Abstract—A method for single-trial dynamical estimation of event related potentials (ERPs) is presented. The method is based on recursive Bayesian mean square estimation and the estimators are obtained with a Kalman filtering procedure. We especially focus on the case that previous trials contain prior information of relevance to the trial being analyzed. The potentials are estimated sequentially using the previous estimates as prior information. The performance of the method is evaluated with simulations and with real P300 responses measured using auditory stimuli. Our approach is shown to have excellent capability of estimating dynamic changes form stimulus to stimulus present in the parameters of the ERPs, even in poor signal to noise ratio conditions.

Index Terms—Event related potentials, single-trial dynamical estimation, recursive Bayesian mean square, Kalman filter.


I. INTRODUCTION

Event related potentials (ERPs) are voltage changes of brain electric activity due to stimulation. The measured responses are often considered as the combination of electric activity resulted by multiple brain generators, active in association with the eliciting event, and noises which are brain activity not related to the stimulus, together with interference from non neural sources, like eye blinks and other artifacts. Usually ERPs are considered as transient-like smooth waveforms. Even though they are dominated by lower frequencies, compared to background electroencephalogram (EEG), due to poor signal to noise ratio conditions their form is difficult to be estimated in a trial to trial scheme.

The simplest model for ERPs consists of the sum of an invariant signal and random noise. Thus, the most common way to measure their parameters is averaging time locked single-trial measurements. This assumption may be a reasonable one for some but not all cases. For example, changes in degree of fatigue, habituation, or level of attention of the subject can affect the ERPs. Therefore, averaging implies a loss of information related to trial to trial variability. The fact that the same event can elicit somewhat different signals has been evident for a few decades [1] and, thus, the analysis of ERPs is focused on single-trial based estimation. In that sense, the investigation of the variability of the ERP parameters reveals more information related to changes of the cognitive state.

Several techniques have been proposed in order to denoise and estimate ERPs. In general, the performance of the estimator or the filter is naturally dependent on the prior information related to the statistical properties of the signal and the background noise. Digital filtering is sometimes used for single responses [2], [3]. The simplest filters are moving average FIR filters. The main problems in linear time invariant filtering are that usually the spectra of ERPs and background noise overlap heavily and that since the evoked potential is a transient-like smooth waveform with no periodicity its spectrum is not defined properly. Wiener filtering is also possible for single measurements with some specific structure of the filter, which provides an optimal filtering in the mean square error sense [3], [4]. However, these approaches have the drawback of considering the ERP signal as a stationary process, but since the ERPs are compositions of transient responses with different time and frequency localizations, they are not likely to give optimal results.

Other approaches involve time-varying filtering of the single-trials and in particular they use techniques based on Wiener formalism [5]. A crucial problem in time-varying mean square error filtering is to obtain a good model for the cross covariance between the ERP and the measurement. This is a difficult estimation task, especially when there exists a significant level of correlation between the hypothetical ERP and the noise. Later methods include, for example, time-frequency decomposition of the signal based on wavelet transform [6] and regularization based methods [7]. Different estimation methods also exist for the case that multiple channels are used in the analysis. Some of the most promising are probably methods based on independent component analysis [8] or methods based on subspace regularization [9]. However, in this study, we are mainly concentrated in single channel estimation techniques.

Usually the measurements can be thought to be sampled from the same joint probability density and the set of measurements is assumed to be homogeneous. Then all measurements can be taken into account with equal weights in estimation. Of special interest is the case when some parameter of the ERP changes dynamically from stimulus to stimulus. This kind of situation can be, for example, a trend such as a change of the latency of some specific component of the ERP. In that situation, the joint density has some parameters with trendlike variations during the test and, therefore, the order that the ERPs occur contains additional information to be used in the estimation. The approaches mentioned above do not take into account such temporal information. A dynamical estimation method in which preceding or following estimates are used in
the estimation procedure can give additional information about the ERPs and improve the trial to trial estimation.

In this paper, we assume that the ERP is a vector valued random process with slow dynamic variability during the repetitions of the test. Then past realizations contain information of relevance to future realizations and these changes can be modeled with a state-space model. The recursive mean square estimate for the state is then given by Kalman filter. A brief idea of this approach was introduced in [10] and further discussed in [11], together with other systematic procedures for ERPs estimation. In this paper, we present a denoising method for ERPs based on Kalman filtering and generic observation models. We also discuss the connection of recursive mean square estimation with time-varying Wiener filtering, as well as with other dynamical estimation methods. We evaluate the performance of the presented approach to produce reliable estimates for single-trials with simulations and with real P300 responses. Recursive estimation by Kalman filtering has many theoretical benefits compared to other dynamical methods and shows to have exceptional capability of estimating dynamic changes present in the measurements.

II. METHODS

In this section, we first give an overview on time varying estimation of ERPs. Then the recursive mean square estimator with Kalman filter is obtained and its connection with time varying Wiener filtering is shown. Finally the connection of Kalman filter with other dynamical estimation procedures is presented and its applicability to single-trial estimation is discussed.

A. Time varying estimation of ERPs with linear observation model

We use the following notation for ERP measurements. The sampled potential after the $t$th stimulus is denoted with a column vector of length $M$

$$ z_t = \begin{pmatrix} z_t(1) \\ z_t(2) \\ \vdots \\ z_t(M) \end{pmatrix}. $$

(1)

Using these measurements we want to estimate a vector of random parameters $\theta_t$, by restricting the estimator to be a linear mapping from $z_t$ to $\theta_t$. The estimator that minimizes the mean square Bayes cost $B_{MS} = E\{||\theta_t - \hat{\theta}_t||^2\}$, is $\hat{\theta}_t = E\{\theta_t|z_t\}$ and can be shown to be

$$ \hat{\theta}_t = \eta_0 + C_{\theta_t,z_t}^{-1}(z_t - \eta_z), $$

(2)

where $\eta_z$ is the mean and $C_{\theta_t,z_t}$ the covariance of the column vector $z_t$ of the measurements, $\eta_0$ is the mean of the parameter vector, and $C_{\theta_t,z_t}$ is the cross-covariance of the measurements and the parameters to be estimated. The estimator is optimal among all possible estimators, not only linear, if $\theta_t$ and $z_t$ are jointly Gaussian [12]. This form is independent of the relation between the observations $z_t$ and parameters $\theta_t$ and the estimation error $\hat{\theta}_t = \theta_t - \hat{\theta}_t$ has covariance $C_{\hat{\theta}_t}$.

$$ C_{\hat{\theta}_t} = C_{\theta_t} - C_{\theta_t,z_t}C_{z_t}^{-1}C_{\theta_t,z_t}. $$

(3)

In practice, we can not calculate $C_{\theta_t,z_t}$ without assuming some model for the observations or without some prior knowledge about the ERPs. A widely used model for ERP estimation is the linear additive noise model. The observations are then assumed to be of the form

$$ z_t = s_t + v_t. $$

(4)

The response $s_t$ corresponds to the part of the activity that is related to the stimulus and the rest of the activity $v_t$ is usually assumed to be independent of the stimulus and the ERP. With that model the ERP $s_t$ equals the parameter vector $\theta_t$ and equation (2) is the time-varying Wiener filter used, e.g., in [5].

The ERPs can be further modeled as a linear combination of some pre-selected basis vectors. Then, the observation model takes the form

$$ z_t = H_t\theta_t + v_t, $$

(5)

where $H_t$ is a deterministic observation matrix, which contains the basis vectors $\psi_{t,1}, \ldots, \psi_{t,k}$ of length $M$ in its columns, and $\theta_t$ is a parameter vector of length $k$. The estimated ERP $\hat{s}_t$ can then be obtained by using the estimated parameters $\hat{\theta}_t$

$$ \hat{s}_t = H_t\hat{\theta}_t. $$

(6)

Now, the linear mean square estimator with observation model $H_t$, in the special case that $\theta_t$ and $v_t$ are uncorrelated, i.e. $C_{\theta_t,v_t} = 0$, can be written in the form [12]

$$ \hat{\theta}_t = (H_t^T C_{v_t}^{-1} H_t + C_{\theta_t}^{-1})^{-1}(H_t^T C_{v_t}^{-1} z_t + C_{\theta_t}^{-1} \eta_\theta), $$

(7)

where $C_{\theta_t}$ and $\eta_\theta$ are the covariance and the mean of $\theta_t$, respectively. $C_{v_t}$ is the covariance of the zero mean measurement noise and $(\cdot)^T$ denotes transpose. Again this estimator is optimal among all possible estimators, not only linear, if $\theta_t$ and $z_t$ are jointly Gaussian, but with the requirement of uncorrelated parameters and noise. In Bayesian estimation this is also called the maximum a posteriori estimator (MAP) and $C_{\theta_t}$ and $\eta_\theta$ represent the prior information about the parameters $\theta_t$. If they are not available, we can assume $C_{\theta_t}^{-1} = 0$ corresponding to infinite prior variance for the parameters. In this case, the estimator reduces to the ordinary minimum variance Gauss-Markov estimator. If we assume that the errors are independent with equal variances $C_{v_t} = \sigma_v^2 I$, the estimator is identical to the ordinary least squares estimator

$$ \hat{\theta}_t = (H_t^T H_t)^{-1} H_t^T z_t. $$

(8)

In general, most of the methods for single-trial ERP analyses and estimation try to decompose the measurements into relevant components or to explain the data through some parameters. These parameterizations give the necessary means to investigate, for example, the properties of the changes that the stimulus caused to the ongoing EEG signal or that the repetition of the test caused from trial to trial to the responses. Most of the methods are based on an explicit model or on
some specific assumptions for the EEG. Every decomposition is then involved in at least two main considerations. On the one hand, if the resulting estimates follow too closely the measurements, it is probable that some features of the data are still going to be hidden by phenomena unrelated to the stimuli. On the other hand, if the estimates are not following closely the measurements, some features probably have been neglected and lost or some extra features have been added by the model itself. Usually a balance between these considerations is made and care is given to the correct interpretation of a parametrization that is able to reveal and investigate some specific features related to the experiment.

Several parameterizations have been used for single-trial estimation of ERPs in different studies. Common ones are based on different types of Gaussian shaped components and exponentially damped sinusoidal functions [13]. Other choices are Fourier basis or several possibilities of wavelet basis. While these choices are not based on the measurements it is possible to form decompositions based on the data. For example, it is possible to use the $k$ first eigenvectors of the data correlation matrix as a basis, and then the least squares solution is equivalent to the so-called principal component regression approach [14].

Here we focus on the case that dynamic changes exist from stimulus to stimulus. Although some of the methods that are briefly mentioned here and in the introduction could be used to estimate dynamically changing features, the possibility that previous trials and estimates can contain information of relevance to the next trials and estimates is not taken explicitly into account in the estimation procedure. One way to model such information is to use of a state-space representation for the signals and then estimate the states with recursive mean square estimation and Kalman filter.

B. Recursive mean square estimation and Kalman filter

Among many other forms the state-space model for linear dynamical systems can be written as follows. Let $z_t \in \mathbb{R}^M$ and $\theta_t \in \mathbb{R}^k$ be vector-valued processes. The state $\theta_t$ evolves according to the linear difference equation

$$\theta_{t+1} = F_t \theta_t + G_t \omega_t$$  \hspace{1cm} (9)

with some initial distribution for $\theta_0$. The state is not observed directly, but instead, the measurements $z_t$ are available at discrete sampling times and are described as

$$z_t = H_t \theta_t + \nu_t.$$  \hspace{1cm} (10)

This is clearly a linear observation model. The other assumptions for the model are as follows

- $F_t$, $G_t$, and $H_t$ are known sequences of matrices,
- $(\theta_0, \omega_t, \nu_t)$ is a sequence of mutually uncorrelated random vectors with finite variance,
- $E\{\omega_t\} = 0$ and $E\{\nu_t\} = 0$ for every $t$,
- the covariances $C_{\omega_t}$, $C_{\nu_t}$, and $C_{\omega_t \nu_t}$ are known sequences of matrices.

The Kalman filtering problem is now to find the linear mean square estimator $\hat{\theta}_t$ for state $\theta_t$ given the observations $z_1, \ldots, z_t$. This is equal to the conditional mean [15]

$$\hat{\theta}_t = E\{\theta_t | z_1, \ldots, z_t\}. \hspace{1cm} (11)$$

For the derivation of Kalman filter it is required that $\theta_t$, $\nu_t$, $\omega_t$ are uncorrelated and that $C_{\nu_t \nu_i} = 0$, $C_{\omega_t \omega_j} = 0$ for every $i \neq j$. Based on these assumptions there are two approaches to obtain the mean square estimator. The first is to specify a linear conditional mean and find the best linear form. The second is to assume that $\nu_t$ and $\omega_t$ are Gaussian. In this case the conditional mean is again linear. The results of these two approaches are identical and the required estimator can be shown to be equal to

$$\hat{\theta}_t = (H_t^T C_{\nu_t}^{-1} H_t + C_{\theta_{t_{t-1}}}^{-1})^{-1} (H_t^T C_{\nu_t}^{-1} z_t + C_{\theta_{t_{t-1}}}^{-1} \hat{\theta}_{t-1}), \hspace{1cm} (12)$$

where $\hat{\theta}_{t_{t-1}} = E\{\theta_t | z_{t-1}, \ldots, z_1\}$ is the prediction of $\theta_t$ based on $\hat{\theta}_{t-1}$ and $\hat{\theta}_{t_{t-1}} = E\{\theta_{t_{t-1}} | z_{t-1}, \ldots, z_1\}$ is the optimal MS estimate at time $t-1$. Clearly this estimator is of the form (7), which is the Bayesian MAP estimate using the last available estimate as prior information. In other words, the Kalman filter is the best sequential estimator if the Gaussian assumption is valid and it is the best linear estimator whatever the distributions are. After adding the initializations we can summarize the Kalman filter algorithm

$$C_{\theta_0} = C_{\theta_0} \hspace{1cm} (13)$$

$$\hat{\theta}_0 = E\{\theta_0\} \hspace{1cm} (14)$$

$$\hat{\theta}_{t_{t-1}} = F_{t-1} \hat{\theta}_{t-1} \hspace{1cm} (15)$$

$$C_{\theta_{t_{t-1}}} = F_{t-1} C_{\theta_{t-1}} F_{t-1}^T + G_{t-1} C_{\omega_{t-1}} G_{t-1}^T \hspace{1cm} (16)$$

$$K_t = C_{\theta_{t_{t-1}}} H_t^T (H_t C_{\theta_{t_{t-1}}} H_t^T + C_{\nu_t})^{-1} \hspace{1cm} (17)$$

$$C_{\theta_t} = (I - K_t H_t) C_{\theta_{t_{t-1}}} \hspace{1cm} (18)$$

$$\hat{\theta}_t = \hat{\theta}_{t_{t-1}} + K_t (z_t - H_t \hat{\theta}_{t_{t-1}}). \hspace{1cm} (19)$$

In this study, we use a simpler model for the dynamic estimation of evoked potentials. In the state-space equations we set $F_t = I$ and $G_t = I$. Then the state-space equations are of the form

$$\theta_{t+1} = \theta_t + \omega_t \hspace{1cm} (20)$$

$$z_t = H_t \theta_t + \nu_t. \hspace{1cm} (21)$$

The time variations of the state vector are then following the so-called random walk model. If we denote the conditional covariance matrix of the parameter estimation error as $P_t = C_{\theta_t}$, the Kalman filter equations can be written in the form

$$K_t = P_{t-1} H_t^T (H_t P_{t-1} H_t^T + C_{\nu_t})^{-1} \hspace{1cm} (22)$$

$$P_t = (I - K_t H_t) P_{t-1} + C_{\omega_t} \hspace{1cm} (23)$$

$$\hat{\theta}_t = \hat{\theta}_{t_{t-1}} + K_t (z_t - H_t \hat{\theta}_{t-1}), \hspace{1cm} (24)$$

where $K_t$ is the so-called Kalman-gain matrix. Estimators for the ERPs can be directly obtained from equation (6).

We can also write equation (24) in the following form

$$\hat{\theta}_t = \hat{\theta}_{t_{t-1}} + K_t \epsilon_t. \hspace{1cm} (25)$$
where the residual or prediction error $\epsilon_t = z_t - H_t \hat{\theta}_{t-1}$ is the estimator of the unknown noise random vector $v_t$. With different choices or assumptions for $C_{\omega_1}$, $C_{v_1}$, and $P_0$ several algorithms can be written in the form (25). Kalman gain matrices $K_t$ and recursive covariance estimates $P_t$ for different recursive algorithms, namely recursive least squares (RLS), least mean square (LMS), and normalized least mean square (NLMS), are presented in Table I. In that sense these recursive algorithms can be optimal in the mean square sense if the specific choices or assumptions about the parameters are valid. The connections of RLS, LMS, and NLMS algorithms to Kalman filtering are discussed e.g. in [16], [17].

### C. Dynamical estimation of ERPs

The most obvious way to handle time variations between single-trial measurements is sub-averaging of the measurements in groups. Sub-averaging is used, e.g., in [18] to demonstrate the decrease of amplitude in visual ERPs. Sub-averaging could give optimal estimators if the ERPs are assumed to be deterministic within the sub-averaged groups. Another method for the dynamical estimation of ERPs is the windowed averaging of the measurements. This can also be called sliding window averaging. The estimator then takes the form of a moving average filter of the measurement values in every time lag. In vector form the moving window average filter for dynamical estimation of ERPs is

$$\hat{s}^\text{MWA}_t = \sum_{i=0}^{n-1} w_i z_{t-i}. \quad (26)$$

In [19], this filter average was used with equal weights $w_i = 1/n$. Another method which was used in [19], [20] is exponentially weighted average (EWA) in which the weights are of the form $w_i = \gamma^i / \sum_{j=0}^{i-1} \gamma^j$ for some $0 < \gamma < 1$. It can be shown that the equivalent form for EWA is then

$$\hat{s}^\text{EWA}_t = \gamma z_t + (1-\gamma) \hat{s}^\text{EWA}_{t-1}. \quad (27)$$

The common disadvantage of these moving averages is that they cannot be adaptively tuned based on the data. Another is that neither their statistical properties nor the assumptions imposed to the ERPs and background EEG can be directly defined. However, their statistical properties can be investigated through the Kalman filter equations. For example, by choosing $H_t = I$ for every $t$ we have from equation (24)

$$\hat{s}_t = \hat{\theta}_t - 1 + K_t (z_t - \hat{\theta}_{t-1}) = K_t z_t + (1-K_t) \hat{s}_{t-1}. \quad (28)$$

If we compare now with equation (27) we can see that exponentially weighted averager can be obtained by choosing some fixed value for the Kalman gain such as $K_t = \gamma I$. Actually it is enough to choose $H_t = I$, $C_{v_1} = C$, $C_{\omega_1} = \gamma^2 / (1-\gamma) C$ for every $t$ and $P_0 = \gamma / (1-\gamma) C$ to the Kalman filter equations and obtain exponential weightedaverager.

In all the estimators based on weighted averagers, the background EEG is assumed Gaussian and stationary both within a specific trial and from trial to trial. As it is shown, these assumptions are verified and made explicit through the connection to Kalman filtering. We also explained that different recursive algorithms can be compared to recursive mean square estimation. They can be derived from Kalman filter equations either with some specific choices for the Kalman gain and conditional covariance, which involves the hyper-model of the state-space representation, or with some specific choices for the covariances $C_{v_1}$ and $C_{\omega_1}$.

Kalman filtering gives a good theoretical base for sequential estimation and a good starting point for investigation of realistic models for dynamical estimation of ERPs. For the methods based on Kalman filtering, a common issue is the investigation of optimal choices for the covariance matrices $C_{v_1}$ and $C_{\omega_1}$ in an unknown environment. The adaptation or the speed-of-change of the parameters is described by the state noise covariances $C_{\omega_1}$ and $C_{v_1}$ involves the prediction error. Another issue in Kalman filtering is the discussion of the validity of the relevant assumptions in relation to the field of application and how the violation of the assumptions influence the resulting estimates.

One very common strategy in state-space modeling is the choice of a diagonal matrix for $C_{\omega_1}$. This choice is made so that the individual parameter evolutions are assumed and allowed to be independent. This assumption could also be seen in relation to the selected observation model. In the following section we use a simple observation model based on shifted Gaussian shaped functions. These functions are simply selected so that the residual of the equation (8) is small and that the least squares estimates follow closely the data. Then the only assumption of this model is that the ERPs are forced to be smooth waveforms. This choice is not affecting the exact shape of the ERPs from trial to trial and independent evolutions seem logical. If some parts of the ongoing EEG are dependent within a trial, this is expected to be modeled by the parameters $\theta_i$.

The step-sizes $\omega_i$ of the states are assumed uncorrelated from trial to trial and from the parameters. This is an easy requirement and is not related to the data, but to the parametrization. For dynamical estimation of ERPs we further assume or allow that the interesting phenomena should be slowly varying from stimulus to stimulus in a nearly constant way. The Gaussian assumption for the vector $\omega_1$ is then also a simple requirement. Then the selection of $C_{\omega_1} = C_{\omega}$ for every $t$ can be applied. This assumes that every sudden variation is due to the background noise and it should be filtered out.
from the estimates. In the following section, we make the simplifications \( C_{\omega t} = \sigma^2_{\omega t} I \), i.e. all parameters are allowed to change similarly as slow from trial to trial. Of course how sudden these transitions are allowed to be is tightly related to the selection of \( C_{\omega t} \).

The observation noise covariance matrices \( C_{v t} \), although assumed to be known until now, could be estimated based on the data, which is a difficult estimation task. Background EEG, even not related to the stimulation, is a non stationary process. Therefore, its properties cannot be estimated accurately from a pre-stimulus sample or from ensemble data. Estimation based on the prediction error creates extra considerations since it is related to the selection of \( \sigma^2_{\omega t} \). When recursive algorithms are applied for adaptive filtering or modeling of time series, different strategies exist for time varying selection, which in that case implies \( C_{v t} = \sigma^2_{v t} I \). Although these methods are not directly applicable here, for a review see for example [21]. Usually the selection or the inaccurate estimation of non diagonal matrices for the noise covariances can lead to completely inaccurate results. It is then necessary to consider the simplification that the noise covariances are diagonal matrices, at least by assuming that the time varying variances are much larger than the covariances.

The assumptions imposed to the background noise are more essential in dynamical estimation of ERPs. The Gaussian assumption can be easily relaxed, if we only require the optimality among all the linear estimators. Furthermore, it is well known that EEG measures activity generated from different independent brain locations and activity arising from nonneural sources such as muscles. Then, we can consider that a significant part of the variability in ERP data is the effect of an almost Gaussian random vector, created by a nearly infinite sum of noisy independent processes that are not relevant to the stimulation. Additionally, the experiments involving ERPs are usually well randomized, and some part of the activity must be uncorrelated from stimulus to stimulus. Then with appropriate choices for \( \sigma^2_{\omega t} \) and \( C_{v t} \), it is possible to filter out, firstly, the nearly Gaussian noisy contributions and, secondly, all the other uncorrelated from trial to trial contributions and then be sure for the optimality of the estimates. For some sources of variation in the data we can also assume that their randomly occurring trial to trial correlations are randomized enough so that they are self canceled and they do not create significant bias to the estimates by the violation of the uncorrelatedness requirement for \( v_t \).

In practice, in ERP analysis it is difficult to decide which part of the activity is related to the stimuli and which trial to trial correlations are not due to the stimulation. From this arises the following intuitional approach for dynamic estimation of ERPs when slow changes of interest exist from stimulus to stimulus. With appropriate tuning of the relevant parameters state-space representation and Kalman filter first model the most dominant phenomena correlated to the previous trials present in the data. Any other nearly randomly occurring phenomena will be filtered out. Any significant violation of the assumption for \( v_t \) will create bias in the estimates which can hide some interesting results of the experiment.

\[ s_t = \sum_{i=1}^{3} a_t(i)e^{-(T-b_t(i))^2/c_t^2(i)}, \]  

(29)

where \( a_t(i) \) is the amplitude, \( b_t(i) \) is the latency, and \( c_t(i) \) is the width of the ith Gaussian component at the tth stimulus. The Gaussian components were computed over the vector \( T = \{-49, \ldots, 0, \ldots, 300\} \) such that the theoretical stimulus occurs at time point zero. Therefore, and being consistent to the real measurements sampling rate each simulated ERP vector has three Gaussian peaks, a negative around 100 ms after the stimulus, a negative around 200 ms, and a positive around 300 ms.

III. CASE STUDIES

In this section, we present the performance of Kalman filter on tracking dynamic variations and estimating single-trial ERPs in three simulated and one real data sets. Although the proposed method is capable of estimating different kinds of ERPs, we have simulated data resembling the P300 peak. The P300 peak is one of the most extensively studied cognitive potentials [22] and there exist many works where the trial to trial variability of the component is discussed, for example [23], [24].

A. Measurements and Simulations

Real EEG data were used as background EEG activity or noise in the simulations. The EEG measurements and P300 responses were obtained from an odd-ball paradigm with auditory stimuli. The sampling rate of the EEG was 500 Hz. From the recordings, we used only the channel CZ, after digital filtering in the range 1-40Hz. As background EEG activity for the simulations we selected epochs from -700 ms to 0 ms relative to the standard stimulus onset having amplitudes between +40 mV to -40 mV. To evaluate the method to real activations we took epochs from -100 ms to 600 ms relative to the stimulus onset of each deviant stimulus from the same subject and channel.

Simulated ERPs were constructed according to the additive noise model (4) by superimposing upon the selected real EEG epochs, that represent the background activity, linear combinations of 3 Gaussian shaped functions of the form

![Fig. 1. Amplitudes and latencies for the third peak in the three simulated cases. Case 1: linear and random variations, Case 2: sinusoidal and random variations, Case 3: only random variations uniformly distributed.](image-url)
In order to evaluate the method for different kinds of dynamic variations, we have created three sets of simulations for the noiseless third peak. In the first case (Case 1), the trial to trial dynamic variations of the amplitudes and latency of the third peak were set to be linear functions of time. In the second case (Case 2), dynamic variations were represented by two different sinusoidal functions, one for the amplitudes and one for the latencies. Random variations, uniformly distributed, were further added both to the amplitudes and latencies in the two cases. In the third case, for the amplitudes and latencies of the third peak uniform variations in a larger range were allowed. The widths of the third peak in all the cases were also uniformly distributed but in a smaller interval. The amplitudes and latencies for the three cases are presented in Fig. 1 as functions of the stimulus number.

The other two peaks were selected to be the same for the three cases having random variations for the amplitudes, latencies and widths, and slow linear trends in amplitudes. We also used exactly the same background noise for all the cases in order to better compare the noisy simulated ERPs and the estimates.

B. Results

For the ERPs estimation we considered the linear observation model given by equation (5) for the simulated and the real data sets. Recursive estimators in the mean square sense were computed for the parameters \( \theta_t \) with Kalman filter equations (22)-(24). Estimates for ERPs were then obtained from equation (6). For the three simulated cases as well as for the real data set, we used the same generic observation model. The model was selected to be the same for every stimulus, and had as columns 30 shifted Gaussian functions with maximum values at 0.15 and latencies of 10 points.

For the covariances we used \( C_{w_1} = \sigma^2_1 I \) and \( C_{w_2} = \sigma^2_2 I \). The selection of the last variance term is not essential since only the ratio \( \sigma^2_2/\sigma^2_1 \) has effect on the estimates. When Kalman filtering techniques are applied, usually the choice \( C_{w_i} = I \) is made and care is given to the selection of \( \sigma^2_i \). In general, if it is tuned too small, the estimates have bias towards the previous estimates and if it is selected too big, the estimates have too much variance and they will tend to be similar to the ordinary least squares solution (8). Usually the selection is based on experience and visual inspection of the estimates. If the variances of the background noise change randomly from trial to trial, a time-varying selection is not expected to improve the overall properties of the estimates. It is expected to improve some individual single-trial estimates mainly related to much lower or much higher signal to noise ratio conditions. Intuitively, very noisy single-trials are better estimated based on the past, while less noisy based on the present. If the background noise is changing with a clear pattern, for example, if its absolute value grows for trial to trial, the noise variances should be selected accordingly. The effect of the initial values \( \theta_0 \) and \( P_0 \) can be removed by using the algorithm first backward in time. The last estimates of the backward run can then be used as initial estimates in the forward run. Usually around 10 backward steps are enough for a good starting point.

To identify a good value for the variance term \( \sigma^2_1 \) we calculated root mean square errors (RMSEs) between the estimates based on the noisy data and the noiseless simulated ERPs for the three different cases. In order to investigate the performance of the filter when the noise is not present, we also computed new estimates based on the noiseless data and new RMSEs between these estimates and the noiseless simulated ERPs. Since we focus on the third peak, the RMSEs were computed over the time interval shown at the upper left part of Fig. 2. The means of the RMSEs for different values of the variance and for the three different cases are presented in Fig. 2. For the first case, with the linear variability, the mean of the RMSEs for the noisy estimator (solid line) takes a clear minimum. Since we wish to keep the error related to the noiseless estimator small (dashed line), a value around 0.01 seems to be optimal for the selection of the variance term. Similarly, for the second case with the faster sinusoidal variations, a higher value around 0.05 seems to be optimal, as it was expected. For the third case, with only larger random variability, the mean (solid line) is not taking any clear minimum. In order to keep both means in an acceptable level, the value 0.5 was selected for the variance. We also computed the mean of RMSEs between the ordinary least squares estimators based on the noisy data and the noiseless ERPs and it was found to be around 8 for all the three cases.

By construction, the synthetic datasets have similar means but different higher order statistics. Different plots describing the simulated ERPs are presented for Case 1, Case 2, and Case 3 in Figures 3 (a), 4 (a), and 5 (a), respectively. From the epoch plots (upper right), only Case 1 can be easily distinguished from the others. Similar observations can be made by comparing the different images of the covariance matrices of the data (bottom right). Only the comparison of the epochs as image plots (left) reveal their true differentiation. After adding the noise (see Figs. 3 (b), 4 (b), and 5 (b))
this differentiation can barely be observed only in the images of the epochs. A single-trial estimation method is obviously necessary for the differentiation of the three cases.

The estimates with Kalman filter are presented in Figs. 3 (c), 4 (c), and 5 (c). Excellent estimators are obtained for the first case, even under poor signal to noise ratio conditions (epochs for stimuli 80-100). The amplitudes and the latencies were computed simply as the maximum in the time interval given at the upper left part of Fig. 2. Their dynamic variability is clearly estimated (Fig. 3 (d)) and clear single-trial estimates are obtained (Fig. 3 (e)). For the second case, some noise is still present in the estimates. Although the sinusoidal variabilities both in amplitudes and latencies are well estimated (Fig. 4 (d)), the effect of the noise is still visible around stimulus number 70. However, very good single-trial estimates were obtained (Fig. 4 (e)). For the third case the noise is still highly present in the estimates (Fig. 5 (a)). The sudden amplitude changes are better estimated than the sudden latency changes (Fig. 5 (b)). In the estimated single-trials (Fig. 5 (e)) the noise is significantly decreased and the estimators could be considered still acceptable.

From the analysis of the simulations, we can conclude some general observations consistent to the theory of recursive mean square estimation and state-space representation. Recursive estimation by Kalman filter can model dynamic variability and creates accurate estimates for the simulated ERPs. When slow variations exist, Kalman filter gives excellent estimates for single-trials and the random walk model follows the underlined trends well. Relatively slow variability permits the choice of small values for the state variances which leads to highly improved signal to noise ratio, even though the Gaussian assumption for the background noise is not valid. Bigger values for the state noise variance can still give good estimates, in the expense of noise still present, especially in the case that the noise has some significant degree of correlation from trial to trial. When only sudden and large random variations exist, bigger values for the state covariance are appropriate in order to keep the random structure of the peak. Even in this case, reasonable estimates especially for the amplitudes can be obtained with significant improvement to the signal to noise ratio.

The effect of the filter for different values of the state covariance matrix can be observed by comparing the estimates for the first two peaks which are the same for the three simulated cases, as well as, by the data samples after the third peak. The first two peaks are, by construction, better estimated in Case 1.

With the same generic observation model, we computed estimates for the real P300 measurements. The real measurements are presented in Fig. 6 (a). By visual inspection, as well as by using the results obtained by the simulations, a value around 0.1 for the state variance was selected. The results for that choice are presented in Fig. 6 (b). We further observed in Fig. 6 (b) that the first two peaks vary in a smaller range, especially in terms of their latencies, comparing to the P300. We also considered that the data samples before the stimuli and after 450 ms do not correlate to the stimuli. Then we created a diagonal state covariance matrix with maximum values 0.1 associated to the parameters related to the P300, smaller values for the N200, and even smaller for the N100. Much smaller values were selected for the area outside the estimated ERPs. The estimated ERPs with that covariance matrix are presented in Fig. 6 (c). The estimates for the P300 are the same for both cases, but the second selection improved the covariance structure of the estimators. The dynamic variability of the P300 is presented in Fig. 6 (d) and single-trial estimates in Fig. 6 (e).

In different studies, related to single-trial ERP analysis and estimation, evidences have been presented that stimulation is followed by a strong phase synchronization in different bands of the ongoing EEG; for example in [25] involving auditory responses. This was compared to almost no phase synchronization to prestimulus data. Based on these findings, we could further assume that these phase synchronizations create the most dominant correlations from trial to trial present in the data. If this is happening in a dynamic way, it is expected to be modeled by Kalman filter. According to this interpretation, Kalman filter is able to remove form the data the activity that is less phase-locked during the repetition of the test. This becomes possibly more visible with the diagonal matrix selected and the estimates presented in Fig. 6 (c), when larger parts of activity were removed before the stimuli and after the P300 responses. Based on these assumptions, the resulting estimates are not solely interpreted through the additive noise model, but they could be described as the effects of these time-varying phase reorganizations. Bias in the estimates is then created by other phase synchronizations, possibly present in the data, which are not related to the stimulation, or, if these reorganizations are not changing slowly and dynamically from trial to trial. However, with the parametrization used here, the frequency bands in which the synchronizations may occurred are not directly observable and possible phase synchronizations can not be directly verified.

The results presented here for P300 suggest that even in simple experiments a dynamical behavior from trial to trial could be expected. Furthermore, the assumption that brain responses in relation to different attention levels are not a completely randomly changing process, should be studied in more specific experiments. For example, studies could be done with subjects allowed to fall asleep or with time-varying levels of extra distraction. Effects of the level of attention, alertness, or degree of habituation and fatigue could then be estimated with dynamical estimation. Then the results could be better explained and validated with independent physiological measures.

IV. CONCLUSION

We have applied the Kalman filtering algorithm for dynamical estimation of ERPs in simulated and real data sets. The proposed dynamical estimation method seems to give excellent estimates for single-trials, if previous trials contain prior information of relevance to the trial being analyzed. When poor signal to noise ratio conditions exist, the filter still gives reasonable estimates. If the peak under consideration is dominated mainly by sudden and random variations, the
for the same value of state covariance.

Fig. 3. Simulated Case 1. Linear and random variations for the third peak. Recursive mean square estimator (Kalman filter). Dotted lines represent the simulated noisy ERPs, thin lines the simulated ERPs, thick lines the estimator based on noisy ERPs and dashed lines the estimator based on noiseless ERPs for the same value of state covariance.

Fig. 4. As Fig. 3 but for simulated Case 2. Sinusoidal and random variations for the third peak.
thin lines the least squares estimates and thick lines the estimator with diagonal state covariance.

Fig. 5. As Fig. 3 but for simulated Case 3. Only random variations uniformly distributed for the third peak.

Fig. 6. Real ERPs. Dynamic variations for the P300 peak. Recursive mean square estimator (Kalman filter). Dotted lines represent the measurements, solid thin lines the least squares estimates and thick lines the estimator with diagonal state covariance.
filter could be used only with large values for the state noise covariances, leading to increased noise still being present in the estimates.

In general, the state-space representation for the evolution of ERPs and recursive mean square estimation have many additional benefits compared to other dynamical estimation methods. Perhaps the most important is the optimality of the mean square estimator among a wide class of estimators. The optimal recursive estimate in the mean square sense is given by the Kalman filter and for different choices of the relevant parameters different recursive algorithms can be obtained for dynamical estimation. Then their statistical properties, as well as the relevant assumptions about the parameters, can be compared with mean square estimation.

Kalman filtering also directly allows the use and investigation of more realistic models for ERP fluctuations than the random walk model. The Bayesian aspect of the method is another benefit of Kalman filtering, because it allows a feasible modeling of prior knowledge about the parameters. For example, this can permit the combined use of Kalman filter with some fixed interval smoother [26]. It is also possible to use the so-called state-space identification methods or hyper-parameter models for the state equations [27].

ACKNOWLEDGMENT

The authors would like to thank the Department of Clinical Neurophysiology of the University Hospital of Kuopio for the provision of the EEG measurements. Stefanos Georgiadis acknowledges support from the Foundation of Sciences of Instrumentarium and from Magnus Ehrnrooth Foundation administrated by the Finnish Society of Sciences and Letters.

REFERENCES