SINGLE TRIAL ESTIMATION OF EVENT-RELATED POTENTIALS USING PARTICLE FILTERING

Hamid R. Mohseni, Edward L. Wilding and Saeid Sanei

Schools of Engineering and Psychology
Cardiff University, Wales, UK
{mohsenihr,wildinge,saneis}@cf.ac.uk

ABSTRACT

In this paper a method based on wavelet transform (WT) and particle filtering (PF) for estimation of single trial event-related potentials (ERPs) is presented. The method is based on recursive Bayesian mean square estimation of wavelet coefficients of the ERPs, using PF as the estimator. Simulation results are provided to demonstrate the superior performance of PF over Kalman filtering (KF) for non-Gaussian and non-linear electroencephalography (EEG) signals. The methods were also applied to the real data in an *odd-ball* paradigm to explore the changes in the P300 component from trial to trial.

***Index Terms*—** Event related-potentials, single trial estimation, P300, particle filtering, Kalman filtering.

1. INTRODUCTION

Event related-potentials (ERPs) are the activity of the brain as the response to a kind of stimuli and can provide records of brain activity at any reasonable scale of temporal resolution [1]. Conventional methods for analyzing ERPs involve averaging time-locked segments of the EEG signal over many trials to obtain the ERP waveform. These methods assume that the statistical parameters of the ERP waves are constant over time and the background EEG is a random process that is attenuated by averaging over trials. This is acceptable in some cases. However, there is evidence that ERP waves may vary considerably over time [2] due to changes in the degree of fatigue, habituation, or level of attention of the subject. Therefore, a method for estimation of single trials to investigate the variability of ERPs from trial to trial is desirable.

Statistical signal processing methods including maximum a posterior (MAP) solutions [3] and Wiener filtering [4] are the major approaches that have been widely used in ERP single trial estimation. In [5] maximum likelihood (ML) is formulated yielding the estimators of amplitude and latency jitter of single trials. A Bayesian approach for construction of a single-trial estimator for the ERPs by a subspace regularization method has been proposed in [6]. In this approach the second-order statistical information extracted from a set of measured potentials was used to represent the prior information for the estimation. An extension of this single-channel subspace regularization method to multi-channel is presented in [7].

Another promising method in statistical signal processing, Kalman filtering (KF), has been developed and used for the separation of each single measured response (trial) into background activity and ERP parts [8]. KF has been also employed for estimation of the dynamic changes of amplitude and latency of ERPs by considering the entire epoch of the ERP trials as the input to the KF [9]. Unfortunately, these techniques fail to estimate single trial ERP in many cases, because of the very low ERP signal to background noise power ratio, the non-Gaussian and stochastic nature of the EEG, and the inter-trial variability of the recorded ERPs.

In this study we present an algorithm for estimation of single trial ERPs, where particle filtering (PF) is an estimator for coefficients of discrete wavelet transform (DWT). Based on the concept of *sequential importance sampling* and the use of Bayesian theory, PF is particularly useful in dealing with nonlinear and non-Gaussian problems [10]. Combining DWT and PF is an effective method for denoising a non-stationary signal such as the ERP in a non-linear and non-Gaussian EEG environment.

2. METHODS

2.1. Problem Formulation in State Space

The M wavelet coefficients of time locked ERP signals in the k_{th} trial are formulated as

$$\mathbf{y}_k = [\ y_k(1) \quad y_k(2) \quad \dots \quad y_k(M)\ ]^T \tag{1}$$

where $[.]^T$ indicates the transpose operation. By modeling the wavelet coefficients in the state space, the evolution of the state $\{\mathbf{x}_k, k \in \mathbb{N}\}$, and the relation between the state and the estimated wavelet coefficients (the measurement equation) are respectively given by

$$\mathbf{x}_k = f_{k-1}(\mathbf{x}_{k-1}, \mathbf{w}_{k-1}) \tag{2}$$

This work has been supported by the School of Engineering and Psychology, Cardiff University, Wales UK.

$$\mathbf{y}_k = h_k(\mathbf{x}_k, \mathbf{v}_k) \tag{3}$$

where f_k and h_k are generally nonlinear functions of the state $\mathbf{x}_k$, and $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ are i.i.d. noise processes. We search for the filtered estimates of $\mathbf{x}_k$ based on a set of all available wavelet coefficients $\mathbf{y}_{1:k} = \{\mathbf{y}_i, i = 1, \ldots, k\}$, up to the k_{th} trial.

By recursive calculation of the posterior density function $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ of state $\mathbf{x}_k$ at trial k, the estimation of state $\mathbf{x}_k$ can be the expected value of its posterior density at trial k. Via Bayes rule, an available measurement $\mathbf{y}_k$ at time k is used to update the posterior density

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} \tag{4}$$

where $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ is computed in the prediction stage using the Chapman-Kolmogorov equation and $p(\mathbf{y}_k|\mathbf{y}_{1:k-1})$ is a normalizing constant.

Kalman and particle filtering are two major approaches for solving the recursive equations (4). In the following sections these algorithms are summarized.

2.2. Kalman Filter

For a Kalman filter if f_k and h_k assumed to be known and linear functions and random sequences $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ be mutually independent zero mean white Gaussian noise with known covariances, the posterior density becomes Gaussian and one can evaluate only the mean and the covariance matrices in closed form.

In the linear Gaussian environment where the (states' distribution is Gaussian) Kalman filtering is an optimal solution. This cannot be achieved using other recursive algorithms. Also, considering any other distribution for states, the Kalman filter is the best linear estimator in the presence of Gaussian white noise.

2.3. Particle Filter

In PF, the posterior distributions are approximated by discrete random measures defined by particles $\{\mathbf{x}^{(n)}, n = 1, \ldots, N\}$ and their associated weights $\{w^{(n)}, n = 1, \ldots, N\}$. The distribution based on these samples and weights at the k_{th} trial is approximated as

$$p(\mathbf{x}) \approx \sum_{n=1}^{N} w^{(n)} \delta(\mathbf{x} - \mathbf{x}^{(n)}) \tag{5}$$

where δ is the Dirac delta function and $(.)^n$ refers to the n_{th} weight.

If the particles are generated according to the distribution $p(\mathbf{x})$, the weights are equal and will be $1/N$. When generation of the particle by direct sampling from unknown distribution $p(\mathbf{x})$ is impossible, the particles are generated from a known distribution $\pi(\mathbf{x})$ called importance density. This concept, known as importance sampling, results in the following weights [10]:

$$w^{(n)} \propto \frac{p(\mathbf{x}^{(n)})}{\pi(\mathbf{x}^{(n)})} \tag{6}$$

Suppose at k_{th} trial we want to approximate the posterior distribution $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ subject to having $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$. If the importance density is chosen such that it can be factorized to

$$\pi(\mathbf{x}_k|\mathbf{y}_{1:k}) = \pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_{1:k})\pi(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \tag{7}$$

then the new samples $\mathbf{x}_k^{(n)}$ can be obtained according to the importance density $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(n)}, \mathbf{y}_k)$ which depends on the old samples and the new measurements.

Using Bayes' rule (4) and equations (6 and 7), the new weights are updated as follows [10]:

$$w_k^{(n)} \propto w_{k-1}^{(n)} \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(n)})p(\mathbf{x}_k^{(n)}|\mathbf{x}_{k-1}^{(n)})}{\pi(\mathbf{x}_k^{(n)}|\mathbf{x}_{k-1}^{(n)}, \mathbf{y}_{1:k})} \tag{8}$$

The choice of importance density is one of the most crucial issues in the design of PF and plays a significant role in its performance. This function must have the same support as the probability distribution to be approximated. In general, the closer the importance function to the distribution, the better the approximation is. The most popular choice for the prior importance function, also used in this paper, is given by

$$\pi(\mathbf{x}_k|\mathbf{x}_{k-1}^{(n)}, \mathbf{y}_{1:k}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(n)}) \tag{9}$$

This choice of importance density implies that we need to sample $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(n)})$. A sample can be obtained by generating a noise sample $\mathbf{w}_{k-1}^{(n)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_w)$ and setting $\mathbf{x}_k^{(n)} = f_{k-1}(\mathbf{x}_{k-1}^{(n)}, \mathbf{w}_{k-1}^{(n)})$. Also, it implies that particle weights from equation (8) can be updated by

$$w_k^{(n)} \propto w_{k-1}^{(n)} p(\mathbf{y}_k|\mathbf{x}_k^{(n)}) \tag{10}$$

The importance sampling weights indicate the level of importance of the corresponding particle. A relatively small weight implies that the sample is drawn far from the main body of the posterior distribution and has a small contribution in the final estimation. Such a particle is said to be ineffective. If the number of ineffective particles is increased, the number of particles contributing to the estimation of states is decreased, so the performance of the filtering procedure deteriorates. The degeneracy can be avoided by a resampling procedure. Resampling is a scheme that eliminates the particles with small weights and replicates those with large weights according to their weights.

3. EXPERIMENTAL RESULTS

3.1. Simulated Data

To generate simulated data, two Gaussian functions were used resembling P300 and N400 components. P300 is a positive wave which occurs with a latency of approximately 300 ms after rare or task relevant stimuli and N400 is a negative wave which occurs approximately 400 ms after stimuli. The amplitudes and latencies of the first peak were assumed to change sinusoidally from trial to trial by adding Gaussian white noise (GWN). The amplitudes and latencies of the second peak were assumed to change according to a uniform random distribution, during trials.

For better assessment of the proposed algorithms, GWN and real background EEG activity were considered as two different kinds of noise. Therefore the signal-to-noise ratio (SNR) and the signal-to-background ratio (SBR) can be defined as follows:

$$SNR = 10\log[\frac{P(signal)}{P(GWN)}] \tag{11}$$

$$SBR = 10\log[\frac{P(signal)}{P(background)}] \tag{12}$$

where $P(.)$ indicates the power of the signal. SNR represents the amount of GWN and both KF and PF are trying to suppress this kind of noise, theoretically. SBR is a measure of background activity of the brain which is a nonlinear and non-stationary noise and is of main interest in this study.

In all the algorithms, the f_{k-1} and g_k functions in equations (2) and (3) are assumed to be identity functions and the covariances of $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ in KF were assumed to be $\mathbf{Q}_w = \sigma_w\mathbf{I}$ and $\mathbf{Q}_v = \sigma_v\mathbf{I}$ respectively. Whereas, in the PF the covariance of noise matrices were assumed to be $\mathbf{Q}_v = q_v\mathbf{I}$ and $\mathbf{Q}_w = q_w\mathbf{I}$ respectively, where σ_v, σ_w, q_v, q_w were known and constant parameters, and $\mathbf{I}$ is the identity matrix. In the KF, only the σ_v/σ_w ratio is important so $\sigma_v = 1$ was chosen for convenience. In the PF q_v and q_w play different roles and a proper combination of them can lead to better results, but here we fixed $q_v = 5$ and only q_w was adjusted.

Fig. 1 shows the output SNR and SBR vs. input SNR and SBR in dB. They are obtained by finding the best σ_w and q_w parameters in each SNR and SBR. To tune the parameters, each parameters was increase by the step of 3 units and the algorithms were running 100 Monte Carlo trials in each step, and the best parameters among them were selected. Using wavelet coupled with PF or KF, as two stages of denoising, causes all algorithms to have proper performance in high input SNR. Comparing the SBR for the PF and KF shows that they have the same performance in the high input SBR, but when the amount of background EEG increases and the noise becomes more non-Gaussian, performance of the PF improves accordingly. In this linear estimation, if the added noise is GWN, the KF is the best estimator and the superior performance of PF over KF in Fig. 1 is the result of added non-Gaussian background noise.

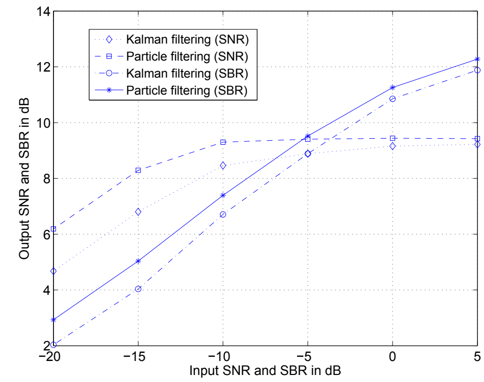

Fig. 1. Output SNR and SBR vs. input SNR and SBR in the simulated data.

3.2. Real Data

Real data was obtained using an odd-ball paradigm in the Cognitive Electrophysiology Laboratory, School of Psychology, Cardiff University. Subjects heard in total 300 tones, 240 of which were frequent (80%) and 60 of which were infrequent (20%). The frequency bandwidth of the linear bandpass filter was set to 0.03-40 Hz and the sampling rate was 250 Hz. A Fz reference was employed during acquisition, and the data were re-referenced off-line to the average of the left and right mastoids.

Epochs from 200ms to 500ms time-locked to stimulus onset were extracted. A 150ms pre-stimulus interval was used for baseline correction. Since the estimation of first trials is important, the initial filter values should be selected appropriately. Each algorithm was run two times: in the first run the initial filter values were set to zero and in the second run they were set to the last estimated trial of the first run.

Fig. 2 shows the results of the proposed algorithms for real data. The Cz site, at which the P300 component amplitude in the odd-ball paradigm is prominent, was chosen for analysis. Fig. 2 (a) shows all the epochs of the original data and their mean signal. The epoch signals are also presented in Fig. 2 (b) in the form of colored images, in which epochs are plotted vertically with time on the horizontal axis and color represents the epoch amplitudes. These images can be used to visualize variability in the amplitudes and latencies of ERPs. Fig. 2 ((c), (d)) shows estimated ERPs with KF and PF, and ((e), (f)) shows ERP images for estimated data using KF and PF. The estimated amplitudes of the P300 waves for consecutive trials were obtained by finding the maximum value and the results are shown in Fig. 2 ((g), (h)).

In previous studies, there is evident for decreases in P300 amplitude over trials in the odd-ball paradigm [11]. A decrease in the P300 amplitude can be seen for the results of

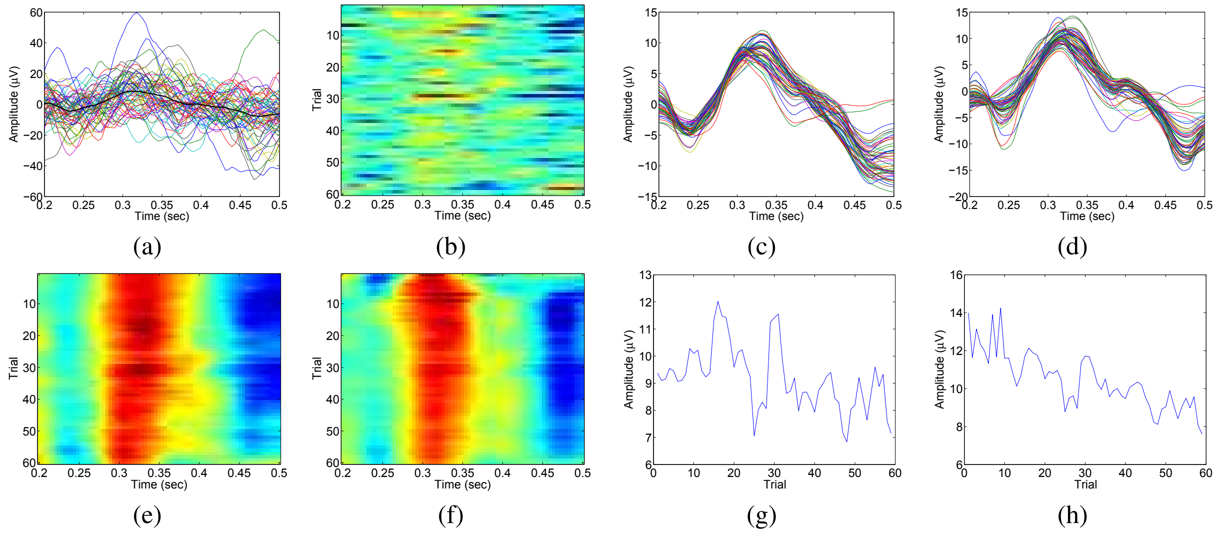

Fig. 2. Results for real data. (a) Data epochs with their average (thick line), (b) ERP image for real data, ((c), (d)) Estimated ERPs with KF and PF, ((e), (f)) ERP image for estimated data using KF and PF, ((g), (h)) amplitudes of the P300 waves for consecutive trials extracted by KF and PF.

the PF (Fig. 2 (h)) but not for the KF method (Fig. 2 (g)). The changes in amplitude has a constant rate over the 60 infrequent trials for this subject. Also a small increase in the latency can be seen in the ERP images.

4. CONCLUSIONS

We have proposed a new method for single-trial estimation of ERPs. The method is based on discrete wavelet transform and particle filtering. The main benefits of the proposed method are application of wavelet-based PF together with the use of sequential importance sampling concept, as well as Bayesian theory and its accurate performance in non-Gaussian and nonlinear data like EEG signals. The effectiveness of the methods is illustrated for both simulated and experimental data. As a specific application, the method was applied to the estimation of the P300 component. These observation demonstrate the potential for this approach to single trial analysis of EEG and ERP data.

5. REFERENCES

[1] M. D. Rugg and M. G. H. Coles, *Electrophysiology of mind*, Oxford Science publications, 1995.

[2] M. L. A. Jongsma, T. Eichele, C. M. Van Rijn, A. M. L. Coenen, K. Hugdahl, H. Nordby, and R. Q. Quiroga, "Tracking pattern learning with single-trial event-related potentials," *Clinical Neurophysiology*, vol. 117, pp. 1957–1973, 2006.

[3] W. A. Truccolo, K. H. Knuth, A. Shah, S. L. Bressler, C. E. Schroeder, and M. Ding, "Estimation of single-trial multicomponent erps: Differentially variable component analysis (dvca)," *Biological Cybernetics*, vol. 89, pp. 426–438, 2003.

[4] S. Cerutti, V. Bersani, A. Carrara, and D. Liberati, "Analysis of visual evoked potentials through wiener filtering applied to a small number of sweeps," *J. Biomed. Eng.*, vol. 9, pp. 3–12, 1983.

[5] P. Jaskowski and R. Verleger, "Amplitudes and latencies of single trial erps estimated by a maximum likelihood method," *IEEE Trans. Biomed. Eng*, vol. 46, pp. 987–993, 1999.

[6] P. Karjalainen, J. Kaipio, A. Koistinen, and M. Vauhkonen, "Subspace regularization method for the single trial estimation of evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 849–860, 1999.

[7] P. O. Ranta-aho, A. S. Koistinen, J. O. Ollikainen, J. P. Kaipio, J. Partanen, and P. A. Karjalainen, "Single-trial estimation of multichannel evoked-potential measurements," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 2, pp. 189–196, 2003.

[8] M. V. Spreckelsen and B. Bromm, "Estimation of single-evoked cerebral potentials by means of parametric modeling and kalman filtering," *IEEE Trans. Biomed. Eng*, vol. 35, no. 9, pp. 691–700, 1988.

[9] S. D. Georgiadis, P. O. Ranta-aho, M. P. Tarvainen, and P. A. Karjalainen, "Single-trial dynamical estimation of event-related potentials: a kalman filter-based approach," *IEEE Trans. Biomed. Eng*, vol. 52, no. 8, pp. 1397–1406, 2005.

[10] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, 2002.

[11] J. Pan, T. Takeshita, and K. Morimoto, "P300 habituation from auditory single-stimulus and oddball paradigms," *International Journal of Psychophysiology*, vol. 37, pp. 149–153, 2000.