modeled parametrically by a finite sum of damped sinusoids. The parameters are estimated by weighted nonlinear least-

square procedures to fit the data at hand, given the model order (number of sinusoid terms). The weight matrix is estimated from the covariance matrix of the prestimulus signals. This means in a sense that EEG signals are decorrelated and normalized such that EEG is then removed by the least-square procedure. They used a model of order nine for their data according to the previous results [5,6]. The model order is not chosen by a data-adapted method, and it is not easy to get a good estimate of the covariance matrix of the background EEG from the short prestimulus signals.

In this paper we propose a nonparametric procedure for estimating single trial ERPs. This procedure is based on the model (1) and multiresolution wavelet analysis. Since the background EEG is a correlated noise process, the size of noise contained at different level of wavelet coefficients is different. Hence a level-dependent shrinkage is applied to remove noise. At each wavelet decomposition level, the noise level is computed from the estimated autocovariance function of the background EEG. This estimation of the noise level is more accurate than the estimations obtained from the size of the wavelet coefficients at the decomposition level. To enhance the estimation of ERP by wavelet shrinkage, a local linear fitting of the shrunk wavelet coefficients across trials is performed to remove the low frequency EEG and to enforce that the estimated ERPs are slowly changing from trial to trial. Finally, the shrunk and smoothed wavelet coefficients are transformed into signals in the time domain.

In the rest of the paper we will use simplified notation. Since the method works on repeated trials from a given subject and electrode, the subscripts i, j of model (1) will be dropped out. That is, the model can be written as

$$s_k(t) = p_k(t) + n_k(t). \tag{2}$$

2. Method

For some details on wavelet theory and applications, we refer to [8] and [9]. The latter presents a software package *Wavelet Toolbox*. The wavelet transforms in this paper are performed by using the toolbox. Some applications of the multiresolution analysis to ERP data are reviewed in [12].

Step 1: Estimating the autocovariance function of EEG

Since the stimulus application time is independent of any components in the ongoing background EEG and since the stimuli (target, nontarget, novel) are presented in random order, it is reasonable to assume that the noise is independent from trial to trial under a given stimulus type. Therefore,

$$E[n_k(t)n_j(u)] = 0, \quad j \neq k, \quad \text{all} \quad t, u.$$

The within-trial autocovariance function of EEG is

$$\sigma(u) = E[n_k(t)n_k(t+u)], \quad \sigma(-u) = \sigma(u), \quad \text{all} \quad k.$$

The ERPs are slowly changing signals in the following sense. First, the ERPs change smoothly in the time direction. Secondly, the trial-to-trial change in consecutive trials is small. For such slowly changing ERPs, the successive difference method is effective for estimating the autocovariance function $\sigma(u)$ [11]. The estimator is

$$\hat{\sigma}(u) = \frac{\sum_{t=1}^{T-u}\sum_{k=1}^{K-1} Q_u(t,k)}{2(T-u)(K-1)}, \quad u = 0, 1, 2, \cdots$$

with $Q_u(t,k) = [s_{k+1}(t+u) - s_k(t+u)][s_{k+1}(t) - s_k(t)]$ and $\hat{\sigma}(-u) = \hat{\sigma}(u)$.

Step 2: Wavelet transform and shrinkage

The discrete wavelet transform is an orthogonal linear transform

$$c_k = W s_k.$$

Here $s_k = (s_k(1), \cdots, s_k(T))'$ is the recorded signal of the trial k, W is the transform matrix depending on a specified wavelet and is orthogonal ($WW' = I$). The rows of W are simply the coefficients of the lower-pass filters (for extracting trend) or the high-pass filters (for extracting details). These filters are constructed from the specified wavelet, scaled to have longer support (zoom-out) or shorter support (zoom-in), and shifted to center at different locations. Specifically, W can be written as

$$W = \begin{pmatrix} W_0 \\ W_1 \\ \vdots \\ W_J \end{pmatrix}$$

if the wavelet decomposition is performed at J scale levels. The rows of W_0 are the coefficients of the lower-pass filters at the coarsest scale level and W_0 is a circulant matrix where one row is a circularly shifted version of another row. The rows of W_j, $j \geq 1$, are the coefficients of the high-pass filters at the j^{th} scale level and W_j is a circulant matrix. This is the structure of the multiresolution wavelet analysis.

The maximum possible scale level J depends on the length of the signal s_k and the specified wavelet. We will taken $J = 6$ in this paper according to our simulations and the suggestion of [9].

With the decomposition of W, we can write c_k as

$$c_k = \begin{pmatrix} c_k^0 \\ c_k^1 \\ \vdots \\ c_k^J \end{pmatrix}$$

with $c_k^j = W_j s_k$, $j = 0, 1, \cdots, J$. Note that $c_k^j = W_j p_k + W_j n_k$ where $p_k = (p_k(1), \cdots, p_k(T))'$ is the ERP to be estimated and $n_k = (n_k(1), \cdots, n_k(T))'$ is the noise of trial k. The term $W_j p_k$ is the wavelet transform of the ERP of trial k, while $W_j n_k$ is the wavelet transform of the noise. It is straightforward to compute the covariance matrix

$$COV[c_k^j] = COV[W_j n_k] = W_j \Sigma W_j'$$

where Σ is a $T \times T$ circulant matrix whose first row is $(\sigma(0), \sigma(1), \cdots, \sigma(T-1))$. Since W_j is also a circulant matrix, all the diagonal elements of $COV[c_k^j]$ are the same and the square root of their value

$$v_j = \sqrt{\text{mean of the diagonal elements of } COV[c_k^j]}$$

gives the size of the noise contained in the coefficients c_k^j. The off-diagonal elements of $COV[c_k^j]$ is almost zero. This decorrelation property of wavelet transform is a consequence of the fact that wavelets are "almost-eigenfunctions" of many operators [2]. This decorrelation justifies the one-by-one shrinkage of the wavelet coefficients for removing noise. Furthermore, because the mapping from the measured signals to the wavelet coefficients on any particular level is essentially a band-pass filter, there will tend to be little or no correlation between the wavelet coefficients at different levels [15]. That is, c_k^j and c_k^i are almost uncorrelated for $i \neq j$.

To remove noise from c_k^j, we apply a soft shrinkage with a universal threshold [1]

$$\lambda_j = v_j \sqrt{2 \log(d_j)}$$

where d_j is the length of c_k^j. The shrunk coefficients are given by

$$\hat{c}_k^j(i) = \text{sign}(c_k^j(i)) \max(0, |c_k^j(i)| - \lambda_j), \; i = 1, \cdots, d_j.$$

Perform this shrinkage at all J levels and keep c_k^0 unchanged, because it is the wavelet coefficients of the trend of ERPs. Set

$$\hat{c}_k = \begin{pmatrix} c_k^0 \\ \hat{c}_k^1 \\ \vdots \\ \hat{c}_k^J \end{pmatrix}.$$

Note that the universal threshold of [1] is defined by $\lambda_j = v_j \sqrt{2 \log(T)}$ where T is the length of the signal s_k. In view of the results of [15], the optimal threshold should be $\lambda_j = v_j \theta$ where θ is chosen to minimize a loss function and $\theta \in [0, \sqrt{2 \log(T)}]$. Our choice $\theta = \sqrt{2 \log(d_j)}$ is in the above range and can be thought of as "universal threshold" at each decomposition level.

Step 3: Extracting slowly changing ERPs

Let N denote the length of c_k (the number of wavelet coefficients of s_k). Then the shrunk wavelet coefficients of the ERPs from K trials can be written in a $N \times K$ matrix

$$C = (\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_K).$$

The shrinkage removes high frequency EEG while some low frequency EEG remains because the wavelet shrinkage is not performed for the coefficients c_k^0 of the ERP trend. In addition, the high frequency EEG can not be completely removed by the wavelet shrinkage. The remaining EEG has to be removed by utilizing cross trial properties.

The ERPs change from trial to trial, and they changes slowly across trials in both latency and amplitude. This means that the wavelet coefficients of the ERPs at a given scale level and a given time also changes slowly from trial to trial. This suggests the following linear modeling for extracting slowly changing ERPs.

Let $z_i = (z_i(1), \cdots, z_i(K))$ be the i^{th} row of C. Since the ERPs change slowly from trial to trial, the wavelet coefficients of any few consecutive trials can be approximated by a linear fit. That is,

$$z_i(j + d) \approx a + bd, \qquad |d| \leq h$$

for two constants a and b, where h is a bandwidth.

For any j and a given bandwidth h, solve the minimization problem

$$\min_{a,b} \sum_{k=1}^{K} Q(j, k, h)[z_i(k) - (a + b(k - j))]^2,$$

where Q is a weight function assigning weight of fitting at each point. We will take

$$Q(j, k, h) = \max\left(0, 1 - \frac{(j-k)^2}{h^2}\right)$$

which assigns higher weight if $|k-j|$ is small, smaller weight if $|k-j|$ is larger, and zero weight if $|k-j| \geq h$. Let $(\hat{a}, \hat{b})$ be the minimizer of the above minimization problem, then the wavelet coefficient of trial j at the given frequency level and the given time is estimated by

$$\hat{z}_i(j) = \hat{a}.$$

Repeat this linear fitting procedure for all rows of C and let

$$\hat{C} = \begin{pmatrix} \hat{z}_1 \\ \vdots \\ \hat{z}_N \end{pmatrix}.$$

Then $\hat{C}$ is the estimated wavelet coefficients of the ERPs from K trials. Computationally, this is obtained by

$$\hat{C} = CS_h$$

where S_h is the smoothing matrix depending on the bandwidth h.

If one performs such smoothing across trials without doing wavelet shrinkage first, the poor SNR of the recorded signals requires a large bandwidth in order to remove noise. This usually leads to the conventional averaging discussed in the introduction.

Step 4: Inverse wavelet transform

Recall that W is orthogonal. Then

$$\hat{P} = (\hat{p}_1, \cdots, \hat{p}_K) = W'\hat{C}$$

is the estimated ERPs from K trials and $\hat{n}_j = s_j - \hat{p}_j$, $j = 1, \cdots, K$, is the estimated noise of trial j.

3. Experiment

We apply the proposed method to the ERP data set plotted in Figure 2. The wavelet 'db8' (see [9] for details of the wavelet) and bandwidth $h = 3.5$ are used based on our simulations. The results are plotted in Figure 3.

The estimated SNR for this data set is 0.32, typical in evoked potential recording. The conventional average of the data is still quite noisy and the background EEG is not cleared (see Figure 3). Figure 4 plots several cross correlation functions, where prestimulus signals are the recordings of the 200 msec signals before the 'X' appears on screen.

Figure 4 shows that the prestimulus signal and the estimated ERPs are almost uncorrelated. It also shows that the cross correlation between the prestimulus signal and the data is almost identical with

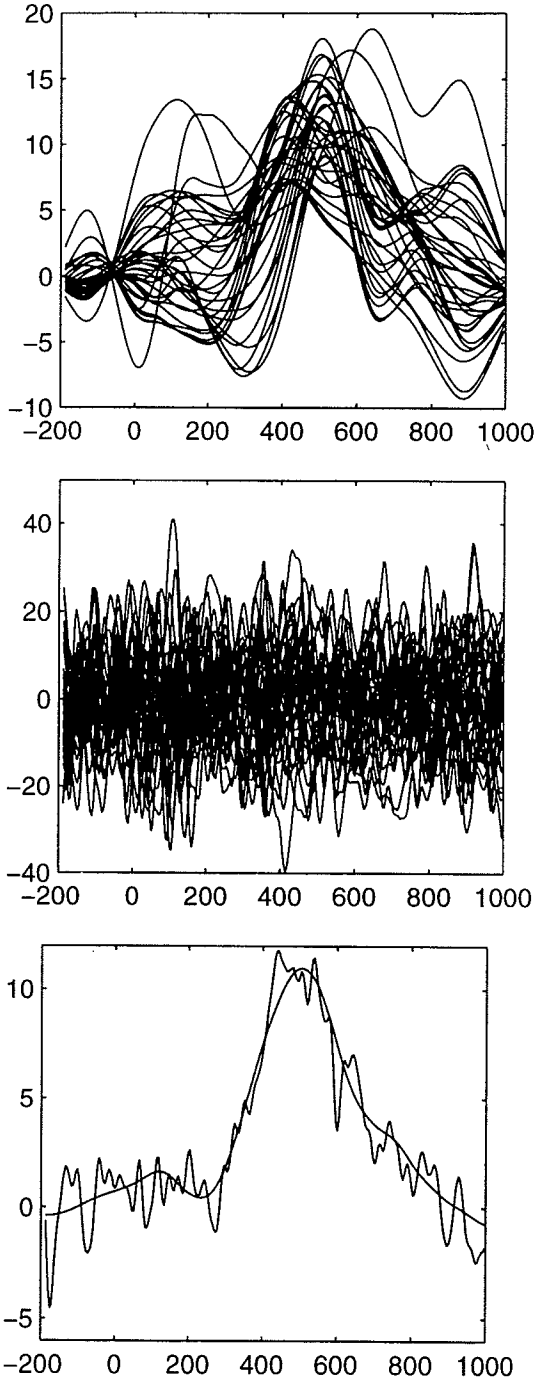

Figure 3: Top – the estimated ERPs from the 35 trials plotted in Figure 2; Middle – the estimated noise or EEG; Bottom – the average of the data (rough line) and the average of the estimated ERPs (smooth line).

the cross correlation between the prestimulus signal and the estimated noise. Therefore, the correlation analysis suggests that the estimated ERPs are the evoked brain potentials due to the application of the stimuli.

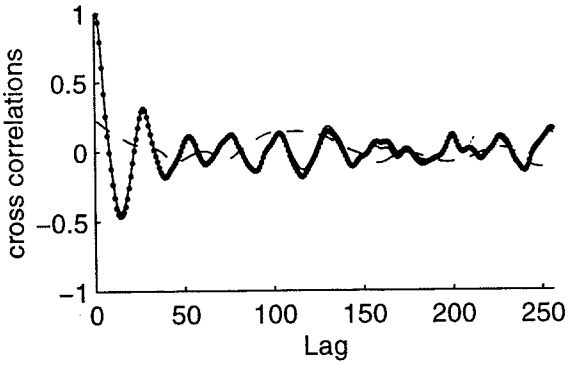

Figure 4: The cross correlations between prestimulus signal and the data (solid line), between prestimulus signal and the estimated ERPs (dashed line), and between prestimulus signal and the estimated noise (dotted line).

References

[1] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrica*, vol. 81, pp. 425-455, 1994.

[2] M. Frazier, B. Jawerth, and G. Weiss, *Littlewood-Paley theory and the study of function spaces*, NSF-CBMS Regional Conference Series in Mathematics 79, Providence, RI: American Mathematical Society, 1991.

[3] A.S. Gevins and A. Rémond, *Methods of Analysis of Brain Electrical and Magnetic Signals*, Elsevier, 1987.

[4] M. Hansson, T. Gänser, and G. Salomonsson, "Estimation of single event-related potentials utilizing the Prony method," *IEEE Trans. Biomed. Eng.*, vol. 43, pp. 973-981, 1996.

[5] B.H. Jansen, J.R. Bourne, and J.W. Ward, "Autoregressive estimation of short segment spectra for computerized EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 9, pp. 630-638, Sept. 1981.

[6] A. Isaksson, A. Wennberg, and L.H. Zetterberg, "Computer analysis of EEG signals with parametric models," *Proc. IEEE*, vol. 69, pp. 451-461, Apr. 1981.

[7] N. Laskaris, S. Fotopoulus, P. Papathanasopoulos, and A. Bezerianos, "Robust moving averages, with Hopfield neural network implementation, for monitoring evoked potential signals," *Evoked Potentials*, vol. 104, pp. 151-156, 1997.

[8] Y. Meyer, *Wavelets Algorithms & Applications*, SIAM, Philadelphia, 1993.

[9] M. Misiti, Y. Misiti, G. Oppenheim, and J.M. Poggi, *Wavelet TOOLBOX*, The MathWorks, Inc., 1996.

[10] R. Nätäänen, "The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive functions," *Behavioral and Brain Sciences*, vol. 13, pp. 201-288, 1990.

[11] J. Raz, B. Turetsky, and G. Fein, "Selecting the smoothing parameter for estimation of slowly changing evoked potential signals," *Biometrics*, vol. 45, pp. 745-762, Sept. 1989.

[12] V. Samar, K. Swartz, and M. Raghuveer, "Multiresolution analysis of event-related potentials by wavelet decomposition," *Brain and Cognition*, vol. 27, pp. 398-438, 1995.

[13] J. Sgro, R. Emerson, and T. Pedley, "Real-time reconstruction of evoked potentials using a new two-dimensional filter method," *Electroenceph. and Clin. Neurophysiol.*, vol. 62, pp. 372-380, 1985.

[14] W. Wang, H. Begleiter, and B. Porjesz, "Multivariate composite estimators: new ways to track signals with applications to human cerebral potentials," *Brain Topography*, vol. 7, pp. 9-15, 1994.

[15] I. Johnstone and B. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Statist. Soc. B*, vol. 59, pp. 319-351, 1997.