ERP Detection Based on Smoothness Priors

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY-NC-SA 4.0

SUBMISSION DATE / POSTED DATE

24-11-2021 / 06-12-2021

CITATION

Mobaien, Ali; Boostani, Reza; Mohammadi, Mokhtar; Sanei, Saeid (2021): ERP Detection Based on Smoothness Priors. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.17074772.v1>

DOI

[10.36227/techrxiv.17074772.v1](https://doi.org/10.36227/techrxiv.17074772.v1)

ERP Detection Based on Smoothness Priors

Ali Mobaen, Reza Boostani, Mokhtar Mohammadi and Saeid Sanei, *Senior Member, IEEE*

Abstract—Objective: Detection of event-related potentials (ERPs) in electroencephalography (EEG) is of great interest in the study of brain responses to various stimuli. This is challenging due to the low signal-to-noise ratio of these deflections. To address this problem, a new scheme to detect the ERPs based on smoothness priors is proposed. **Methods:** The problem is considered as a binary hypothesis test and solved using a smooth version of the generalized likelihood ratio test (SGLRT). First, we estimate the parameters of probability density functions from the training data under Gaussian assumption. Then, these parameters are treated as known values and the unknown ERPs are estimated under the smoothness constraint. The performance of the proposed SGLRT is assessed for ERP detection in post-stimuli EEG recordings of two oddball settings. We compared our method with several powerful methods regarding ERP detection. **Results:** The presented method outperforms the competing algorithms and improves the classification accuracy. **Conclusion:** The proposed SGLRT could be employed as a powerful means for different ERP detection schemes. **Significance:** ERP-based systems (e.g. brain-machine interfaces) mainly suffer from the lack of classification accuracy, hence, the proposed method is an important step toward real-life applicability of these systems.

Index Terms—Event related potentials, P300, smooth signal, detection, generalized likelihood ratio test.

I. INTRODUCTION

EVENT related potentials (ERPs) are small deflections in the brain activity signals which are believed to be related to the attention and perception levels of the brain processes. These small potentials can be evoked using sensory stimuli or even by motion and cognitive events [1]. Detecting and studying ERPs are of great interest in the field of biomedical signal processing, and can be used for medical diagnostics [2] or brain-computer interface (BCI) systems [3]–[5]. Electroencephalography (EEG) is one of the most convenient tools to capture these potentials. EEG signals can be seen as a combination of desired signals (i.e. ERPs), the background EEG (the constant brain waves related to normal physiological activities), and noise (e.g. the measurement noise, eye blinks, and movement artifacts). If the non-ERP components are counted as noise, then the signal-to-noise ratio (SNR) of ERP is low, and therefore, the other sources often dominate the ERP waveform. Thus, detecting ERPs and tracking their parameters through the ongoing EEG are challenging [1]. Hence, new and effective signal processing techniques to more accurately detect and estimate the ERPs become necessary.

ERP sequences are composed of several components and each component is related to a specific level of brain process [6]. EEG-ERP frequency domain analyses have indicated that these components have a smooth behavior compared to the wide-band background activities [7], [8]. This a priori on the desired signal structure can be used to improve the conventional ERP detection techniques.

Contribution: In this research, we propose a new scheme to detect ERPs from multichannel EEG signals based on smoothness priors. The main contribution of this work is to mathematically model and embed the smoothness constraint into an ERP detection scheme and provide a unified powerful method. To reach this goal, the problem is considered as a binary hypothesis test, and it is solved by a two-step generalized likelihood ratio test (GLRT). In the first step, the GLRT solution, without considering smoothness priors, is obtained. It is shown that the achieved detector tends to be a linear discriminant analysis (LDA) method. In the second step, smoothness priors are added to the problem. Consequently, the solution leads to a smoothing operator that maps the data into a smooth subspace to better estimate the desired signal. Then, this operator is imposed on the GLRT solution and the proposed smooth GLRT (SGLRT) is achieved. Monte Carlo simulations over different ERP datasets show improved detection rates for the new method. The proposed SGLRT mainly relies on a single parameter that specifies the smoothness level of the desired signal, and it can easily be tuned for different ERPs.

The proposed method is evaluated over two ERP components namely P300 and mismatch negativity (MMN), focusing on P300 as one of the most studied ERPs. These components are more pronounced during oddball paradigm. The subjects are exposed to a series of frequent (non-target) and infrequent (target) visual or auditory stimuli in such settings. Research findings indicate that when a subject is exposed to an infrequent stimulus, there is a small deflection in his/her EEG [7], [9]. The brain response to a rare stimulus can be used as a cue to control a BCI system. Generally, the studies on ERP signals can be divided into two categories. The first category focuses on ERP detection and tries to solve a binary classification problem to determine the target stimuli [10], suitable for BCI systems (e.g. a speller) [11]. The second category focuses on the estimation of ERP subcomponents and tracking their parameters (such as amplitude, latency, and width). In these studies, usually the target trials are known and the inter-trial variability of ERPs is of interest [12], suitable for diagnostic purposes [2]. The scope of this contribution is the detection of ERPs. According to the review article by Philip et al. (2019) [11], the most frequent employed classification methods for ERP (P300) elicitation are discriminant analysis (DA), support vector machines (SVM), distance-based classifiers, artificial

Manuscript received November 23, 2021.

A. Mobaen and R. Boostani are with the Department of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran (e-mail: a.mobaen@shirazu.ac.ir; boostani@shirazu.ac.ir).

M. Mohammadi is with the Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, KR-Iraq (e-mail: mokhtar@lfu.edu.krd).

S. Sanei is with the School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, U.K. (e-mail: saeid.sanei@ntu.ac.uk).

neural networks (ANN) and ensemble classifiers [10], [13], [14]. Among recent studies regarding ERPs, the two following works can be named. Yang et al. (2019) [15] proposed a simultaneous spatiotemporal equalization (STE) technique for whitening EEG data based on a multivariate autoregressive (MVAR) model, followed by a correlation detection (CD) method (STE-CD) for P300 detection. Campos et al. (2020) [16] developed an efficient algorithm to extract the underlying waveforms of ERPs using a three-step spatial filtering method. These waveforms are called principal ERPs (pERPs), and the method to represent ERP signals based on the pERPs is called pERP-reduction (pERP-RED). Most of these studies consider the ERP smoothness by roughly applying a low pass filter to the data, and none of them consider the smoothness as a built-in property of their method. In contrast, in our method, the smoothness is an intrinsic property, which we believe, can be beneficial by enhancing the estimation of ERPs.

SGLRT is compared to LDA, SVM, STE-CD, and pERP-RED. The methods are evaluated over three publicly available datasets on P300 and MMN. The results demonstrate that the presented method outperforms the competing algorithms in the sense of classification accuracy.

The rest of the article is as follows. In section II, the proposed method is rigorously investigated. In section III, the evaluation schemes and parameter selection are discussed and the comparative results of the different methods are presented. Section IV includes the discussion over the competing methods. Finally, the paper is concluded in section V.

II. METHOD

Here, a two-step GLRT approach is developed to detect the activity of an ERP component in the background EEG. This method is usually used when the probability density function (PDF) parameters of observations are unknown. For a binary hypothesis test, GLRT is extracted as the likelihood ratio of observations, where the PDF unknown parameters are calculated using maximum likelihood (ML) estimation under the corresponding hypothesis. Then, the likelihood ratio is compared with a predefined threshold and the hypothesis is determined. The threshold is calculated based on a desirable false alarm rate [17].

A. The Proposed Detection Scheme

The framework for the proposed SGLRT is summarized as follows. In the first step, the detection problem is solved by a GLRT approach without considering smoothness priors. For this,

- The problem is formulated as a binary hypothesis test.
- The unknown PDF parameters are estimated under each hypothesis; then, the likelihood ratio is formed and the GLRT solution is found.

In the second step, the smoothness priors are added to the problem and the solution is found. For this,

- A model is defined for the observed smooth signal, and its PDF is found.
- The desired signal is estimated under smoothness constraint which leads to a smoothing operator.

- The smoothing operator is imposed on the GLRT solution and it is updated to SGLRT.

1) *GLRT Solution without Smoothness Priors:* Here, we should decide whether the post-stimulus EEG recordings include ERP. This is to separate the target and non-target trials derived from an oddball setting while non-target trials are considered as the background activity. Furthermore, the background activity is modeled as a multivariate Gaussian noise with an unknown structure; The mixing model of signal and noise is assumed to be additive, widely used in the literature [18].

Let $X \in \mathbb{R}^{J \times N}$ be a post-stimulus data matrix, where J is the number of samples in time and N is the number of recorded channels. The data matrix is then vectorized by concatenating its columns, leading to $JN \times 1$ column vector $\underline{x}$. Each vectorized or matrix form of post-stimulus observations is known as a trial. Let Class c_0 be a collection of trials excluding ERP, and class c_1 be a collection of trials including ERP.

Problem Formulation: Consider $\underline{x}_0$ as a vector of observations to be tested, and $\underline{x}_k^i$ as k 'th training trial of class $i \in \{0, 1\}$. Here, the problem can be seen as a binary hypothesis test,

$$H_0 : \underline{x}_0 = \underline{v}_0 \quad \text{vs.} \quad H_1 : \underline{x}_0 = \underline{p} + \underline{v}_0$$

and for both hypotheses we have the training trials as

$$\begin{cases} \underline{x}_k^0 = \underline{v}_k^0, & k = 1, 2, \dots, K_0 \\ \underline{x}_k^1 = \underline{p} + \underline{v}_k^1, & k = 1, 2, \dots, K_1 \end{cases}$$

where K_0 and K_1 are the numbers of available training data samples for the classes c_0 and c_1 , respectively. In the above representation, $\underline{v}_0$, $\underline{v}_k^0$ ($k = 1, 2, \dots, K_0$) and $\underline{v}_k^1$ ($k = 1, 2, \dots, K_1$) are $JN \times 1$ column vectors of noise which are assumed to be drawn from a multivariate normal distribution with unknown mean and covariance. These vectors are assumed to be independent and identically distributed (i.i.d.); however, the samples within each vector are correlated as represented by the covariance matrix. The noise vectors are assumed to be a combination of background EEG and other noise sources (e.g. the noise of the recording device). $\underline{p}$ is a column vector of length JN which indicates the presence of ERP; also, $\underline{p}$ is a smooth sequence.

Formation of the likelihood ratio: To derive GLRT, the likelihood ratio of PDFs under two hypotheses should be formed. To this aim, the test and train data samples are concatenated in a long vector and the PDF parameters of all data under each hypothesis are found. Since we have assumed that the trials are independent, there are only three parameters to be estimated under each hypothesis; mean of the data that does not contain ERP, mean of the data that contains ERP, and covariance matrix of normal distribution. Although these three parameters are the same for the two hypotheses, their estimations are different under each hypothesis. Here, the estimation of $\underline{p}$ is embedded in the estimation of mean.

Let $\underline{x}_{all} = [\underline{x}_0^T, \underline{x}_1^{0T}, \underline{x}_2^{0T}, \dots, \underline{x}_{K_0}^{0T}, \underline{x}_1^{1T}, \underline{x}_2^{1T}, \dots, \underline{x}_{K_1}^{1T}]^T$ ($[.]^T$ denotes transpose operation). Then, PDF of observations

under H_j , $j \in \{0, 1\}$ is,

$$f(\underline{x}_{all}; \underline{\mu}_{0j}, \underline{\mu}_{1j}, R_j, H_j) = f(\underline{x}_0; \underline{\mu}_{jj}, R_j, H_j) \times \prod_{k=1}^{K_0} f(\underline{x}_k^0; \underline{\mu}_{0j}, R_j, H_j) \times \prod_{k=1}^{K_1} f(\underline{x}_k^1; \underline{\mu}_{1j}, R_j, H_j) = \alpha_j \exp \left(-\frac{1}{2} \left[(\underline{x}_0 - \underline{\mu}_{jj})^T R_j^{-1} (\underline{x}_0 - \underline{\mu}_{jj}) + \sum_{k=1}^{K_0} (\underline{x}_k^0 - \underline{\mu}_{0j})^T R_j^{-1} (\underline{x}_k^0 - \underline{\mu}_{0j}) + \sum_{k=1}^{K_1} (\underline{x}_k^1 - \underline{\mu}_{1j})^T R_j^{-1} (\underline{x}_k^1 - \underline{\mu}_{1j}) \right] \right) \quad (1)$$

where $\alpha_j = (1/(\sqrt{2\pi}^{JN} |R_j|^{1/2}))^{1+K_0+K_1}$ is the normalization coefficient of PDF function ($|\cdot|$ denotes determinant operator), and $\underline{\mu}_{0j}$, $\underline{\mu}_{1j}$ and R_j are mean of the data that does not contain ERP, mean of the data that contains ERP, and the covariance matrix of normal distribution, respectively. For ML estimation of unknown parameters under H_j , we have:

$$\hat{\underline{\mu}}_{0j} = \frac{1}{K_0 + 1 - j} \left((1-j)\underline{x}_0 + \sum_{k=1}^{K_0} \underline{x}_k^0 \right) \quad (2)$$

$$\hat{\underline{\mu}}_{1j} = \frac{1}{K_1 + j} \left(j\underline{x}_0 + \sum_{k=1}^{K_1} \underline{x}_k^1 \right) \quad (3)$$

$$\hat{R}_j = \frac{1}{K_0 + K_1 - 1} \left((\underline{x}_0 - \underline{\mu}_{jj})(\underline{x}_0 - \underline{\mu}_{jj})^T + \sum_{k=1}^{K_0} (\underline{x}_k^0 - \underline{\mu}_{0j})(\underline{x}_k^0 - \underline{\mu}_{0j})^T + \sum_{k=1}^{K_1} (\underline{x}_k^1 - \underline{\mu}_{1j})(\underline{x}_k^1 - \underline{\mu}_{1j})^T \right) \quad (4)$$

By likelihood ratio and using the estimated parameters, GLRT can now be extracted as:

$$L_{GLRT}(\underline{x}_{all}) = \frac{f(\underline{x}_{all}; \hat{\underline{\mu}}_{01}, \hat{\underline{\mu}}_{11}, \hat{R}_1, H_1)}{f(\underline{x}_{all}; \hat{\underline{\mu}}_{00}, \hat{\underline{\mu}}_{10}, \hat{R}_0, H_0)} \quad (5)$$

The above equation is the exact GLRT solution for our hypothesis test. However, it can be simplified by excluding $\underline{x}_0$ for ML estimation of unknown parameters. It can be shown that under Gaussian assumption, the ML estimation of mean and covariance are unbiased and consistent (i.e. by increasing the size of the dataset, the estimation mean square error tends to zero) [19]. Hence, for large datasets, $\underline{x}_0$ can be ignored. In this case, $\hat{\underline{\mu}}_{00} = \hat{\underline{\mu}}_{01} = \hat{\underline{\mu}}_0$, $\hat{\underline{\mu}}_{10} = \hat{\underline{\mu}}_{11} = \hat{\underline{\mu}}_1$ and $\hat{R}_0 = \hat{R}_1 = \hat{R}$, and we have:

$$\hat{\underline{\mu}}_i = \frac{1}{K_i} \sum_{k=1}^{K_i} \underline{x}_k^i \quad (6)$$

$$\hat{R} = \frac{1}{K_0 + K_1 - 2} \left(\sum_{k=1}^{K_0} (\underline{x}_k^0 - \hat{\underline{\mu}}_0)(\underline{x}_k^0 - \hat{\underline{\mu}}_0)^T + \sum_{k=1}^{K_1} (\underline{x}_k^1 - \hat{\underline{\mu}}_1)(\underline{x}_k^1 - \hat{\underline{\mu}}_1)^T \right) \quad (7)$$

where $i \in \{0, 1\}$. Now, using simplified estimations to form likelihood ratio and taking natural logarithm, GLRT is achieved as,

$$L_{GLRT}(\underline{x}_0) = \left[\frac{1}{2} (\underline{x}_0 - \hat{\underline{\mu}}_0)^T \hat{R}^{-1} (\underline{x}_0 - \hat{\underline{\mu}}_0) - \frac{1}{2} (\underline{x}_0 - \hat{\underline{\mu}}_1)^T \hat{R}^{-1} (\underline{x}_0 - \hat{\underline{\mu}}_1) \right] \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{GLRT} \quad (8)$$

In the above equation, the weighted distance between the trial under test (i.e. $\underline{x}_0$) and mean of each class is calculated, then their difference is compared with a threshold and $\underline{x}_0$ is assigned to the closer class. The representation of GLRT in (8) is equal to an LDA approach [20]. Note that LDA is suboptimal to Bayes classifier when the distributions are Gaussian. In the Bayes classifier, a priori information (the probability of each class) and loss coefficients are used to form a fixed threshold. To designate each test data category, the likelihood ratio is compared with this threshold. Here, η_{GLRT} is considered as a floating threshold; however, it can be set by considering a priori information and loss coefficients. For example, if π_i is the occurrence probability of hypothesis $i \in \{0, 1\}$ and l_{ij} , $(i, j) \in \{0, 1\}$ is the loss of selecting H_i in case of H_j , then an appropriate value for the threshold is $\eta_{GLRT} = \ln((l_{10} - l_{00})\pi_0 / (l_{01} - l_{11})\pi_1)$ [17].

2) *SGLRT - Improving GLRT by Exploiting ERPs Smoothness*: Here, it is proposed to use a smooth estimation of the desired signal from observation vector $\underline{x}_0$ in GLRT instead of $\underline{x}_0$ itself. This idea comes from the fact that ERPs have slow fluctuations compared to the background EEG. Furthermore, to test each incoming data vector $\underline{x}_0$, it is compared with the mean signals of each class which are achieved by their temporal averaging. Hence, we suggest to map $\underline{x}_0$ into a smooth subspace and increase the SNR by eliminating high frequency and noisy contents. Therefore, the EEG under test is pulled closer to its corresponding class mean.

Smooth signal PDF: It is assumed that $\underline{x}_0 = \underline{s} + \underline{n}$, where $\underline{s}$ is an unknown smooth template and $\underline{n}$ is a noise component that covers the corresponding frequency band and is sampled from a zero-mean multivariate Gaussian distribution with known covariance matrix R (calculated by ML estimation from training data). The smooth template includes the ERP component under H_1 and excludes it under H_0 . In either case, it can be estimated from $\underline{x}_0$ under smoothness constraint. Here also an ML estimation approach is employed. Note that in this representation, $\underline{s}$ is the unknown mean of the multivariate normal distribution that $\underline{x}_0$ comes from, and we are going to estimate this mean by using only one trial. It is straightforward to show that without smoothness constraint $\hat{\underline{s}} = \underline{x}_0$.

Consider $\underline{s} = \Phi \underline{s}'$, where Φ is a structure matrix chosen to form $\underline{s}$ based on $\underline{s}'$. This matrix gives the freedom to form a convenient structure for $\underline{s}$ based on prior information about $\underline{s}'$ morphology. For example, consider that P300 has the same shape in all channels and differs only in scale. In this situation, P300 can be modeled as $\underline{s}'$ and be projected to all channels by the structure matrix Φ . Now the PDF of each observation vector $\underline{x}_0$ can be modeled as:

$$f(\underline{x}_0; \Phi \underline{s}', R) = \beta \exp \left(-\frac{1}{2} (\underline{x}_0 - \Phi \underline{s}')^T R^{-1} (\underline{x}_0 - \Phi \underline{s}') \right) \quad (9)$$

where $\beta = 1/(\sqrt{2\pi}^{JN} |R|^{1/2})$ is the normalization factor of PDF function. In the above representation, $\underline{s}'$ is the unknown smooth template to be estimated.

Desired signal estimation under smoothness constraint: To model the smoothness constraint, the fact that the j th derivative of a smooth sequence is smaller than a threshold,

is used [21], [22]. Here, a matrix form of discrete approximation of smoothing operator is adopted. Let d_j be the finite impulse response approximation of length M for the j th order derivative operator. For $j > 2$, this operator can be achieved iteratively by $d_j = d_{j-1} * d_1$, where $d_1 = [1, -1]$ and “ $*$ ” indicates the convolution operator. For a signal length $L > M$, $D_j \in \mathbb{R}^{(L+M-1) \times L}$ is defined as a Toeplitz matrix form of d_j (the derivative operator $d_j = [d_{j1}, d_{j2}, \dots, d_{jM}]$ is shifted along rows as well as columns of D_j). We have,

$$D_j^T = \begin{bmatrix} d_{j1} & d_{j2} & \cdots & d_{jM} & 0 & \cdots & 0 \\ 0 & d_{j1} & d_{j2} & \cdots & d_{jM} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_{j1} & d_{j2} & \cdots & d_{jM} \end{bmatrix}$$

The left multiplication of D_j by any column-wise signal of length L is equivalent to the j th order derivative approximation of the signal [22].

The ML estimation of $\underline{s}'$ is equal to the following optimization problem which is regularized by smoothness priors.

$$\hat{\underline{s}}' = \arg \min_{\underline{s}'} (\underline{x}_0 - \Phi \underline{s}')^T R^{-1} (\underline{x}_0 - \Phi \underline{s}'), \text{ s.t. } \|D_j \underline{s}'\|^2 \leq \delta^2 \quad (10)$$

Here δ^2 indicates the smoothness bound. The problem is a convex optimization problem with a quadratic constraint, and the Lagrangian form of the problem is:

$$\hat{\underline{s}}' = \arg \min_{\underline{s}'} \left\{ (\underline{x}_0 - \Phi \underline{s}')^T R^{-1} (\underline{x}_0 - \Phi \underline{s}') + \lambda \underline{s}'^T D_j^T D_j \underline{s}' \right\} \quad (11)$$

where $\lambda \geq 0$ is the Lagrange coefficient. This problem is in the form of a CWLSE problem [23] and its solution is obtained by:

$$\hat{\underline{s}}' = (\Phi^T R^{-1} \Phi + \lambda D_j^T D_j)^{-1} \Phi^T R^{-1} \underline{x}_0 \quad (12)$$

and $\hat{\underline{s}} = \Phi \hat{\underline{s}}'$. The Lagrange coefficient can be obtained in two ways; when the smoothness bound is presumed or known, an optimal λ can be calculated using singular value decomposition (SVD) [22], [23], and when δ^2 is unknown, methods such as the L-curve are used [22], [24]. Let $\Psi = \Phi(\Phi^T R^{-1} \Phi + \lambda D_j^T D_j)^{-1} \Phi^T R^{-1}$; then, Ψ can be seen as a projection matrix which maps $\underline{x}_0$ into a smooth subspace and pulls it in a direction to minimize the CWLSE. By increasing λ from 0, the smoothness of the estimated signal increases. Fig. 1b shows the estimated signal $\hat{\underline{s}}$ from $\underline{x}_0$ for different values of λ , $\Phi = I_{JN}$ (I_{JN} is unity matrix of JN dimension) and derivative order $j = 2$.

Imposing smoothness priors on GLRT to achieve SGLRT: Note that it is necessary to map the mean vectors into the same space as data is mapped. This guarantees that the weighted distance is measured in a common space. Now, the proposed SGLRT method can be obtained by replacing $\Psi(\underline{x}_0 - \hat{\underline{\mu}}_i)$ by $(\underline{x}_0 - \hat{\underline{\mu}}_i)$, $i \in \{0, 1\}$ in (8), and performing some simplifications,

$$L_{SGLRT}(\underline{x}_0) = \underline{x}_0^T \Psi^T \hat{R}^{-1} \Psi \hat{\underline{p}} + c \underset{H_0}{\overset{H_1}{\geq}} \underset{H_0}{\eta_{SGLRT}} \quad (13)$$

where $\hat{\underline{p}} = (\hat{\underline{\mu}}_1 - \hat{\underline{\mu}}_0)$ is an estimation of ERP component from training data, and $c = (1/2)(\hat{\underline{\mu}}_0^T \Psi^T \hat{R}^{-1} \Psi \hat{\underline{\mu}}_0 -$

Fig. 1. (a) Averaged signal of all target versus non-target trials (from dataset II of BCI competition III, subject B and channel C_z). The P300 component can be identified as a smooth positive signal, centered around 300-400ms post-stimuli. (b) Estimated signal of observations (from subject A and channel C_z) for different values of λ . The estimated signal for $\lambda = 5$ is similar to the grand average of target trials.

$\hat{\underline{\mu}}_1^T \Psi^T \hat{R}^{-1} \Psi \hat{\underline{\mu}}_1$) is a constant value. Equation (13) calculates a weighted correlation of observation vector $\underline{x}_0$ and $\hat{\underline{p}}$ in a smooth subspace to detect the presence of ERP. One may consider that $(\underline{x}_0 - \hat{\underline{\mu}}_i)$ is in fact an estimation of centralized (zero mean) noise under H_i ; therefore, $\Psi(\underline{x}_0 - \hat{\underline{\mu}}_i)$ can be seen as the projection of noise into a smooth subspace which increases SNR.

B. Validity of the Assumptions

An underlying assumption of this study is the smoothness of ERPs compared to background activities. This is supported by several studies. For instance, consider P300. It consists of two subcomponents namely p3a and p3b where p3a arrives first and is mostly related to the attention, and p3b is mostly related to memory update. The brain response to the rare stimuli is mostly related to p3b. In an extensive review article by Polich [7], the strong relationship between the theta (4-8Hz) and alpha (8-12Hz) band activities of EEG signals and the P300 subcomponents are discussed. In another study by Spencer and Polich [25], spectral analysis of the p3b from an auditory oddball task showed a strong increase in the theta band activity. Furthermore, other studies reported an interesting event in terms of alpha suppression, observed during sensory stimuli and also the cognitive tasks related to attention and memory [8]. This phenomenon is called alpha event-related desynchronization (ERD). Alpha ERD along with delta (1-4Hz) and theta-range event-related synchronization (ERS) reveals an energy shift from higher toward lower frequencies when an ERP is elicited [7]. This indicates the slow variations of induced ERPs compared to wide-band (0-60Hz) background activities.

In the method derivation, the non-target trials are considered as the background activity. This may be questionable as we know there are also ERP components related to attention in response to non-target stimuli. For instance, in a P300 oddball scheme, there is also a p3a in the non-target data with lower amplitude compared to the target. Note that, brain fatigue and habituation can decrease P300 amplitude; however, utilizing a paradigm in which the targets are presented with a probability lower than 20%, habituation occurs only for frequent stimuli while target stimuli induce no decrease in P300 amplitude [9].

Hence, considering non-target data as background activity is a reasonable assumption in our case (Fig. 1a). Furthermore, the background activity is modeled as Gaussian noise. This choice is based on the idea that the background activity results from the superposition of the signals from numerous neurons which are quasi-randomly activated. Based on the central limit theorem, its distribution tends toward a Gaussian function. This assumption is also tested using Henze-Zirkler's and Mardia's multivariate normality tests [26] on the employed datasets. For P300 data, the first test approved Gaussian assumption at significance level $p = 0.05$, while the other approved skewness test. However, for MMN data, only Mardia's skewness test was approved.

In the derivation of GLRT solution, it is mentioned that the test trial can be excluded from the estimation of PDF parameters if the number of training trials is large enough. For a specific dataset, the sufficient number of data can be investigated using a T-test on the estimated parameters, including and excluding the test trial. When there are no meaningful changes in the estimated parameters by increasing the data size from a certain point, the sufficient number of data is found. For the studied datasets in this work, using more than twenty trials is enough to approve the assumption at $p = 0.05$.

III. METHOD EVALUATION

A. Data

1) *P300*: The presented method is investigated using two publicly available P300 datasets, from [27] and [28] (dataset II of BCI competition III). The first dataset consists of ten subjects while trying to spell 10 characters. Subjects were exposed to a 6×6 matrix of characters while focusing on the desired one. The third to tenth subjects (S3, S4, ..., S10) used the row-column speller scheme where all rows and columns of this matrix were successively and randomly intensified for 100ms, followed by the blank matrix for 60ms. Two out of 12 intensifications (one row and one column) contained the desired character. This led to 20 target and 100 non-target trials. Each stimulus was repeated 15 times. 8 channels captured EEG signals at 256Hz. In the second dataset, data was recorded from two subjects (SA and SB) while trying to spell 185 characters. Again, subjects used the row-column speller with 100ms intensification of rows/columns followed by the blank matrix for 75ms. This led to 370 target and 1850 non-target trials. Each stimulus was repeated 15 times. 64 channels captured EEG signals which then passed through a 0.1 to 60Hz bandpass filter, and were finally digitized at 240Hz.

2) *MMN*: This dataset is obtained from a publicly released project called SPM [29], [30]. It is a 128-channel single-subject EEG dataset acquired from an auditory oddball paradigm. There are 480 non-target trials and 120 target trials in the dataset. SPM12 software is used to preprocess and segment the data. The data are passed through a 0.1 to 60Hz bandpass filter and down-sampled to 240Hz.

B. Evaluation Scheme and Parameter Selection

The proposed SGLRT is compared with DA, SVM, STE-CD, and pERP-RED (the represented data by the pERP-RED are fed to an LDA classifier) in terms of classification

accuracy. In this regard, a 10-fold cross-validation approach is applied to two equal-size groups of randomly chosen trials where one contains the ERP component and the other one does not. Training sets of different sizes are used to perform this task, and for each case, at least four random groups of data are generated. The final results have been defined as the average results of these groups. The competing methods are also evaluated in different SNRs using the P300 dataset. For this, the trials are averaged over the first k_r repetitions which increase SNR (or decrease noise level) about k_r times. As a result of equal-size groups and considering uniform loss coefficients, η_{SGLRT} is set to zero (the same as DA, STE-CD, and pERP-RED threshold). Moreover, finding a proper λ value and its effect on the classification performance has been investigated. Consequently, $\lambda = 5$ and 2 are used for P300, and MMN datasets. The derivative order is set to 2 and the structure matrix $\Phi = I$.

For the P300 dataset, trials are extracted using 700ms windows of post-stimuli signals, while for the MMN dataset, windows from -100ms to 400ms regarding stimuli are used. For P300, data from eight channels (F_z , C_z , P_3 , P_z , P_4 , PO_7 , PO_8 , and O_z) have been used since they are suggested as P300 related informative channels [27], and for MMN data, ten channels from all across the scalp (every 14'th channel from the first one) including C_{21} (corresponding to F_z [30]) are considered. For DA, SVM, and STE-CD, the data are bandpass filtered, using a 0.1 to 15Hz (suggested by [31]) Butterworth filter of order 4, however, the raw data are fed to SGLRT and pERP-RED as they benefit from built-in noise reduction procedures. Finally, the P300 and MMN data are down-sampled at rates 1/4 and 1/3, and data trials in each channel are demeaned and normalized to unit variance.

The pERP-RED-based classifier is implemented in R using the pERPred package [16]. Here, the set of selected parameters for pERP-RED is shown as $\{N_p, V_p\}$ where N_p is the number of pERPs and V_p is the percent variation used in the electrode PCA step. The set of chosen parameters is $\{4, 85\}$ for P300 data, and $\{3, 80\}$ for MMN data. The MATLAB platform is used for the implementation of other methods. To realize the SVM classifier, the "svmtrain" function is employed while using the least square method as its optimization technique, and a linear (SVM-lin), also an order three polynomial (SVM-pol) function as its kernels. For the realization of STE-CD, the Arfit package [32] is used. The minimum and maximum orders of the MVAR model are set to 4 and 16, respectively (suggested by [15]). For each case, the optimizer of Schwarz's Bayesian criterion is chosen as the best order. Note that, for each method, different settings have been examined for several sample sizes and the best set of parameters is chosen based on paired T-tests. Here, some of the related classification accuracy results are reported in form of the mean (significance level of ref.), where ref. stands for the reference settings or method which performs the best. For pERP-RED and MMN data, parameter set $\{3, 80\}$ (68.5 (ref.)) outperforms $\{2, 80\}$ (65.3 (0.04)), and $\{4, 80\}$ (64.7 (0.01)) for sample size 60, and $\{3, 80\}$ (73.6 (ref.)) outperforms $\{3, 70\}$ (71.1 (0.02)) and $\{3, 95\}$ (71.7 (0.07)) for sample size 100. Similar results are obtained for other sets of data. For DA, two linear (LDA) and

Fig. 2. The performance of the competing methods in different SNRs, and sample size 20. (a) P300-S4 data; (b) P300-S8 data.

Fig. 3. The performance of the competing methods in different amounts of training data. (a) P300-SB data. (b) MMN data.

quadratic (QDA) versions are tested where LDA outperforms QDA with a huge gap for all datasets (e.g. LDA (82.3 (ref.)) and QDA (73.5 (0.01)) for P300-SA data and sample size 360). A linear versus an order three polynomial function is tested for the SVM kernel, and the results of both are reported.

Note that, the SGLRT, LDA, and pERP-RED are covariance-based methods, hence, a good estimation of the covariance matrix plays an important role in their performances. Covariance estimation is challenging for high dimension data and usually requires a large dataset. For example, in the case of JN data dimension, $K > JN$ independent trials are required to gain a full rank and positive definite estimation. One approach to address this issue is the matrix normal distribution assumption for EEG data [33], [34]. This assumption reduces the required number of trials for a positive definite covariance estimation to $K > \max\{J/N, N/J\}$. Here, this model is used for P300 data, however, for MMN data it is used whenever it could provide better results for each method and each sample size. Furthermore, the first k_r trials are treated independently for covariance estimation (rather than using the average of them) which provides more data. From a practical point of view, an infinitesimal value could be also added to the diagonal elements of the covariance matrix to stabilize its inversion.

C. Results

Fig. 2 shows the performance of competing methods in different SNRs (i.e. different k_r values). The results have been presented for P300-S4 and S8 data, and sample size 20. The proposed method provides higher accuracy almost in all SNR levels (except for S4 and $k_r < 5$, where pERP-RED performs slightly better). By increasing the SNR, the performance of all methods increases, as expected.

TABLE I
MEAN (P-VALUE $\leq$) RESULTS OF THE COMPETING METHODS FOR DIFFERENT DATASETS IN FIVE SAMPLE SIZES

Data	Method	Size of Training Set				
		30	60	120	240	360
P300, SA	SGLRT	77.1 (ref.)	78.9 (ref.)	80.9 (ref.)	82.3 (ref.)	82.5 (ref.)
	LDA	74.1 (0.02)	76.3 (0.01)	78.6 (0.01)	81.1 (0.01)	82.3 (0.23)
	SVM-lin	70.9 (0.01)	69.1 (0.01)	70.0 (0.01)	75.3 (0.01)	78.7 (0.01)
	SVM-pol	66.0 (0.01)	71.1 (0.01)	74.2 (0.01)	76.1 (0.01)	76.5 (0.01)
	STE-CD	73.5 (0.01)	76.0 (0.01)	78.2 (0.01)	80.2 (0.01)	81.1 (0.01)
	pERP.	77.9 (0.78)	78.4 (0.28)	79.4 (0.01)	79.7 (0.01)	80.2 (0.01)
P300, SB	SGLRT	83.3 (ref.)	85.1 (ref.)	87.8 (ref.)	87.9 (ref.)	88.0 (ref.)
	LDA	76.6 (0.01)	80.5 (0.01)	84.7 (0.01)	85.7 (0.01)	86.1 (0.01)
	SVM-lin	74.4 (0.01)	72.4 (0.01)	75.8 (0.01)	81.4 (0.01)	83.1 (0.01)
	SVM-pol	65.5 (0.01)	72.6 (0.01)	77.3 (0.01)	79.7 (0.01)	80.0 (0.01)
	STE-CD	79.9 (0.01)	82.7 (0.01)	85.9 (0.01)	86.3 (0.01)	84.4 (0.01)
	pERP.	80.7 (0.01)	82.9 (0.01)	85.8 (0.01)	85.9 (0.01)	86.1 (0.01)
MMN	SGLRT	74.3 (ref.)	75.9 (ref.)	78.8 (ref.)	NA	NA
	LDA	72.7 (0.01)	69.0 (0.01)	70.4 (0.01)	NA	NA
	SVM-lin	72.1 (0.05)	68.9 (0.01)	65.7 (0.01)	NA	NA
	SVM-pol	60.0 (0.01)	66.1 (0.01)	73.1 (0.01)	NA	NA
	STE-CD	69.1 (0.01)	71.1 (0.01)	75.8 (0.02)	NA	NA
	pERP.	61.7 (0.01)	68.5 (0.01)	71.7 (0.01)	NA	NA

Fig. 3 indicates the average performance (over different SNRs) of different methods in different sizes of the training set. In Fig. 3a, the results for P300-SB are presented. It can be seen that SGLRT outperforms all of the other methods in different amounts of training data. Among all the competing methods, pERP-RED has the closest performance to SGLRT. SVM-pol has the weakest average results for small amounts of training data; however, by increasing the size of the training set, its performance improves. Fig. 3b, shows the results for MMN dataset. Again, SGLRT is superior compared to the other methods. Overall, it can be seen that SGLRT has good performance in detecting a weak ERP component such as MMN, even if the Gaussian assumption is not fully satisfied (see subsection II-B).

Table I indicates the averaged results of different methods for P300-SA, P300-SB, and MMN data, and for five sample sizes. The significance of the SGLRT results are also tested by paired T-tests and the p-values are reported for each method and each set of data. As it can be seen, SGLRT provides meaningfully higher detection rates compared to all the other methods for all sets of data (except for P300-SA and sample size 30, where pERP-RED performs slightly better). Note that, for P300 data the performance gap between SGLRT and LDA is higher for smaller amounts of training data, however, by increasing the size of the training set, LDA approaches SGLRT. This is due to better estimation of LDA parameters (i.e. mean and covariance) for larger amounts of training data.

Table II presents the averaged results of different methods for P300-S3 to S10, and sample size 20. In most cases, the proposed SGLRT outperforms the competing methods. The average results of all eight subjects indicate the close performance of SGLRT compared to pERP-RED and its superior performance compared to the other methods. These

TABLE II
MEAN (P-VALUE $\leq$) RESULTS OF THE COMPETING METHODS FOR P300
DATA (S3 TO S10) AND SAMPLE SIZE 20

Sbj.	Method					
	SGLRT	LDA	SVM-lin	SVM-pol	STE-CD	pERP-RED
S3	78.7 (ref.)	77.7 (0.21)	69.9 (0.01)	74.0 (0.02)	76.6 (0.09)	83.9 (1.00)
S4	91.0 (ref.)	86.5 (0.01)	84.0 (0.01)	77.7 (0.01)	87.2 (0.01)	90.7 (0.38)
S5	90.9 (ref.)	86.1 (0.01)	86.0 (0.01)	73.7 (0.01)	90.0 (0.17)	88.7 (0.02)
S6	68.9 (ref.)	69.8 (0.74)	61.0 (0.01)	53.8 (0.01)	71.8 (0.97)	65.4 (0.04)
S7	78.6 (ref.)	71.9 (0.01)	77.2 (0.26)	66.0 (0.01)	76.4 (0.08)	82.5 (1.00)
S8	91.8 (ref.)	87.4 (0.01)	81.9 (0.01)	71.1 (0.01)	88.7 (0.01)	88.4 (0.01)
S9	86.0 (ref.)	87.8 (1.00)	88.5 (0.01)	65.3 (0.01)	83.6 (0.04)	85.5 (0.29)
S10	88.0 (ref.)	83.8 (0.01)	85.9 (0.04)	73.1 (0.01)	89.5 (0.90)	89.4 (0.88)
Avg.	84.2 (ref.)	81.4 (0.02)	78.3 (0.01)	69.3 (0.01)	83.0 (0.09)	84.3 (0.53)

results along with previously mentioned results indicate the generalization of SGLRT over different data from different subjects.

The effect of λ on SGLRT performance is assessed using a small and a large training set of P300-SA data (Fig. 4). The proper value of λ can be found based on the data structure. Here, results suggest $\lambda = 5$ as a good value. Note that for λ values near to zero and for $\Phi = I$, SGLRT tends to LDA (it can be seen from (12) and (13)). By increasing λ from the proper value, the performance of SGLRT decreases; the drop in performance is even more noticeable for higher SNRs. This is due to signal and noise interference, caused by averaging over more trials (by averaging, the smoothness of desired signal and background activity increases). Results interpretation suggests an interesting solution in terms of adaptive smoothness for improving SGLRT performance. This means an adaptive λ value can be used based on the level of data smoothness. For example, higher values can be used for lower numbers of averaged trials and smaller ones for higher numbers of averaged trials.

IV. DISCUSSION

This work presents a new scheme to detect ERPs based on smoothness priors. The main goal of this study is to improve conventional methods by adding this extra information to the desired signal morphology. The proposed SGLRT benefits from an internal smoothing operator that can estimate the desired signal based on an adjustable smoothness level. This allows for a more specific and accurate trial-by-trial ERP estimation and estimation of the unknown PDF parameters. SGLRT mainly relies on these parameters to separate target from non-target trials; hence, enhancing their estimations leads to improved detection rates. The results suggest that SGLRT provides higher detection rates for small training sets compared to LDA; however, the performance gap between these methods decreases for larger amounts of training data. This can be explained by the fact that the ML estimation of mean and covariance matrix are unbiased and consistent (see section II). Hence, by increasing the size of the training set, their estimations tend to real values. The proposed SGLRT method can be directly applied to many ERP-based BCI systems (e.g.

Fig. 4. The effect of λ on SGLRT performance for P300-SA data. (a) Training-set size is 20. (b) Training-set size is 360.

P300 mind speller). Furthermore, accurate detection of ERPs can help better the separation of ERP subcomponents (e.g. p3a and p3b in P300) in a later stage, which can then be used by various BCI applications and mental activity evaluation. It can also be used as a template for ERP localization in a forward source localization solution [12], [35]. In addition, the proposed method can be categorized in the general framework of signal activity detection. Since by increasing the sampling frequency from a certain point, all signals can be considered smooth, SGLRT can be seen as an interesting solution for detecting the desired signal in wide-band noises.

Conventional methods usually consider smoothness priors by roughly passing the data through a bandpass filter. In contrast, the proposed method employs a smoothing filter which is specifically designed to maximize the separability of target and non-target trials. Some of the new methods use dynamic models for ERPs and incorporate temporal smoothness and spatial correlations. For example, the authors in [2] developed a cooperative particle filtering approach for ERP tracking over trials. These methods proved to be very effective; however, those models might not exactly follow the structure of ERPs. In our method, there is no constraint on ERP morphology and different components with different shapes (polynomial of any order, Gaussian, etc.) and latencies can be considered. The only constraint is the smoothness of the ERP component, and its level is adjustable for different ERPs through λ value. It should be noted that the proposed method filters the data in the time domain, and tries to find a smooth subspace to enhance ERPs. Several approaches in the literature use spatial filtering to find a subspace for better emerging of ERPs (for instance, the well-known xDAWN algorithm [36]), or to estimate underlying sources of ERPs (for instance, pERP-RED method). Although these methods are very effective, smoothness priors are not considered for signal structure in time. For instance, in [16], the authors say "A second limitation is that the derived pERPs are not penalized in time to attain a desired degree of smoothness.". One idea that comes to mind is to consider smoothness in a joint time-space domain, or add smoothness priors to the previously developed spatial filtering methods and achieve even more powerful schemes regarding ERP estimation or detection.

Finally, SGLRT has some limitations. First, the estimated smooth ERPs are a combination of all ERP subcomponents; hence, this method is useful for an ERP detection scheme and

cannot estimate separate ERP subcomponents (such as p3a and p3b in P300). Second, the proposed method is covariance-based, hence, increasing the number of EEG channels dramatically increases the computational burden of inverting a very large covariance matrix. A solution to this could be using a set of spatial filters before the proposed method to reduce space dimension (for instance, by projecting the channel data onto their principal eigenmodes [30]). However, how these spatial filters affect smoothness priors is an interesting question to be answered in future work.

V. CONCLUSION

In this study, a powerful ERP detection method based on smoothness priors was presented. A binary hypothesis test was defined, and it was solved utilizing a two-step GLRT method. A smoothing operator, which can estimate the signal with a tunable smoothness level was derived from the solution. The resulting method was called SGLRT. We applied SGLRT to real ERP datasets and compared it with conventional methods as well as two recent studies regarding ERP detection. The proposed method showed superior results in terms of classification accuracy. This is very promising as the ERP-based BCI systems mainly suffer from lack of classification accuracy, hence, it is an important step toward real-life applicability of these systems. In future studies, an approach based on adaptive smoothness corresponding to different noise levels will be investigated. Moreover, a joint spatiotemporal smoothing filter, which may improve the results is of interest.

REFERENCES

- [1] S. J. Luck, *An Introduction to the Event-Related Potential Technique*. MIT press, 2014.
- [2] S. Monajemi, D. Jarchi, S.-H. Ong, and S. Sanei, "Cooperative particle filtering for tracking ERP subcomponents from multichannel EEG," *Entropy*, vol. 19, no. 5, p. 199, 2017.
- [3] M. Dehghani, A. Mobaeni, and R. Boostani, "A deep neural network-based transfer learning to enhance the performance and learning speed of BCI systems," *Brain-Computer Interfaces*, vol. 8, pp. 14–25, 2021.
- [4] A. Mobaeni and R. Boostani, "ACSP: Adaptive CSP filter for BCI applications," in *2016 24th Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2016, pp. 466–471.
- [5] F. Alimardani, R. Boostani, and B. Blankertz, "Weighted spatial based geometric scheme as an efficient algorithm for analyzing single-trial EEGs to improve cue-based BCI classification," *Neural Networks*, vol. 92, pp. 69–76, 2017.
- [6] G. F. Woodman, "A brief introduction to the use of event-related potentials in studies of perception and attention," *Atten. Percept. Psycho*, vol. 72, no. 8, pp. 2031–2046, 2010.
- [7] J. Polich, "Updating P300: An integrative theory of p3a and p3b," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, 2007.
- [8] F. Boiten, J. Sergeant, and R. Geuze, "Event-related desynchronization: The effects of energetic and computational demands," *Electroen. Clin. Neuro.*, vol. 82, no. 4, pp. 302–309, 1992.
- [9] S. H. Patel and P. N. Azzam, "Characterization of N200 and P300: Selected studies of the event-related potential," *Int. J. Medical Sci.*, vol. 2, no. 4, p. 147, 2005.
- [10] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, 2010.
- [11] J. T. Philip and S. T. George, "Visual P300 mind-speller brain-computer interfaces: A walk through the recent developments with special focus on classification algorithms," *Clin. EEG Neurosci.*, 2019.
- [12] D. Jarchi, S. Sanei, J. C. Principe, and B. Makkiabadi, "A new spatiotemporal filtering method for single-trial estimation of correlated ERP subcomponents," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 132–143, 2010.
- [13] J. Qu, F. Wang, Z. Xia, T. Yu, J. Xiao, Z. Yu, Z. Gu, and Y. Li, "A novel three-dimensional P300 speller based on stereo visual stimuli," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 4, pp. 392–399, 2018.
- [14] S. Kundu and S. Ari, "P300 detection with brain-computer interface application using PCA and ensemble of weighted SVMs," *IETE J. Res.*, vol. 64, no. 3, pp. 406–414, 2018.
- [15] C. Yang, H. Zhang, S. Zhang, X. Han, S. Gao, and X. Gao, "The spatiotemporal equalization for evoked or event-related potential detection in multichannel EEG data," *IEEE Trans. Biomed. Eng.*, 2019.
- [16] E. Campos, C. Hazlett, P. Tan, H. Truong, S. Loo, C. DiStefano, S. Jeste, and D. Şentürk, "Principle ERP reduction and analysis: Estimating and using principle ERP waveforms underlying ERPs across tasks, subjects and electrodes," *NeuroImage*, vol. 212, p. 116630, 2020.
- [17] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [18] J. Candy, *Model-Based Signal Processing*. Wiley-IEEE Press, 2005.
- [19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall PTR, 1993.
- [20] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley & Sons, 2004, vol. 544.
- [21] W. Gersch, "Smoothness priors," in *New Directions in Time Series Analysis*. New York: Springer, 1993, pp. 113–146.
- [22] R. Sameni, "Online filtering using piecewise smoothness priors: Application to normal and abnormal electrocardiogram denoising," *Signal Process.*, vol. 133, pp. 52–63, 2017.
- [23] K. W. Cheung, H.-C. So, W.-K. Ma, and Y.-T. Chan, "A constrained least squares approach to mobile positioning: algorithms and optimality," *EURASIP J. Adv. Sig. Pr.*, vol. 2006, no. 1, p. 020858, 2006.
- [24] P. C. Hansen, "The L-curve and its use in the numerical treatment of inverse problems," *IMM, Department of Mathematical Modelling, Technical University of Denmark*, 1999.
- [25] K. M. Spencer and J. Polich, "Poststimulus EEG spectral analysis and P300: attention, task, and probability," *Psychophysiology*, vol. 36, no. 2, pp. 220–232, 1999.
- [26] S. Korkmaz, D. Goksuluk, and G. Zararsiz, "MVN: an R package for assessing multivariate normality," *R J.*, vol. 6, no. 2, p. 151, 2014.
- [27] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Caraballona, F. Gramatica, and G. Edlinger, "How many people are able to control a P300-based brain-computer interface (BCI)?" *Neurosci. Lett.*, vol. 462, no. 1, pp. 94–98, 2009.
- [28] J. R. Wolpaw, G. Schalk, and D. Krusienski, (2005) BCI competition III webpage. [Online]. Available: <http://www.bbci.de/competition/iii/>
- [29] J. Ashburner, G. Barnes, C.-C. Chen, J. Daunizeau, G. Flandin, K. Friston, S. Kiebel, J. Kilner, V. Litvak, R. Moran *et al.*, "SPM12 manual," 2014. [Online]. Available: https://www.fil.ion.ucl.ac.uk/spm/data/eeeg_mmn/
- [30] M. I. Garrido, J. M. Kilner, S. J. Kiebel, K. E. Stephan, and K. J. Friston, "Dynamic causal modelling of evoked potentials: A reproducibility study," *NeuroImage*, vol. 36, no. 3, pp. 571–580, 2007.
- [31] L. Bougrain, C. Saavedra, and R. Ranta, "Finally, what is the best filter for P300 detection?" in *TOBI Workshop III-Tools for Brain-Computer Interaction-2012*, 2012.
- [32] T. Schneider and A. Neumaier, "Algorithm 808: Arfit—A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models," *ACM Trans. Math. Software. (TOMS)*, vol. 27, no. 1, pp. 58–65, 2001.
- [33] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. CRC Press, 2018, vol. 104.
- [34] P. Gonzalez-Navarro, M. Moghadamfalahi, M. Akçakaya, and D. Erdogmus, "Spatio-temporal EEG models for brain interfaces," *Signal Process.*, vol. 131, pp. 333–343, 2017.
- [35] L. Spyrou and S. Sanei, "Source localization of event-related potentials incorporating spatial notch filters," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 9, pp. 2232–2239, 2008.
- [36] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: Application to brain-computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2035–2043, 2009.