

Regularization and a general linear model for event-related potential estimation

Emmanuelle Kristensen^{1,2} · Anne Guerin-Dugué^{1,2} · Bertrand Rivet^{1,2}

Published online: 8 March 2017
© Psychonomic Society, Inc. 2017

Abstract The usual event-related potential (ERP) estimation is the average across epochs time-locked on stimuli of interest. These stimuli are repeated several times to improve the signal-to-noise ratio (SNR) and only one evoked potential is estimated inside the temporal window of interest. Consequently, the average estimation does not take into account other neural responses within the same epoch that are due to short inter stimuli intervals. These adjacent neural responses may overlap and distort the evoked potential of interest. This overlapping process is a significant issue for the eye fixation-related potential (EFRP) technique in which the epochs are time-locked on the ocular fixations. The inter fixation intervals are not experimentally controlled and can be shorter than the neural response's latency. To begin, the Tikhonov regularization, applied to the classical average estimation, was introduced to improve the SNR for a given number of trials. The generalized cross validation was chosen to obtain the optimal value of the ridge parameter. Then, to deal with the issue of overlapping, the general linear model (GLM), was used to extract all neural responses inside an epoch. Finally, the regularization was also applied to it. The models (the classical average and the GLM with and without regularization) were compared on both simulated data and real datasets from a visual scene exploration in co-registration with an eye-tracker, and from

a P300 Speller experiment. The regularization was found to improve the estimation by average for a given number of trials. The GLM was more robust and efficient, its efficiency actually reinforced by the regularization.

Keywords Event-related potential · General linear model · Overlap · Regularization · Eye fixation-related potential · P300 Speller

Introduction

The event-related potential (ERP) technique is a classical technique (e.g., Luck, 2014) used to study the neural responses to specific events such as the presentation of visual stimuli or auditory stimuli. Using this technique, many cognitive processes have been explained (e.g., Key, Dove, & Maguire, 2005). One popular way to record brain activity is to use electroencephalography (EEG) at the surface of the scalp (see Niedermeyer & Silva, 2005). However, the resulting signal-to-noise ratio (SNR) is very low due to ongoing brain activities. To overcome this drawback, the stimuli are presented repeatedly (typically a few tens of time at least), so that the neural response evoked by these stimuli is obtained by averaging the epochs that are time-locked on each stimulus onset. However, the ERP estimation is only unbiased if two main conditions are verified. (1) The (zero mean) noise due to ongoing brain activities needs to be decorrelated across the epochs. (2) Only a single potential should be elicited at each epoch and it should be elicited within each epoch. In this article, these essential conditions are discussed in practical situations and a methodology based on the general linear model (GLM) is proposed to tackle experiments where the two required conditions are not satisfied. Moreover, a regularization method to estimate

✉ Emmanuelle Kristensen
emmanuelle.kristensen@grenoble-inp.org

¹ Laboratory GIPSA-lab, University Grenoble Alpes, Grenoble, France

² CNRS, GIPSA-lab, 38000, Grenoble, France

the ERP is proposed, especially if the number of epochs is low.

Firstly, if the condition on the zero mean noise decorrelation across the trials is achieved, the SNR increases linearly with the number of epochs. Thus, the more the stimuli are repeated, the better the SNR. However, with a large amount of stimuli, the experiment becomes more time-consuming and very repetitive for the participants. In addition, and more importantly, in practice the linear increasing is not always verified because ongoing activity may be correlated from one trial to another one and the resulting SNR after averaging is lower than expected for a given number of trials. In this article, the Tikhonov regularization (Tikhonov, 1963) is used in the ERP estimation to increase the SNR for a given number of trials. The strength of the regularization is given by the value of the regularization parameter. The choice of this ridge parameter is a very common but nonetheless a difficult problem requiring the bias and the variance of the estimator to be balanced as well as possible. In the same context as this EEG study, Lalor and colleagues proposed an approach with a second-order Tikhonov regularization to estimate the event-related potential (ERP). However, they empirically chose the ridge parameter (Lalor, Pearlmutter, Reilly, McDarby, & Foxe, 2006); (Lalor, Power, Reilly, & Foxe, 2009) to give a good compromise between the off-sample errors and the weakness of the potential peaks. To estimate this parameter automatically, the two classical methods are the generalized cross-validation (GCV) procedure (Golub, Heath, & Wahba, 1979) and the “L-curve” method (Engl & Grever, 1994). Also for EEG analysis, Subramaniam and colleagues (Subramaniam, Väisänen, Wendel, & Malmivuo, 2010) compared these two methods to estimate the visually evoked cortical potentials. In their study, the Tikhonov regularization was used to solve the inverse problem of the source localization to find the spatial cortical distribution at the presentation of emotional faces. When applied separately on the EEG signals for each subject and each condition, they found that the regularized solutions were more robust with the ridge parameter computed by the GCV procedure rather than by the “L-curve” method. Indeed, the ridge parameter obtained by the “L-curve” method was too large compared to the parameter obtained by the GCV procedure when the SNR was low: as a consequence, the solution was over-smoothed and spatially blurred. Here, in this study, two estimation methods of the regularization parameter are compared but in a context of temporal estimation of the evoked potentials. Moreover, an inter-subject analysis is carried out as is usual for ERP experimental designs instead of an intra-subject study as in (Subramaniam et al., 2010). The regularization process is thus implemented at the overall level for the whole dataset, but also compared with a regularization for each subject.

Secondly, the last condition for an unbiased estimation by averaging can only be achieved if just a single identical potential is elicited in each epoch time-locked on the event’s onset. This condition does not consider possible distortion of the target neural response due to experimental design and/or latent cognitive processes. For example, Luck in (Luck, 2014) described an ERP experiment in which the inter stimuli intervals (ISI) values were between 300 and 500 ms and the studied components were P2, N2, and P3 with latencies between 200 and 400 ms. Luck warned against distortions due to the short ISI values: i.e., latencies of the elicited potentials are shorter than the ISI duration that leads to the overlapping responses evoked by two temporally consecutive stimuli. In this situation, more than one potential is elicited in an epoch so that the estimation by averaging is the summation of the delayed versions of the expected neural components. Consequently, this estimate is not the expected evoked potential itself, but the convolution of this potential with the distribution of ISI values inside one epoch. To reduce overlaps, the experimental design must control the ISI values as far as possible in agreement with the latencies of the expected potentials.

This issue had been addressed in the context of auditory evoked potentials, but it is also a central concern for the estimation of the visual potential elicited by ocular fixations, the eye fixation-related potential (EFRP) (Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Nikolaev, Meghanathan, & Leeuwen, 2016). Tanaka and colleagues (Tanaka, Komatsuzaki, & Hentona, 1996) had shown that recording auditory brainstem responses at high stimulus rate provided valuable information to detect auditory pathologies. Some methods were proposed to face the overlapping problem due to high stimulus rates. In 1982, Eysholdt and colleagues (Eysholdt & Schreiner, 1982) had proposed the maximum length sequence technique, based on a pseudo-random arrangement of stimuli in a maximum sequence length. The evoked potential pattern was obtained by a deconvolution of the averaged responses. Then, another resolution method was developed, in the frequency domain, by adaptive inverse filtering (Hansen, 1983). Three other methods, using cyclic stimuli sequences, were elaborated. (1) The continuous loop averaging deconvolution method (Delgado & Ozdamar, 2004) consisted of acquiring data from non-isochronic stimuli sequences with a high presentation rate. Then, the deconvolution was performed by matrix computations in the time domain. (2) Jewett and colleagues (Jewett et al., 2004) had applied a similar method in the frequency domain and (3) a Wiener filter was used in (Wang, Özdamar, Bohórquez, Shen, & Cheour, 2006) to deconvolve evoked potential from the noisy observation given cyclic stimuli sequences in the frequency domain. All five of these methods have a common major limitation: they only consider one kind of stimulus, and hence a

single elicited potential. Besides, we considered the ADJAR (adjacent response) algorithm developed by Woldorff in 1993 (Woldorff, 1993). This algorithm iteratively estimates the average of previous and subsequent responses that are then subtracted from the target response, for the new iteration, and so on. It is a very popular algorithm and has been successfully used in several ERP studies to correct distortions due to overlap (Bekker, Kenemans, Hoeksma, Talsma, & Verbaten, 2005; Brannon, Roussel, Meck, & Woldorff, 2004; Brannon, Libertus, Meck, & Woldorff, 2008; Fiebelkorn, Foxe, McCourt, Dumas, & Molholm, 2013; Hopfinger & Mangun, 1998; Kenemans et al., 2005; Talsma, Doty, & Woldorff, 2007; Schmajuk, Liotti, Busse, & Woldorff, 2006) by an iterative deconvolution. However, in a previous article (Kristensen, Guérin-Dugué, & Rivet, 2015), we had shown that this algorithm presented some limitations as to the temporal distribution of the ISI values, especially for EFRP studies. With this recent technique, the EEG signal can be time-locked on the onset of the ocular fixations of the participants to extract EFRP (or eye saccade-related potentials if the EEG signal is time-locked on the onset of the saccades). Using ocular events, the ISI becomes the inter fixations interval (IFI), i.e., the duration between two consecutive fixations: this interval is thus defined as the duration of one fixation plus that of one saccade. These IFIs cannot be controlled experimentally, because they depend on the participant's oculomotor pattern. On average, during experiments in reading or visual exploration, the IFIs are about 250–300 ms and thus, precautions are taken to face with overlaps between adjacent EFRPs to allow studies concerning late potentials such as P3 or N400 (latency of between 300 and 600 ms) (Devillez, Guyader, & Guérin-Dugué, 2015; Frey et al., 2013; Kaunitz et al., 2014). More recently, least-square based methods have been proposed to deconvolve multiple overlapping responses (e.g., Lalor et al., 2006; Rivet & Souloumiac, 2007; Burns, Bigdely-Shamlo, Smith, Kreutz-Delgado, & Makeig, 2013; Bardy, Dillon, & Van Dun, 2014; Crosse, Butler, & Lalor, 2015; Congedo, Korczowski, Delorme, et al., 2016; Dandekar, Privitera, Carney, & Klein, 2012). In (Bardy, Dillon, & Van Dun, 2014), a least-square method was proposed for the deconvolution of overlapping multiple responses, and this was applied to auditory potentials (Bardy, Dillon, & Van Dun, 2014; Bardy, Van Dun, Dillon, & McMahon, 2014). This method corresponds exactly to the first step of the xDAWN algorithm (Rivet, Souloumiac, Attina, & Gibert, 2009), which was designed to find optimal spatial filters for best classifying neural responses in the front of brain-computer interfaces. It was used for the first time for the deconvolution of saccadic potentials during free exploration (Dandekar et al., 2012), and for the deconvolution of EFRP during a visual search task on natural scenes in (Devillez, Kristensen, Guyader, Rivet, & Guérin-Dugué, 2015).

In (Lalor et al., 2006; Rivet & Souloumiac, 2007; Rivet et al., 2009; Crosse et al., 2015), the least-square method deconvolved overlapping responses assuming that only a unique class of response was present: it only tackled the issue of intra-class overlapping response (i.e., overlapping of responses to the same class of stimulus). If several classes of stimuli are simultaneously present and overlap each other (i.e., inter-class overlapping responses), they must be considered separately by ignoring all the other ones. The least-square method proposed by Rivet and colleagues in the first stage of the xDAWN algorithm (Rivet, Souloumiac, Gibert, & Attina, 2008; Rivet et al., 2009; Rivet & Souloumiac, 2013), or in (Bardy, Dillon, & Van Dun, 2014; Burns et al., 2013; Congedo et al., 2016) extended the previously mentioned methods to the case of intra- and/or inter-class overlapping issues. It belongs to the general linear model (GLM) (Kiebel & Holmes, 2003) that provides a very large and flexible framework, routinely used for fMRI studies. In fMRI literature, the GLM is used for the deconvolution of hemodynamic responses (Dale, 1999). In the context of EEG studies, the linear analysis of multidimensional EEG signals has been also proposed (Parra, Spence, Gerson, & Sajda, 2005) as a general framework to extract the underlying neural sources or subtract interfering sources like eye movements. Linear discriminant analysis, principal component analysis and independent component analysis were revisited in this context. The linear model between the latent neural sources and the scalp potentials was addressed and solved using only the statistics of observed data with the stimuli and behavioral responses. The deconvolution of temporally overlapping components was mentioned as an application of the proposed methodology but the developments and examples given in (Parra et al., 2005) mainly concerned the estimation of spatial filters for underlying neural source extractions.

As in the work applied to the auditory evoked potentials (Bardy, Dillon, & Van Dun, 2014); (Bardy, Van Dun, et al., 2014), we plan to use the GLM for the deconvolution of the time-locked evoked potentials and thus manage the intra- and/or inter-class overlapping issues. With this GLM, different situations are addressed where the overlapping components may or may not be of the same nature. Besides in our work, which aims to improve the SNR for a given number of trials, the Tikhonov regularization is added to solve the GLM.

In “**Theoretical developments**”, the matrix framework for the usual ERP estimation is presented in more detail followed by the introduction of the regularization parameter and the GLM with the Tikhonov regularization. In “**Methodology**”, the methodology to assess the proposed algorithms is explained, based on artificial data simulations but also on real EEG signals from two experiments (1) free visual exploration to estimate EFRP, and (2) a “P300 Speller” brain

computer interface. The results based on simulated and real data are presented and discussed in “[Results on simulated signals](#)”, “[Results on EFRP estimation during a free visual exploration](#)”, and “[Results on P300 Speller data](#)”, respectively. Finally, “[Conclusions](#)” summarizes the main results achieved and concludes this article.

Theoretical developments

In this section, we first explain the usual ERP estimation: the “Average” method. Then we introduce a matrix framework to provide the regularization procedure and finally, the GLM is described in the context of the temporal deconvolution of overlapped evoked potentials.

Usual event-related potential estimation

To extract the evoked potentials, the temporal window of epochs must be sufficiently long to include the latency and the whole temporal evolution of the potentials of interest. The underlying assumption is as follows: at each presentation of a same stimulus, the same neural potential is evoked. Thus for the i^{th} presentation, the observed neural response $x_i(t)$ time-locked on the i^{th} stimulus onset can be written as:

$$x_i(t) = a(t) + n_i(t) \quad (1)$$

where $a(t)$ is the evoked potential, and $n_i(t)$ a noise corresponding to the on-going brain activity.

As the signal-to-noise ratio is very low, the stimuli have to be repeated several times. In the simplest configuration, one epoch is extracted at each trial. In this way, $a(t)$ is estimated by averaging across these epochs:

$$\hat{a}(t) = \frac{1}{E} \sum_{i=1}^E x_i(t) \quad (2)$$

where E is the number of epochs.

This estimation is non-biased if the following conditions are fulfilled:

- only one deterministic potential, $a(t)$, is elicited during the epoch synchronized with the stimulus;
- this potential has the same amplitude across the epochs;
- the noise is a zero-mean random variable and decorrelated from one trial to another one. As a consequence, the SNR increases linearly with E .

In “[Regularization to reduce the number of trials](#)” and “[General linear model used to manage the overlapping issue](#)”, we will explain why these conditions are not always verified in practice, but in this section, we first consider that the two first conditions have been fulfilled. To facilitate and justify

the proposed theoretical developments, let us reformulate Eq. 2 in a matrix form.

For this purpose, Eq. 1 can be rewritten using matrix notations as:

$$\forall i \in \{1, \dots, N_e\}, \quad \underline{x}_i = \mathbf{D}_i \underline{a} + \underline{n}_i \quad (3)$$

where $\underline{x}_i = [x_i(1), \dots, x_i(N_e)]^\dagger \in \mathbb{R}^{N_e}$ and $\underline{n}_i = [n_i(1), \dots, n_i(N_e)]^\dagger \in \mathbb{R}^{N_e}$ with \cdot^\dagger the transpose operator. $\underline{a} \in \mathbb{R}^{N_a}$ is the vector of the response time-locked on the stimuli onset. N_e is the length of the epoch and N_a is the length of the response $a(t)$. Finally, $\mathbf{D}_i \in \mathbb{R}^{N_e \times N_a}$ is a Toeplitz matrix¹ defined by its first column with entries are all equal to zero except the τ_i^{th} entry equal to one, with τ_i^{th} the onset of the i^{th} epoch. Consequently, the epochs can be concatenated to obtain:

$$\underline{x} = \mathbf{D} \underline{a} + \underline{n}, \quad (4)$$

with $\underline{x} = [\underline{x}_1^\dagger, \dots, \underline{x}_{N_e}^\dagger]^\dagger \in \mathbb{R}^N$ and $\mathbf{D} = [\mathbf{D}_1^\dagger, \dots, \mathbf{D}_{N_e}^\dagger]^\dagger \in \mathbb{R}^{N \times N_a}$ where $N = N_e \times E$. \mathbf{D} is the concatenation of the matrices \mathbf{D}_i with $i = [1, \dots, N_e]$. This equation can be understood as the general linear model (GLM) with the matrix \mathbf{D} as predictors, as it will be developed in “[General linear model used to manage the overlapping issue](#)”. Thus, with these notations, the neural response $\hat{\underline{a}}$ is the least square solution minimizing the noise variance, such as:

$$\hat{\underline{a}} = \arg \min_a \|\underline{x} - \mathbf{D} \underline{a}\|^2 \quad (5)$$

where $\|\cdot\|$ indicates the Euclidean norm. This general solution is then expressed as:

$$\hat{\underline{a}} = (\mathbf{D}^\dagger \mathbf{D})^{-1} \mathbf{D}^\dagger \underline{x} \quad (6)$$

and can be here simplified such as:

$$\hat{\underline{a}} = \frac{1}{E} \mathbf{D}^\dagger \underline{x} \quad (7)$$

since the matrix $\mathbf{D}^\dagger \mathbf{D}$ is a diagonal one whose diagonal elements are all equal to E . Effectively, each matrix \mathbf{D}_i is composed of only one diagonal with values of one because only one potential is elicited at the onset of the stimulus and then $\mathbf{D}^\dagger \mathbf{D} = E \mathbf{I}$, with \mathbf{I} the identity matrix. Figure 1 illustrates this situation, considering the time interval between two consecutive stimuli (ISI) to be greater than or equal to N_e samples. This configuration of the GLM with a diagonal matrix $\mathbf{D}^\dagger \mathbf{D}$ is called here the “Average” method corresponding with the usual estimation by averaging on the time-locked onsets.

¹By definition, a Toeplitz matrix is a descending diagonal-constant matrix.

Regularization to reduce the number of trials

By averaging over all epochs, the consistent ERP waveform is extracted above the noise level. Considering a same variance for the noise decorrelated with the neural response for each trial, the variance of the residual noise is inversely proportional to the number E of epochs and thus the SNR linearly increases with the number E of epochs. However, this number is often a compromise between the signal quality and the duration of the experiment. So, an alternative is to introduce a regularization constraint to estimate smooth solutions. Then for a given number of epochs, the SNR can be improved by regularization. For this, a zero-order Tikhonov regularization is implemented using a ridge parameter λ in the ERP estimation, such as $\hat{\underline{a}}_{reg}(\lambda)$ is the least-square solution minimizing this cost function:

$$CF(\hat{\underline{a}}_{reg}(\lambda)) = \|\underline{x} - \mathbf{D}\underline{a}(\lambda)\|^2 + \lambda N \|\hat{\underline{a}}_{reg}(\lambda)\|^2 \quad (8)$$

with N the total number of samples ($N = N_e \times E$). Consequently, $\hat{\underline{a}}_{reg}(\lambda)$ can be expressed by:

$$\hat{\underline{a}}_{reg}(\lambda) = (\mathbf{D}^\dagger \mathbf{D} + \lambda N \mathbf{I})^{-1} \mathbf{D}^\dagger \underline{x} \quad (9)$$

It should be noted that the classical estimation (6) is for $\lambda = 0$ (i.e., $\lambda = 0$, no regularization).

A correct regularization is a trade-off between the error of the estimation ($\|\underline{x} - \mathbf{D}\hat{\underline{a}}_{reg}(\lambda)\|$) and its power ($\frac{1}{N_a} \|\hat{\underline{a}}_{reg}(\lambda)\|^2$). A low value of λ promotes an estimate with a low error but this solution may be dominated by data errors. Conversely, a high value of λ promotes a low power estimate with a low variance relatively to the fluctuation of the observed data but a high average gap with the observed data. These two quantities must be controlled to find the best trade-off. Usually possible values for λ are iteratively set by plotting for increasing values of λ , the “L-curve” which would be, here, $\|\underline{x} - \mathbf{D}\hat{\underline{a}}_{reg}(\lambda)\|$ against $\|\hat{\underline{a}}_{reg}(\lambda)\|^2$ in a log-log graph (Engl & Grever, 1994; Hansen, 1992). Then, the optimal value, λ_{opt} , is chosen at the “corner of the L” corresponding to the maximum of curvature of the “L-curve” but this “corner” can be difficult to discern (Subramaniyam et al., 2010). To overcome this difficulty in objectively finding the “corner of the L”, we opted for a one-step method by using the generalized cross validation (GCV) (Golub et al., 1979) instead of the L-curve method. In short, the GCV considers the linear regression model and is based on a weighted version of cross-validation. In this case, the cross validation uses the “leave one out” strategy by which each observed point is left out in turn and is estimated by the rest of data. The aim of the GCV is to estimate

the λ_{GCV} , which minimizes the function V related to the mean square error (MSE). The λ_{GCV} is given by:

$$V(\lambda) = \frac{\frac{1}{N} \|\underline{x} - \mathbf{D}\hat{\underline{a}}_{reg}(\lambda)\|^2}{\left(\frac{1}{N} \text{tr}(\mathbf{I} - \mathbf{D}(\mathbf{D}^\dagger \mathbf{D} + \lambda N \mathbf{I})^{-1} \mathbf{D}^\dagger)\right)^2} \quad (10)$$

$$\lambda_{GCV} = \arg \min_{\lambda} V(\lambda)$$

In the ERP technique, the study is generally an inter-subject study and the final ERP is estimated after a “grand average” (average across all the subjects). Consequently, there needs to be an overall regularization which considers each subject. A single λ_{GCV} is computed for the subjects. So, using the matrix framework, it is possible to consider all the subjects by the concatenation of \underline{x}_i signals:

$$\underline{x}_{all} = \mathbf{D}_{all} \underline{a} + \underline{n}_{all} \quad (11)$$

with $\underline{x}_{all} = [\underline{x}_1^\dagger, \dots, \underline{x}_S^\dagger]^\dagger \in \mathbb{R}^{N_s}$, $\underline{n}_{all} = [\underline{n}_1^\dagger, \dots, \underline{n}_S^\dagger]^\dagger \in \mathbb{R}^{N_s}$ and $\mathbf{D}_{all} = [\mathbf{D}_1^\dagger, \dots, \mathbf{D}_S^\dagger]^\dagger \in \mathbb{R}^{N_s \times N_a}$ where S is the number of subjects and $N_s = S \times N$ is the total number of temporal samples. In this way, the “grand average” can be estimated by Eq. 9 with an unique value of λ_{GCV} for all subjects, using \mathbf{D}_{all} , \underline{x}_{all} and N_s , instead of \mathbf{D} , \underline{x} and N .

General linear model used to manage the overlapping issue

The classical ERP method estimates only one evoked potential by epoch. Two situations can occur regarding overlapping. The first is when the experiment has been designed so that a single potential is elicited at each epoch (Fig. 1). In this case, the simplified model (6) where $\mathbf{D}^\dagger \mathbf{D}$ is really a diagonal matrix, is justified. The second situation is when during an epoch, the evoked response is observed along with a part of the previous and/or the subsequent response (Fig. 2). If the ISI duration is significantly shorter than the latency of the evoked potential and the duration of the response, the amount of distortion due to overlapping can no longer be neglected. Consequently, the right model to take these overlaps into account needs to be built with a non-diagonal matrix $\mathbf{D}^\dagger \mathbf{D}$. Its setup will now be explained.

By using the general linear model to take into account overlapping, the Toeplitz matrix \mathbf{D} has to be set considering all stimuli inside each epoch. To do so, the neural responses can be deconvolved (Rivet et al., 2009; Bardy et al., 2014; Congedo et al., 2016). We first consider that the overlapping neural responses are of the same type and, second, that they are of different types.

Therefore, inside one epoch, there is more than one elicited response, but let us consider that all of these neural responses have the same waveform. Overlaps occur because the ISI is shorter than both the latency of the expected neural response \underline{a} and the duration N_e of the epochs. Consequently,

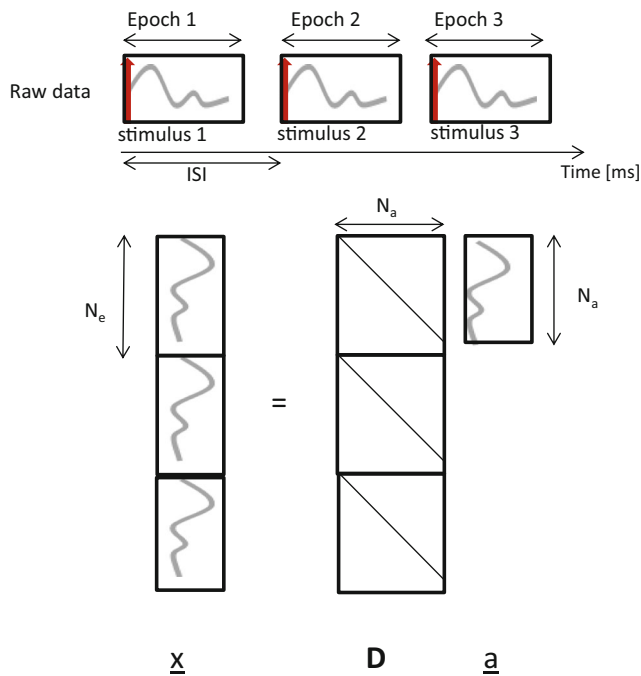


Fig. 1 Construction of the matrix **D** considering one unique neural response per trial (4). Here, a same stimulus (red arrow) repeated three times, elicits an unique ERP, \underline{a} . The raw data are segmented into three epochs (black boxes) time-locked on the stimulus onset. The vector \underline{x} is the concatenation of these epochs. The Toeplitz matrix **D** is null except on the diagonals whose entries are synchronized on the onset of each stimulus. The noise \underline{n} is not illustrated

there is more than one diagonal in each submatrix \mathbf{D}_i , as is illustrated in Fig. 2—the solid line diagonals correspond to the neural responses elicited at the onset of the stimuli (i.e.,

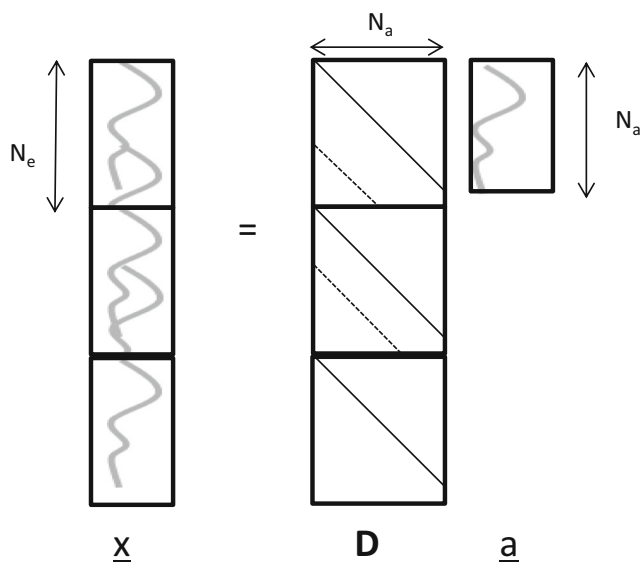


Fig. 2 Construction of the matrix **D** assuming the same evoked potential \underline{a} for all the neural responses in each epoch but with shorter ISIs. The subdiagonals (dotted lines) of the matrix **D** correspond to the other neural responses than the neural response on which the epoch is time-locked (solid lines). The noise \underline{n} is not illustrated

synchronized with them) and the dotted line subdiagonals correspond to another neural response elicited, by another stimulus, in the same epochs. These subdiagonals may or may not be complete depending on the choice of duration of the epoch (N_e) in relation to the duration of the response (N_a) and the ISI values. We also need to remember that even in this situation, the classical estimation of the neural response only takes into account the diagonals of matrix **D** at the onset of the stimuli. Also, the contributions of the previous and subsequent responses are mixed with the time-locked observed response \underline{x} and the potential $\hat{\underline{a}}$ estimated using Eq. 7 ($\mathbf{D}^\dagger \mathbf{D}$ is a diagonal matrix), is biased by these overlaps. To take overlapping into account, the linear model of the observed data, described by Eq. 6, is used with the Toeplitz matrix **D** as described in Fig. 2. Thus, the least-square minimization to estimate the neural response $\hat{\underline{a}}$ as the best linear non-biased estimator with the smaller variance is defined by Eq. 5, and the solution $\hat{\underline{a}}$, is computed by Eq. 6.

Now, let us consider that several types of response are evoked during one same epoch. In this situation, the linear model of the observed data is extended to take each evoked potential in the observed data into account. Let us note C the number of different expected types of response. In the following, C will also be called the number of classes. Thus, the model of the observed data can be rewritten as:

$$\underline{x} = \sum_{c=1}^C \mathbf{D}^{(c)} \underline{a}^{(c)} + \underline{n}$$

$$\underline{x} = \mathbf{D} \underline{a} + \underline{n} \tag{12}$$

where **D** is now the concatenation of the matrices $\mathbf{D}^{(c)}$ ($\mathbf{D} = [\mathbf{D}^{(1)}, \dots, \mathbf{D}^{(C)}]$), and \underline{a} is also the concatenation of the evoked potentials ($\underline{a} = [\underline{a}^{(1)\dagger}, \dots, \underline{a}^{(C)\dagger}]^\dagger$). In this way, all neural potentials inside each epoch are considered instead of a single potential on which the epoch is time-locked.

Some precautions must be taken to avoid an ill-conditioned and noninvertible matrix $\mathbf{D}^\dagger \mathbf{D}$: if the variability of the ISI values (referred to as jitter) is too low, it is not possible to accurately separate the contribution of the different stimuli in the observation of the neural response \underline{x} : therefore, the matrix $\mathbf{D}^\dagger \mathbf{D}$ is ill-conditioned. The amount of jitter must be sufficient to enable the deconvolution between the different types of evoked responses (see Bardy, Dillon, & Van Dun, 2014 for a detailed discussion about the amount of jitter as a balance in the case of the deconvolution of cortical auditory potentials). A minimum amount of jitter is necessary, but if it is too large, the variability of the neural response may increase, because the auditory response is affected by the ISI value from the previous stimulus (Näätänen et al., 1988). Consequently, the neural response is no longer time-invariant, and Eq. 6 becomes incorrect. Besides, the length of the responses to be estimated must be smaller than the number of observed samples ($N_a < N$).

To assess the estimation of one potential relative to the others, we used the SNR, the signal-to-interference ratio (SIR) and the signal-to-artefact ratio (SAR) as they are commonly used in the “Blind Sources Separation” community (Vincent, Gribonval, & Févotte, 2006). Let Eq. 12 be recast in the case of two potentials, \underline{a}_1 and \underline{a}_2 as:

$$\underline{x} = \mathbf{D}^{(1)}\underline{a}_1 + \mathbf{D}^{(2)}\underline{a}_2 + \underline{n} \tag{13}$$

where $\mathbf{D}^{(1)}$ can be split into two parts: $\mathbf{D}^{(1)} = \mathbf{D}_{erp}^{(1)} + \mathbf{D}_{ov}^{(1)}$, with the matrix $\mathbf{D}_{erp}^{(1)}$ constructed by the synchronized onsets and the matrix $\mathbf{D}_{ov}^{(1)}$ constructed by the other neural responses inside the epoch. In the case of estimating \underline{a}_1 , one can write the estimate as:

$$\hat{\underline{a}}_1 = \underline{a}_1 + \underline{Q}_{vl}(a_1) + \underline{Q}_{vl}(a_2) + \underline{n}' \tag{14}$$

where $\underline{Q}_{vl}(a_1)$, $\underline{Q}_{vl}(a_2)$ and \underline{n}' are the remaining overlaps due to \underline{a}_1 and \underline{a}_2 and the remaining noise, respectively. This expression depends on the algorithm used as detailed below (see Appendix 1 for the theoretical developments of these indicators for the simple “Average” method, and for the general linear model. From Eq. 14, one can define the SNR of the estimate $\hat{\underline{a}}_1$ as:

$$\text{SNR}_{\hat{\underline{a}}_1} = 10\log_{10} \left(\frac{\|\underline{a}_1\|^2}{\|\underline{n}'\|^2} \right) \tag{15}$$

This index quantifies the remaining noise in \underline{a}_1 . The higher this is, the better.

The signal-to-artefact ratio (SAR), which quantifies the remaining overlap of \underline{a}_1 in $\hat{\underline{a}}_1$, is defined as:

$$\text{SAR}_{\hat{\underline{a}}_1} = 10\log_{10} \left(\frac{\|\underline{a}_1\|^2}{\|\underline{Q}_{vl}(a_1)\|^2} \right) \tag{16}$$

Again, the higher, the better. Finally, the signal-to-interference ratio (SIR) that quantifies the remaining overlap of \underline{a}_2 in $\hat{\underline{a}}_1$, is defined as:

$$\text{SIR}_{a_1/a_2} = 10\log_{10} \left(\frac{\|\underline{a}_1\|^2}{\|\underline{Q}_{vl}(a_2)\|^2} \right) \tag{17}$$

And yet again, the higher, the better.

Regularization applied to the general linear model

Finally, this formalism serves to set the regularization to reduce the estimation variance overall the classes:

$$\hat{\underline{a}}_{reg}(\lambda) = (\mathbf{D}^\dagger \mathbf{D} + \lambda \mathbf{N} \mathbf{I})^{-1} \mathbf{D}^\dagger \underline{x} \tag{18}$$

This Eq. 18 corresponds to Eq. 9, with the concatenation of the Toeplitz matrices as illustrated in Fig. 3, except that here, the full matrix $\mathbf{D}^\dagger \mathbf{D}$ is no longer a diagonal matrix.

Finally, as previously, λ_{GCv} is computed according to Eq. 10. In the framework of an inter-subject study, combining GLM and regularization is implemented in the same way as in “Regularization to reduce the number of trials” where a single value of λ_{GCv} is computed for all the subjects.

Methodology

To evaluate the efficiency of the proposed approach combining GLM and regularization, a benchmark with four models was set. We shall denote the four models as follows: on the one hand $Av(0)$ the classical “Average” method, and $Gl m(0)$ the GLM, without regularization ($\lambda = 0$), and on the other hand $Av(\lambda)$ the regularized classical “Average”

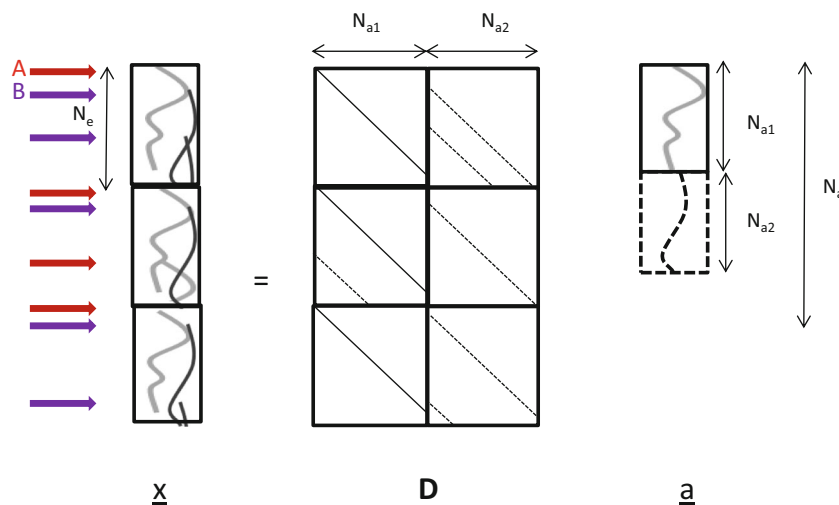


Fig. 3 Construction of the matrix \mathbf{D} assuming two ($C = 2$) different evoked potentials inside the epochs. The potential \underline{a}_1 (resp. \underline{a}_2) is elicited by the stimulus A (red arrow) (resp. stimulus B (purple arrow)). Here, the epochs are time-locked on the stimulus A 's onset. The diagonals and subdiagonals of the matrix D_1 (resp. D_2) are defined according to the onsets of the stimulus A (resp. B) for each epoch. The noise \underline{n} is not illustrated

method and $Glm(\lambda)$ the regularized GLM. To conduct this benchmark, a two-step methodology was followed. Firstly, realistic EEG signals were generated in order to obtain a ground truth by using controlled parameters. Secondly, real EEG signals from two experiments (visual scene exploration, and a classical P300 Speller paradigm) were used to assess the efficiency of the proposed procedures in real situations. Using these two databases, two different configurations for GLM were illustrated. In the first experiment, the GLM was used to estimate EFRP elicited on adjacent fixations, according to a simple assumption where the evoked potentials were identical at each fixation. In other words, the GLM was configured with only one class to estimate overlapping potentials. In the second experiment, the GLM was configured with two classes to highlight the specifics of a given class. These two experiments used an EEG acquisition device. The simulated and real signals are now presented.

Simulated signals

The EEG signals were generated as described in Eq. 1. The potential $a(t)$ was generated with an early waveform $a_{early}(t)$ preceding a late one $a_{late}(t)$. This potential was the ground truth. $a_{early}(t)$ was generated as a white noise filtered by a bandpass filter with a bandwidth between 5 and 10 Hz and multiplied by a Gaussian window whose mean μ was 300 ms and the standard deviation σ , 125 ms. In the same way, $a_{late}(t)$ was a white noise filtered by a low-pass filter with a cutoff of 3 Hz multiplied by a Gaussian window with μ equal to 600 ms and σ to 100 ms. To simulate the ongoing brain activity, some noise, $n(t)$ was added to $a(t)$. $n(t)$ was a white noise filtered by a low-pass filter with a cut-off of 50 Hz. For our simulations, the SNR varied between -20 and 0 dB. The sampling frequency was 1000 Hz, and the length of the temporal observation window was 1000 ms.

Three configurations were studied. In the first, without overlapping, each stimulus was presented with the same ISI value equal to 1000 ms and elicited the unique potential $a(t)$. In a second condition—with overlapping—the stimuli were presented with a random ISI value shorter than the latency of the potential. This ISI value was a uniform random variable between 200 and 400 ms. In the third configuration—with overlapping and multiple responses—two different stimuli were presented, eliciting respectively two potentials, $a_1(t)$ and $a_2(t)$. The ISI values were generated in a similar way, by a uniform random variable between 200 and 400 ms. For the three situations, data were segmented into epochs of 1000 ms time-locked on each stimulus onset. For the two last configurations with overlaps, there were on average, four stimuli, and the averaged ISI value during the epoch was 300 ms. The number E of epochs varied between 10 and 100. These simulations were used to assess the different methods in both situations with

a single evoked potential $a(t)$, and then with two evoked potentials, $a_1(t)$ and $a_2(t)$.

Real signals from a visual scene exploration

During a visual scene exploration, a joint EEG and eye movement database was recorded. For the purpose of this study, only one condition of the whole experiment (Devillez, Guyader, & Guérin-Dugué, 2015), called “the free exploration”, which consisted of freely exploring 60 color scenes, was considered. Each trial started with a white central fixation cross displayed for 800–1200 ms. When the participant had stabilized his/her gaze on this central fixation point, a color scene was displayed for 4 s. Each trial ended with a grey screen for 1 s. Thirty-nine healthy volunteers between 20 and 36 years old participated in the experiment. Eye movements were recorded using Eyelink 1000 (SR Research) and sampled at 1000 Hz, for both eyes. The EEG activity was recorded using 32 active electrodes. The right earlobe and FCz electrodes were used respectively as reference and ground. Data were amplified using a g.USBamp gtec system, and sampled at 1200 Hz. An analog bandpass filter (0.01 – 100 Hz) and a 50-Hz notch filter were applied online. Eye movement and EEG signals were synchronized offline, thanks to the triggers sent simultaneously to both the EEG system and the eye-tracker. EEG data were then re-sampled at 1000 Hz (eye-tracker sampling rate). After visual inspection to reject segments contaminated by muscular activity or non-physiological artefacts, ocular artefacts were then corrected by independent component analysis (infomax ICA) (Bell & Sejnowski, 1995).

Since the task was a free exploration, each ocular fixation was assumed to elicit a similar potential, such as the lambda wave at latency of 100 ms (Yagi, 1981) and the P2 potential at latencies 200 – 300 ms, whatever the fixation rank during the exploration. This EFRP was estimated from -200 ms (including the pre-saccadic potential) up to 800 ms, which was a sufficiently long duration to include these early potentials. For the present study, for each trial, three² ranks of fixation (r_1, r_2, r_3)³ were randomly selected for a given subject. An epoch was defined as the temporal window time-locked on the onset of an ocular fixation, the rank of this fixation was r_1, r_2 or r_3 . So for a given subject, there were 180 epochs (three per scene) at most. For a given subject, the temporal window was defined as $[-200 - \tau; 800 + \tau]$ ms with $\tau = \mu(\text{IFI}) + \sigma(\text{IFI})$ considering the mean and the standard deviation of the distribution of inter fixation

²The number of random draws was fixed at three as a balance between a great number of epochs, and the memory needed for the computation.

³The first two fixations were not considered because their potential was influenced by the potential elicited by the onset of the scene as well as the last fixation, which could be interrupted by the offset of the scene at the end of the exploration.

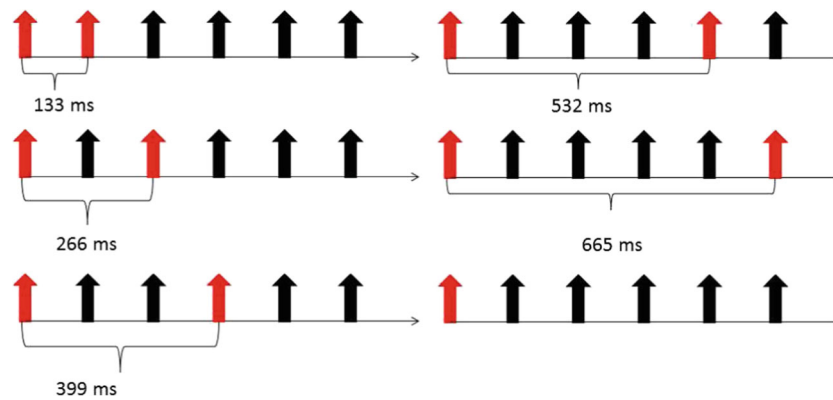


Fig. 4 The six possible configurations of epochs. The *red arrows* (resp. *black arrows*) represent target (resp. non-target) stimuli

interval values. In this way, the fixation on which the epoch was time-locked was preceded and followed by at least a fixation. By averaging across the participants, the temporal window was equal to $[-572; 1171]$ ms and the mean of the inter-fixation interval was equal to 271 ms (std = 91 ms).

In our study, this database was of interest to illustrate the overlapping issue that was one of the main concerns for the eye fixation related potential estimation (Dimigen et al., 2011). Besides, a study was done on real datasets to assess the regularization procedure in two cases. Firstly, the regularization procedure was independently carried out on each subject's dataset, and secondly, the regularization was simultaneously applied on the whole dataset. Moreover, the estimated regularized component vs. the grand average of the regularized components for each subject was compared.

Real signals from a P300 Speller experiment

An EEG database was recorded during a P300 Speller experiment. The P300 Speller is a brain–computer interface based on the oddball paradigm (Farwell & Donchin, 1988) which spells characters. A 6×6 matrix was displayed on a screen computer. The participant focused her/his attention on the target symbol that she/he wanted to spell and had to count how many times this symbol was flashed. Each column and row was randomly intensified several times. According to the oddball paradigm, when the target symbol flashes, a P300 wave is elicited whereas the non-target stimuli elicit only a sensory potential. In this experiment, there were 500 intensifications (or stimuli) of target characters and 3000 intensifications of non-target characters. The ISI value was equal to 133 ms.

Ten healthy volunteers, between 22 and 34 years old, participated in this experiment. The EEG signal was recorded via 16 active electrodes with the g.USBamp device from the g.tec company. The reference electrode was on the ear lobe and the ground electrode was on the forehead. The sampling frequency was 1200 Hz. The signal was filtered by

a four-order Butterworth band-pass filter with a bandwidth between 1 and 12 Hz.⁴

An epoch was defined as the temporal window time-locked on a target stimulus. This window was from -0.5 s to 1 s and thus, within an epoch, each target stimulus was followed by at least five stimuli. There were six possible configurations of epoch as shown in Fig. 4, and there were 500 epochs.

This database was interesting for our study firstly because there were two types of stimuli and thus two types of neural responses and, secondly because the ISI value was shorter than the latency of the expected target evoked potential P300 (300–500 ms). So some overlaps occurred. For the purposes of our study, the methods were assessed for different numbers of trials with $T = [60, 120, 180, 240, 300, 360, 420]$. For each value of T , the epochs were pseudo-randomly chosen in such a way that each configuration gave the same number of epochs (i.e., for $T = 60$, ten epochs from each configuration).

Results on simulated signals

Effect of the regularization parameter

In this section, the effect of the ridge parameter λ is studied when estimating the evoked potential in the simplest case, without overlapping. Since the ground truth is known, the mean square error (MSE) between $a(t)$ and $\hat{a}_{reg}(\lambda)(t)$:

$$\text{MSE}(\lambda) = \frac{1}{N_a} \|\hat{a}_{reg}(\lambda)(t) - a(t)\|^2 \quad (19)$$

can be evaluated as a function of λ (Fig. 5), with a SNR equal to -20 dB and a number of trials equal to 50 and to 100.

⁴12 Hz is a very low cut-off frequency chosen for these data, which were initially recorded for a classification study.

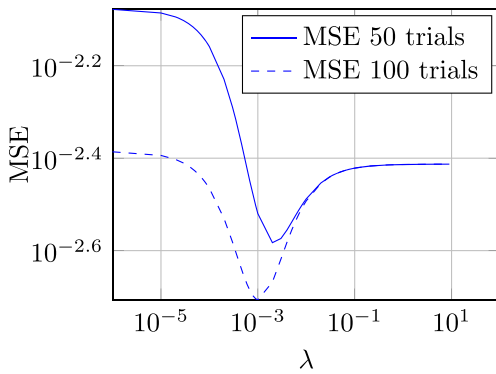


Fig. 5 Evolution of averaged MSE with respect to λ , for SNR = -20 dB, 50 and 100 trials

From these two curves, three areas were clearly noticeable:

- $0 \leq \lambda \leq 10^{-4}$: The regularization was null or very low. The two MSE curves reached a high MSE for $\lambda = 0$ (without regularization) but led a non-biased estimation. This value for $\lambda = 0$ was larger with 50 trials than with 100 because when averaging, the SNR was higher with more trials.
- $10^{-4} < \lambda \leq 10^{-2}$: the MSE decreased to a minimal value corresponding to the optimal λ value (λ_{MSE}). As expected, λ_{MSE} with 50 trials ($\lambda_{MSE} = 2.10^{-3}$) was greater than with 100 trials ($\lambda_{MSE} = 1.10^{-3}$). Indeed, with less trials, the estimated component was less noisy and required more regularization to obtain a smooth estimation.

- $\lambda > 10^{-2}$: The MSE criterion increased and converged towards the same value, independently of the number of trials. According to Eq. 18, $\hat{a}_{reg}(\lambda)(t)$ converged towards zero when λ increased. Consequently, the MSE converged towards the power of $a(t)$ ($\frac{\|a(t)\|^2}{N_a}$).

These three areas are each illustrated by an estimated waveform at Fig. 6 for 50 trials and SNR = -20 dB. On these three graphs, the theoretical expectation $E[\hat{a}]$ and the variability (± 2 standard deviations) are plotted in relation to the ground truth $a(t)$ (see Appendix 2 for the mathematical expressions of the theoretical expectation and variance corresponding to the estimation by Eq. 18). For $\lambda = 0$ (Fig. 6a), the estimation was unbiased (the ground truth $a(t)$ and $E[\hat{a}]$ are identical) but the variance due to the residual noise was large. When the ridge parameter was at the optimal value $\lambda = \lambda_{MSE} = 2.10^{-3}$ (Fig. 6b), the bias was greater but the variance was lower than for the estimation without regularization. In this way, the estimated signal was less noisy, which made this the best compromise between bias and variance. Finally, for a high λ , greater than the optimal value (here $\lambda = 1$), the estimated signal became unusable (Fig. 6c), $\hat{a}_{reg}(t) = 0$. In other words, with a value which is greater than the optimum, the regularization is too strong.

Generalized cross-validation procedure

The aim of this section is to validate the generalized cross validation procedure to estimate the ridge parameter λ when computing the evoked potential $a(t)$ in the simplest case,

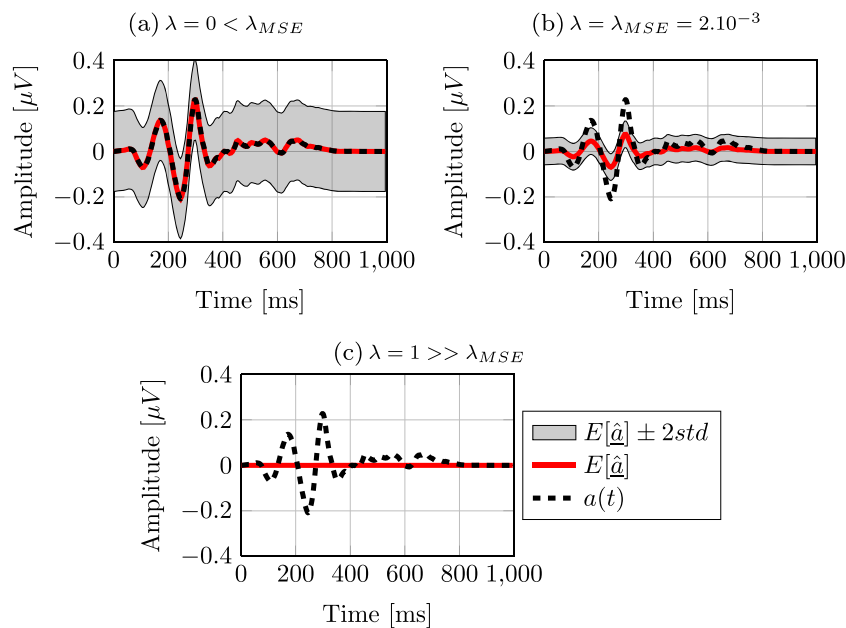


Fig. 6 Illustration of the estimated potential $\hat{a}_{reg}(\lambda)(t)$ with (a) $\lambda = 0$ (no regularization), (b) $\lambda = \lambda_{MSE} = 2.10^{-3}$ (optimal regularization), and (c) $\lambda = 1$ (over-regularized). These three estimations were computed for the same configuration: SNR = -20 dB and 50 trials

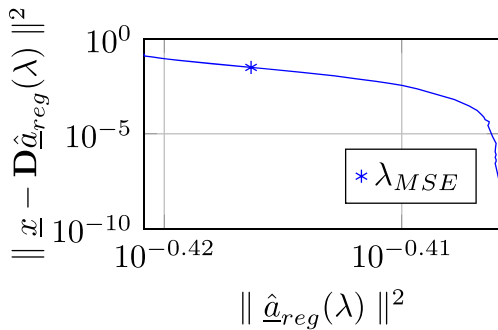


Fig. 7 “L-curve” for 50 trials and SNR = -20dB, $\| \hat{x} - \mathbf{D} \hat{a}_{reg}(\lambda) \|^2$ against $\| \hat{a}_{reg}(\lambda) \|^2$

i.e., a unique potential without overlapping. To begin with, Fig. 7 confirms as in (Subramaniyam et al., 2010) that it is difficult to use the “L-curve” method to choose the optimal λ value. In this figure, the value of λ , λ_{MSE} , which minimizes the MSE, is indicated but does not correspond to a discernible “corner of the L-curve”.

The value of λ given by the minimum of the GCV function, λ_{GCV} , was compared to the value which minimized the MSE, λ_{MSE} . The GCV function, V (10), is plotted for 50 trials and SNR = -20 dB in Fig. 8. Both of the values of λ_{MSE} and the values of λ_{GCV} , are indicated on this graph and are close to each other: $\lambda_{MSE} = 2.10^{-3}$ and $\lambda_{GCV} = 1.10^{-3}$.

In Fig. 9, both the averaged optimal values ($\bar{\lambda}_{MSE}$, $\bar{\lambda}_{GCV}$) computed across ten realizations, are plotted for two SNR values (-20 dB and -10 dB) against the number of trials.

In most cases, these two estimates of the ridge parameter were equivalent. With less noise (SNR = -10 dB), these two estimates were similar even for few trials (≥ 10). However, with more noise (here SNR = -20 dB), the estimation of λ by GCV was over-evaluated if there are few trials. In fact, in the GCV process for the regularization of a signal with a very low SNR, the regularization of the noise outweighs the regularization of the target signal. As a result, this signal is too smoothed in these conditions. Figure 9 also shows that λ decreases with the SNR and with

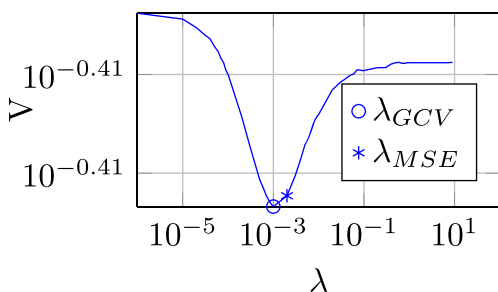


Fig. 8 The GCV function, $V(\lambda)$, for 50 trials and SNR = -20 dB

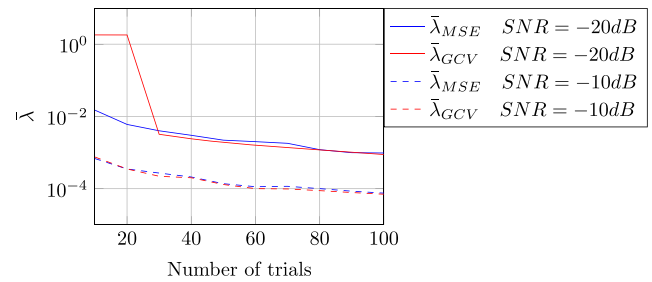


Fig. 9 Averaged optimal ridge parameter given by minimization of the MSE criterion $\bar{\lambda}_{MSE}$ and by minimization of V , $\bar{\lambda}_{GCV}$, function of the number of trials

an increasing number of trials. The less noisy the signal, the less regularization is necessary.

For the following sections, and for all results with regularization, only the GCV method is used to compute the ridge parameter.

Application to overlapped potentials

In this section, the efficiency of the two methods (“Average” and GLM) with regularization ($Av(\lambda_{GCV})$, $Glm(\lambda_{GCV})$) or without regularization ($Av(0)$, $Glm(0)$) is assessed in the case of overlapping potentials, providing four estimation algorithms to benchmark.

First, the situation with only one overlapping potential a , was studied. The four estimated waveforms are illustrated in Fig. 10a ($Av(0)$), 10b ($Glm(0)$), 10c ($Av(\lambda_{GCV})$) and 10d ($Glm(\lambda_{GCV})$). As expected, the GLM estimation without regularization ($Glm(0)$) was an unbiased estimation (Fig. 10b), because the overlap was taken into account in the model (6). The classical estimation by averaging ($Av(0)$) was biased (Fig. 10a): the matrix $\mathbf{D}^\dagger \mathbf{D}$ in the model was only a diagonal matrix (6). Furthermore, the contributions, due to overlap, were not taken into account. Without regularization, the classical estimation by averaging ($Av(0)$) gave a larger variance than the GLM estimation ($Glm(0)$) as shown in Fig. 10a and c. Regularization was found to decrease the variance while increasing the bias. This is what we observed for the two estimates given by $Av(\lambda_{GCV})$ method and by $Glm(\lambda_{GCV})$, but the bias on the estimate potential by averaging ($Av(\lambda_{GCV})$) (Fig. 10c) remained larger than the bias on the estimate by GLM ($Glm(\lambda_{GCV})$) (Fig. 10d).

Secondly, the simulations when two overlapping potentials $a_1(t)$ and $a_2(t)$ can be elicited in one epoch were found to produce the same results and interpretations. To avoid repetitions, the corresponding wave forms were not shown. The evolution of λ_{GCV} and of the estimated SNR (see Eq. 15 and Appendix 1), function of the number of trials, were studied. The evolution of λ_{GCV} , for SNR = -20 dB is plotted in Fig. 11. For more trials, the values for the two models were equivalent and decreased with the

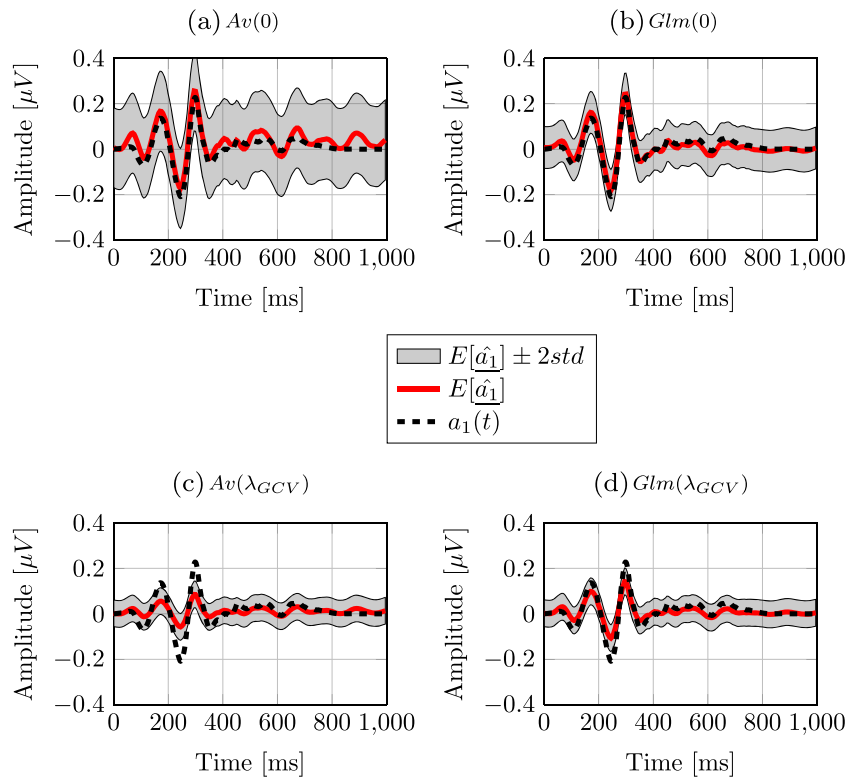


Fig. 10 Illustration of the mathematical expectation $E[\hat{a}]$ in the case of overlapping between epochs. The four models are illustrated: average method and GLM without regularization ($Av(0)$, $Glm(0)$) and with optimal regularization ($Av(\lambda_{GCV})$, $Glm(\lambda_{GCV})$). These four estimations are with the same configuration: SNR = −20 dB and 50 trials

number of trials. This was in line with the evolution of the estimated SNR, plotted against the number of trials in Fig. 12, for the two methods without regularization ($Av(0)$, $Glm(0)$). Both of these SNRs increased with the number of trials, and, consequently, this explains the decreasing of the λ_{GCV} , for the two methods with regularization ($Av(\lambda_{GCV})$, $Glm(\lambda_{GCV})$) in Fig. 11. The higher the SNR, the less regularization is necessary. Consequently, with regularization, the SNR was improved for the two methods for a given number of trials and it was almost constant across the number

of trials. Additionally, the SNR for less trials (i.e., 20), with regularization is better than the SNR for more trials (i.e., 100) without regularization. This is explained by decreased variance using regularization.

Even if the SNR was the same between $Av(0)$ and $Glm(0)$ and, between $Av(\lambda_{GCV})$ and $Glm(\lambda_{GCV})$, the efficiencies of the $Av(0)$ and $Av(\lambda_{GCV})$ models were impacted by distortions due to overlapping. These distortions were evaluated by the signal-to-artifact ratio (SAR) (16) and signal-to-interference ratio (SIR) (17), for the $Av(0)$ and $Glm(0)$ methods (see Appendix 1 for the theoretical developments of these ratios). When only one unique potential is elicited, the SAR assesses the distortions due to the overlaps of the potentials $a_1(t)$ in the estimation of the $a_1(t)$ itself. In

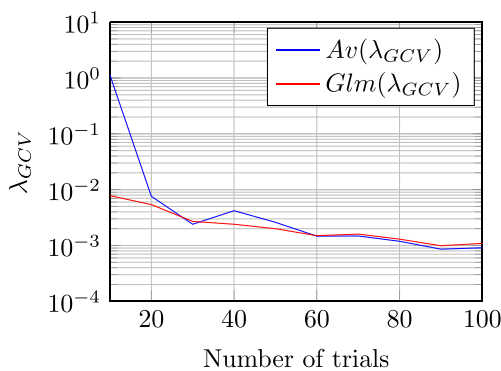


Fig. 11 λ_{GCV} for the “Average” method and GLM, function of the number of trials, SNR = −20 dB

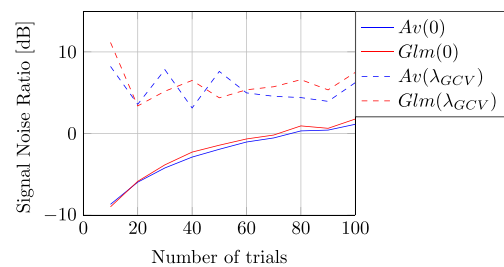


Fig. 12 Estimated SNR function of the number of trials, for the four models

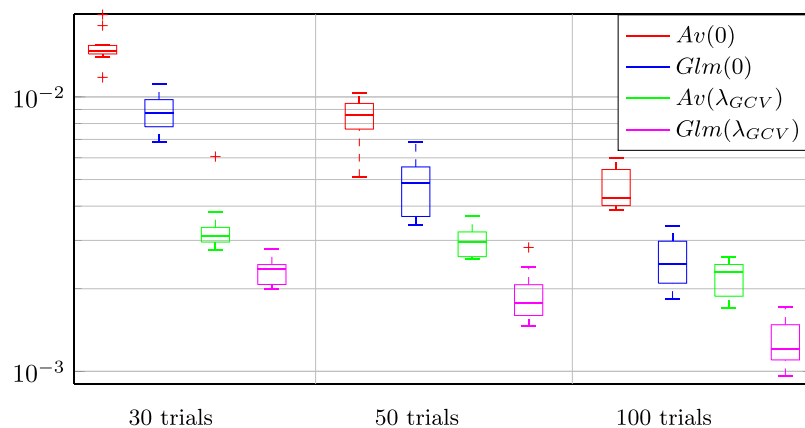
the case of these simulated data, the SAR for $Av(0)$ was 8 dB. The SAR computation for the estimate given by $Glm(0)$ was not relevant as the overlaps were explicitly taken into account. When two overlapping potentials are elicited ($a_1(t)$ and $a_2(t)$), for the $Av(0)$ model, the SAR assesses the distortion due to the overlap of a single potential, and the SIR assesses the distortions due to overlap with the other potential. Applied to the estimation of the potential $a_1(t)$, the SAR for $Av(0)$ was 20 dB and the SIR was 17 dB. For the $Glm(0)$ model, the SAR and SIR were infinite since overlaps between the two potentials were taken into account. Finally, to synthesize results, the MSE illustrated the efficiency of the four models (Fig. 13a), for $SNR = -20$ dB, for one single evoked potential $a(t)$, and for 30, 50, and 100 trials. The MSE decreased with the number of trials and with regularization. Moreover, the estimation by $Glm(0)$ (resp. $Glm(\lambda_{GCV})$) was more efficient than the classical method $Av(0)$ (resp. $Av(\lambda_{GCV})$). This is explained by the smaller bias and variance, as shown in Fig. 10, as well as the reduction of the distortion due to overlaps. The same observations

were made for the situation with two different overlapping potentials, $a_1(t)$ and $a_2(t)$. In this case, the MSE criterion was assessed for the estimation of the potential $a_1(t)$. This MSE (Fig. 13b) had the same behavior function of the number of trials and also function of the four methods, as the MSE evaluated for the estimation of one elicited potential $a(t)$ (Fig. 13a). Even if the SNRs were equivalent, the efficiencies of $Av(0)$ and $Av(\lambda_{GCV})$ models were impacted by the distortions due to overlapping. This is the reason why the $Glm(0)$ (resp. $Glm(\lambda_{GCV})$) was more efficient than the $Av(0)$ (resp. $Av(\lambda_{GCV})$) model.

Results on EFRP estimation during a free visual exploration

In this section, the proposed methods are assessed on the dataset from the free visual exploration. The aim of this study was to extract the neural response to an ocular fixation, $a(t)$, taking into account the contributions of the

(a) Only one elicited potential $a(t)$ per epoch but with overlapping between the epochs.



(b) Two elicited overlapping potentials $a_1(t)$ and $a_2(t)$ per epoch.

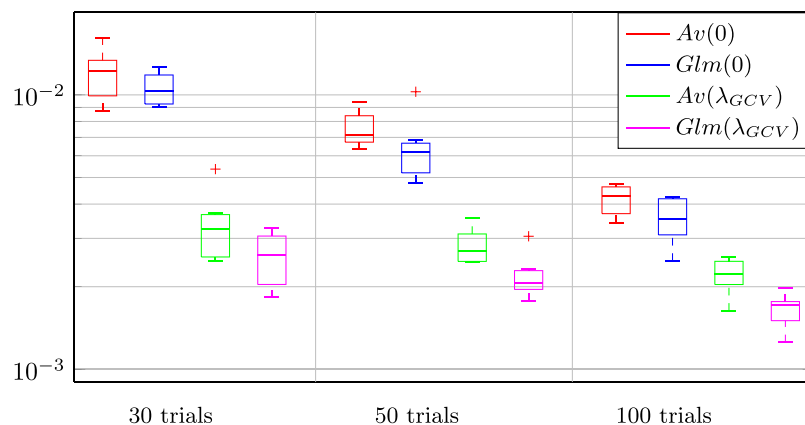


Fig. 13 MSE for $SNR = -20$ dB and for 30, 50, 100 trials in two situations. **a** A single potential. **b** Two potentials. In the second case, MSE assesses the estimation of $a_1(t)$

potential evoked by the immediately precedent and subsequent fixations. The potential was estimated on the temporal window $[-200; 800]$ ms with $N_a = 1000$ samples. The regularization procedures using each subject independently, or using the whole dataset globally, were compared, firstly at the level of the ridge parameter values obtained by the GCV procedure, and secondly at the level of the estimated waveforms by the “Average” method. To avoid repetition, this part was not illustrated here with GLM, because the same conclusions were obtained. In a last subsection, the “Average” method and the GLM are compared with the same regularization procedure (on the whole dataset) to provide results on the quality of the estimations when overlaps occurred.

Comparison between the regularization values on each subject versus on the whole dataset

In this section, the values of the optimal ridge parameter $\lambda_{GCV}^{subject}$ given by the GCV procedure for each subject dataset are compared to the optimal value λ_{GCV} estimated for the whole dataset. The boxplot in Fig. 14 shows the statistical summary of the distribution of the 39 values of $\lambda_{GCV}^{subject}$ compared to the unique value of λ_{GCV} for both the “Average” method and GLM, for the Pz electrode. The values of $\lambda_{GCV}^{subject}$ are higher than the unique value of λ_{GCV} for each of these two models. This was expected because the number of trials was increased (roughly a multiplication by the number of subjects) providing a decreasing value of λ_{GCV} for the whole dataset. Indeed, for the GLM, the value of λ_{GCV} on the whole dataset was $\lambda_{GCV} = 3.10^{-5}$, and $\frac{\lambda_{GCV}^{subject}}{39} = \frac{9.4 \cdot 10^{-4}}{39} = 2.4 \cdot 10^{-5}$. Moreover, the regularization values were lower for the GLM than for the “Average” method. This was in line with the number of

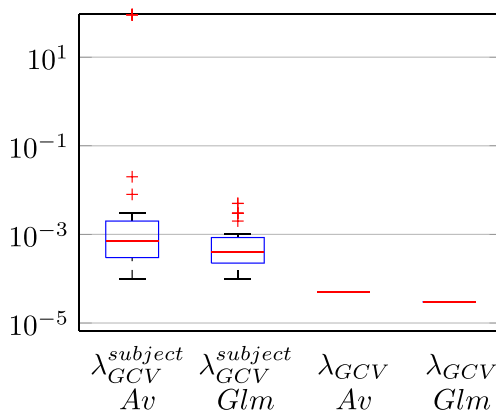


Fig. 14 Comparison between the values of $\lambda_{GCV}^{subject}$ estimated by the GCV procedure for each subject and the value of λ_{GCV} obtained by the GCV on the overall dataset for the “Average” method and Glm, on Pz electrode

synchronization timestamps, which was three times higher for the GLM than for the “Average” method (the GLM took into account three fixations in a given epoch: precedent, current, and subsequent fixations). Lastly, more outlier values were observed for the “Average” method than for the GLM. For six subjects, the search for the optimal value of $V(\lambda)$ (10) did not converge inside the search interval (from 10^{-6} to 100),⁵ and for these subjects, the regularization parameter was set to the maximal value (100). When the SNR was low, searching for the optimal value was more difficult than when the SNR was high, because the valley of the $V(\lambda)$ curve was larger, and less steep with more noise. As a consequence, the regularization was too strong.

Estimation with regularization on each subject vs. on the whole dataset with the “Average” method

Estimation with regularization on each subject

With a regularization per subject, the value of the ridge parameter was adjusted depending on the intrinsic level of noise in the acquired signals and on the number of trials for a given subject. The grand average over all subjects was then estimated by averaging all the regularized estimates. Figure 15 illustrates this procedure for the “Average” method on the Pz electrode. In Fig. 15a, the grand averages given by $Av(0)$ and $Av(\lambda_{GCV}^{subject})$ models, for the complete dataset (39 subjects), are plotted. When the grand average without regularization, obtained directly by the $Av(0)$ model (the unbiased estimator), was compared with the grand average computed using each estimate obtained by the $Av(\lambda_{GCV}^{subject})$ models, we observed an increased bias for the latter (see Fig. 15a for the differences between two different estimates).

To explain this bias, Fig. 15b shows the grand averages given by $Av(\lambda_{GCV}^{subject})$ model for the whole dataset (39 subjects) and for a reduced dataset (33 subjects) without the six subjects’ datasets for which the values of the ridge parameter $\lambda_{GCV}^{subject}$ were too high (see the previous section). A multiplication factor was noticeable between these two estimates for 39 and 33 subjects. This gain factor on the whole temporal window can be estimated by maximum likelihood: $G = 1.181$. This value was equal to $\frac{39}{33}$. In fact, for the complete dataset, the grand average was obtained, as the sum of each estimate by subject divided by 39 (the total number of subjects). However, for six subjects, the predictors in the matrix \mathbf{D} were masked by the high value of $\lambda_{GCV}^{subject}$

⁵according to a logarithmic scale such as for 10^{-6} , from 10^{-5} to 10^{-4} by step 10^{-5} , from 10^{-4} to 10^{-3} by step 10^{-3} , from 10^{-3} to 10^{-2} by step 10^{-2} , from 10^{-2} to 10^{-1} by step 10^{-1} , from 10^{-1} to 1 by step 10^{-1} , from 1 to 100 by step 1

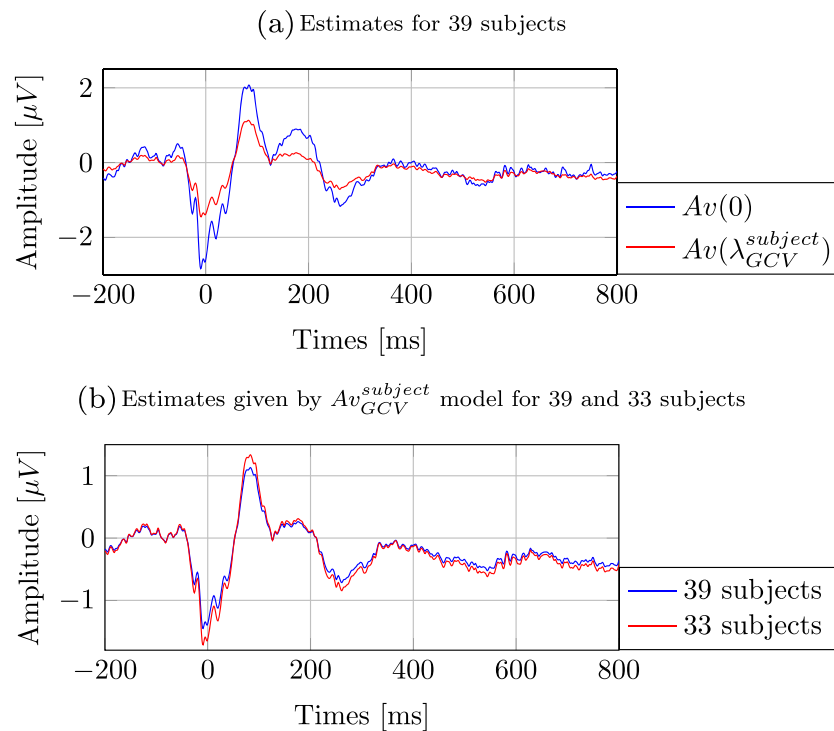


Fig. 15 **a** The estimate on Pz electrode, given by the $Av(0)$ model and by the grand average with the $Av(\lambda_{GCV}^{subject})$ models, for the complete dataset with 39 subjects. **b** The two estimates on Pz electrode, given by the grand average with the $Av(\lambda_{GCV}^{subject})$ models, for the complete dataset with 39 subjects and for the reduced dataset with 33 subjects

($\lambda_{GCV}^{subject} = 100$), and consequently the matrix $(\mathbf{D}^\dagger \mathbf{D} + \lambda \mathbf{N} \mathbf{I})$ in Eq. 9 was close to the diagonal matrix $\lambda \mathbf{N} \mathbf{I}$, and then the resulting estimate was over-regularized near to zero. In addition, to explain completely the bias noticed in Fig. 15a, the impact of the regularization for the 33 other subjects must be considered. Even if the estimates for these subjects were not over-regularized, with an optimal regularization controlled by a suitable ridge parameter, a bias was automatically introduced. This is the reason why the gain factor between the two estimates, in the Fig. 15a, was higher than $\frac{39}{33}$ ($G = 1.58$). To conclude, when the distribution of the ridge parameter's values, $\lambda_{GCV}^{subject}$, has a large variability with high values close to and above 1, the bias of the estimated grand average given by the $Av(\lambda_{GCV}^{subject})$ model increased: the average is computed by dividing with the total number of subjects, but in the sum, some of these subjects' datasets provide over-regularized estimates near to zero. Nevertheless, for the $Av(\lambda_{GCV}^{subject})$ model, the variance of the grand average estimation decreases, the contribution of signals with high noise being reduced by over-regularization.

Estimation with regularization on the whole dataset

With a unique regularization on the whole dataset, the grand average of the potential of interest $a(t)$ was obtained directly by the $Av(\lambda_{GCV})$ method. The value of the ridge

parameter λ_{GCV} was obtained by the GCV procedure applied to the whole dataset. In this way, the same level of regularization has been applied on all subjects. Figure 16 illustrates this regularization procedure for the “Average” method on the Pz electrode, with the complete dataset (39 subjects), compared with the procedure without regularization ($Av(0)$), and the previous result described in “Estimation with regularization on each subject” with the regularization applied to each subject ($Av(\lambda_{GCV}^{subject})$). Due to the increased number of trials, the value of the ridge parameter λ_{GCV} was smaller ($\lambda_{GCV} = 5 \cdot 10^{-5}$) than the average of the values of $\lambda_{GCV}^{subject}$ (cf. Fig. 14), and tended towards zero. The two waveforms were very similar, and the bias of the regularized estimate $\hat{a}(t)$ is lower. However, its variance

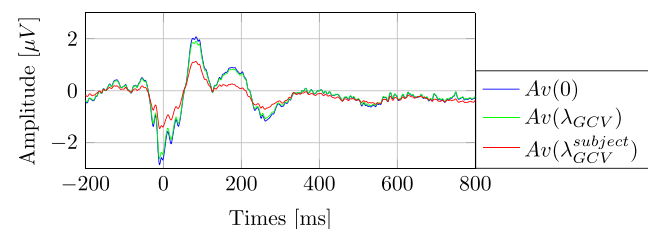


Fig. 16 Estimations of the potential $a(t)$ on Pz electrode, given by $Av(0)$, and $Av(\lambda_{GCV})$ models, and the grand average of $Av(\lambda_{GCV}^{subject})$ for the complete dataset of 39 subjects

was greater because all the trials supplied the same contribution to the estimate $\hat{a}(t)$ of the final potential with the $Av(\lambda_{GCV})$ model, independently of the noise power on the trials for the different subjects.

Comparison between the “Average” method and the GLM with regularization on the whole dataset

In this section, the estimates $\hat{a}(t)$ obtained by the $Av(\lambda_{GCV})$ and $Glm(\lambda_{GCV})$ methods are compared for the whole dataset (39 subjects). In other words, here the lower bias and higher variance provided by regularization on the whole dataset were preferred to the higher bias and lower variance provided by the regularization on each subject. In Fig. 17, the estimate given by the $Av(\lambda_{GCV})$ method is plotted in blue, the estimate given by $Glm(\lambda_{GCV})$ is plotted in red and histograms represent the onsets of precedent and subsequent fixations. In this figure, for both estimates, the lambda wave and the P2 potential were clearly noticeable. The pre-stimulus baseline for the $Av(\lambda_{GCV})$ method was not stabilized due to overlapping of the neural responses to adjacent fixations. Because the baseline is not stabilized, it was difficult to compare the amplitude of different components. Furthermore, the return to zero was faster with $Glm(\lambda_{GCV})$ than with the $Av(\lambda_{GCV})$ method, which was still influenced by the neural response to the subsequent fixations. These results showed the influence of the precedent and subsequent responses on the potentials estimated by the $Av(\lambda_{GCV})$ method and the efficiency of $Glm(\lambda_{GCV})$ to deconvolve the neural responses to successive fixations. Concerning the effect of the regularization, the lambda wave and the P2 potential estimated by the $Av(\lambda_{GCV})$ method were attenuated when compared to the potentials estimated by $Glm(\lambda_{GCV})$. This is explained by a higher ridge parameter with the $Av(\lambda_{GCV})$ method ($\lambda_{GCV} = 5.10^{-5}$) than with $Glm(\lambda_{GCV})$ ($\lambda_{GCV} = 3.10^{-5}$).

Results on P300 Speller data

In this section, the four methods are compared on the dataset from the P300 Speller experiment. For this illustration, the

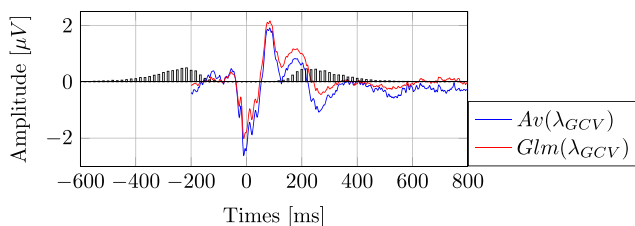


Fig. 17 Estimates on Pz electrode, given by $Av(\lambda_{GCV})$ and $Glm(\lambda_{GCV})$ for the complete dataset of 39 subjects. Histograms represent the onsets of precedent and subsequent fixations

GLM configuration was done to highlight the distinction between two kinds of potential. For instance, this situation is well adapted for classification purposes. In the present case, the aim was to extract the neural response to a target stimuli. For this, two classes were defined (12): with the first class, the potential $a_1(t)$ was estimated on the temporal window $[-500; 800]$ ms, $N_{a1} = 1560$ samples time-locked on the target stimuli. With the second class the potential $a_2(t)$ was estimated with the same duration, $N_{a2} = 1560$ samples, time-locked on each stimulus. This potential $a_2(t)$ must take into account what is common to all the stimuli, both target as well as non-target. The regularization was estimated on the overall dataset providing a single ridge parameter thanks to the concatenation across the subjects as described by Eq. 11. Thus, the regularization was identical for each participant.

Effect of the number of trials on the regularization parameter

We studied the evolution of the optimal value of λ_{GCV} , given by the GCV procedure according to the number of trials. In Fig. 18, the values of λ_{GCV} are plotted against the number of trials for $Av(\lambda_{GCV})$ and for ($Glm(\lambda_{GCV})$) for the Pz electrode. As for the simulated data, and as expected, the ridge parameter λ_{GCV} decreased when the number of trials increased, and as for EFRP data, the value of the ridge parameter was lower for $Glm(\lambda_{GCV})$ than for $Av(\lambda_{GCV})$. The estimation by $Glm(\lambda_{GCV})$ took into account the overlapping from the other stimuli in contrast to the estimation by $Av(\lambda_{GCV})$. In other words, these neural responses were considered as a signal for GLM, and as noise for the “Average” method. In this last case, it was the equivalent to an increase of the noise level providing a higher regularization.

Application of the four models

In this section, the efficiencies of the four models are compared. The potential of interest was the target $a_1(t)$. The

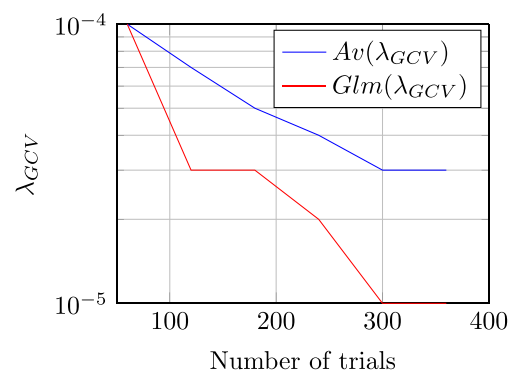


Fig. 18 λ_{GCV} against the number of trials for the $Av(\lambda_{GCV})$ and $Glm(\lambda_{GCV})$ estimations on Pz electrode

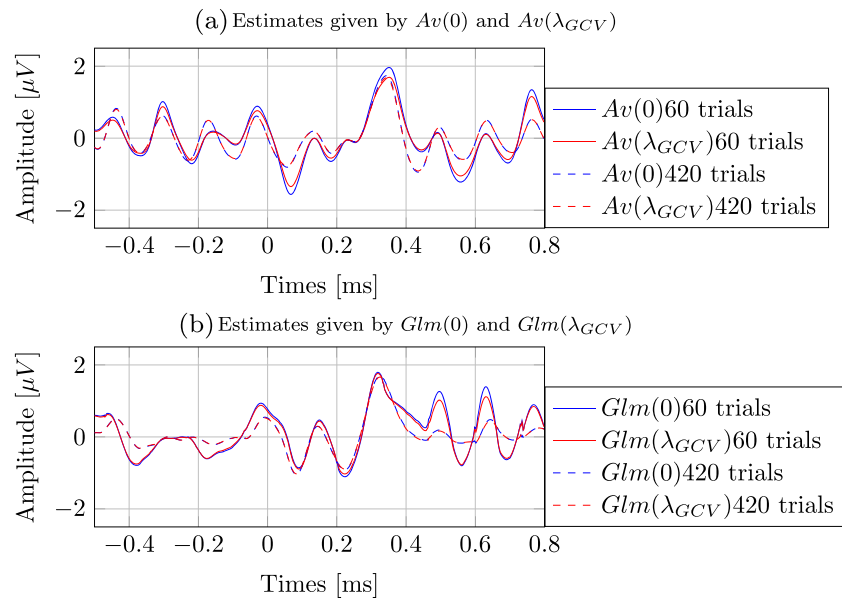


Fig. 19 Estimates on Pz electrode, for 60 and 420 trials. **a** The estimates by $Av(0)$ and $Av(\lambda_{GCV})$ model. **b** The estimates by $Glm(0)$ and $Glm(\lambda_{GCV})$

target estimate $\hat{a}_1(t)$ given by the four models, and for 60 trials (dashed line) or 420 trials (solid lines) are plotted in Fig. 19. Firstly, for the four estimations, oscillations were observed with a periodicity around 130 ms, which corresponds to ISI value between stimuli. These oscillations were due to the steady-state visual evoked potential (Middendorff, McMillan, Calhoun, Jones, et al., 2000) elicited by each stimulus. Concerning the four estimations of the target potential, the P300 wave was observed on the estimated target potential, at a latency of around 300 ms, as expected with this oddball paradigm. For the estimates given by $Av(0)$ and $Av(\lambda_{GCV})$ (Fig. 19a), the regularization had no significant effect for 420 trials: small value of the ridge parameter ($\lambda_{GCV} = 2.10^{-5}$), since there was a large number of trials. However, for 60 trials, the P300 wave was more attenuated by the regularization: the ridge parameter was greater with 60 trials ($\lambda_{GCV} = 1.10^{-4}$) than with 420 trials. The estimates given by $Glm(0)$ and $Glm(\lambda_{GCV})$ are shown in (Fig. 19b). The GLM included the potential $a_2(t)$ ⁶ to estimate the potential of interest $a_1(t)$, this was not the case for the average estimation. Consequently, the remaining oscillations were attenuated compared to these ones providing by averaging (in both $Av(0)$, and $Av(\lambda_{GCV})$ cases): this was clearly noticeable during the pre-stimuli period. This behavior was observed with both a few number of trials (60), and a large number of trials (420).

⁶The discussion concerning the potential $a_2(t)$ is developed in Appendix 3.

Conclusions

The present study develops, applies, and compares four methods (the classical average and the general linear model with and without regularization for both of them) to estimate the evoked potentials in order to improve the SNR and to limit the distortions due to overlapping between adjacent temporally neural responses. To this end, (i) the GLM was proposed to manage the overlapping issue (Kiebel & Holmes, 2003; Dale, 1999; Bardy, Dillon, & Van Dun, 2014) and (ii) the zero-order Tikhonov regularization procedure (Tikhonov, 1963) was used to increase the SNR for a given number of trials. The regularization parameter was estimated by the generalized cross validation procedure (Golub et al., 1979). The interest of this study was to validate and compare these methods on simulated but realistic data, and also on real EEG signals registered during two experiments: a co-registration of eye movements and EEG experiment and a brain–computer interface experiment. To our knowledge, with this article the regularization and the GLM are applied, for the first time, in the context of the EFRP technique. The results demonstrate the interest of the regularization associated to the GLM to estimate evoked potentials in practical contexts (noisy EEG signals, few trials, overlapped potentials). The main assumption of GLM is the linearity of the additive model to take account the different neural responses elicited by a specific event. The non-linearity of the neural responses is a sensitive question, for the relationship between the ERP components through their amplitude/latency and the variables tested by the

experimental conditions (Tremblay & Newman, 2015), but also for the interferences between adjacent responses where the neural responses are no longer time-invariant (Näätänen et al., 1988). Integrate suitable non-linearity remains a great challenge for future work, however this linear model can still be very useful in a wide variety of applications.

Acknowledgments The authors want to thank H el ene Devillez for providing data from a joint EEG and eye movement experiment during a free exploration on natural scenes. This work was supported by a grant from the CNRS funding the PhD of Emmanuelle Kristensen, and partially funded by a grant from the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01).

Appendix 1: Theoretical developments to estimate the signal-to-noise ratio (SNR), signal-to-artifact ratio (SAR), signal-to-interference ratio (SIR)

The signal noise ratio, signal-to-artifact ratio, and signal-to-interference ratio are defined by Eqs. 15, 16 and 17, respectively.

The remaining overlaps, $\underline{Q}_{vl}(a_1)$ and $\underline{Q}_{vl}(a_2)$, and the remaining noise, \underline{n}' , due to \underline{a}_1 and \underline{a}_2 , for ‘‘Average’’ method and GLM are estimated as followed:

1. ‘‘Average’’ method

In the case of the ‘‘Average’’ method, $\hat{\underline{a}}_1$ is expressed as:

$$\begin{aligned} \hat{\underline{a}}_1 &= \frac{1}{E} \mathbf{D}_{1erp}^\dagger \underline{x} \\ &= \underline{a}_1 + \frac{1}{E} \mathbf{D}_{1erp}^\dagger \mathbf{D}_{1ov} \underline{a}_1 + \frac{1}{E} \mathbf{D}_{1erp}^\dagger \mathbf{D}_2 \underline{a}_2 + \frac{1}{E} \mathbf{D}_{1erp}^\dagger \underline{n} \end{aligned} \tag{20}$$

Consequently

$$\begin{aligned} \underline{Q}_{vl}(a1) &= \frac{1}{E} \mathbf{D}_{1erp}^\dagger \mathbf{D}_{1ov} \underline{a}_1 \\ \underline{Q}_{vl}(a2) &= \frac{1}{E} \mathbf{D}_{1erp}^\dagger \mathbf{D}_2 \underline{a}_2 \\ \underline{n}' &= \frac{1}{E} \mathbf{D}_{1erp}^\dagger \underline{n} \end{aligned}$$

2. General linear model

In the case of the GLM method, $\hat{\underline{a}}_1$ is expressed as

$$\hat{\underline{a}}_1 = \mathbf{B}_1 \mathbf{D}^\dagger \underline{x} \tag{21}$$

where $\mathbf{B}_1 \in \mathbb{R}^{N_{a1} \times N_a}$ are the first rows of the matrix $(\mathbf{D}^\dagger \mathbf{D})^{-1}$, with $\mathbf{D} = [\mathbf{D}_1 \ \mathbf{D}_2]$. Consequently, $\hat{\underline{a}}_1 = \underline{a}_1 + \underline{n}'$ with $\underline{n}' = \mathbf{B}_1 \mathbf{D}^\dagger \underline{n}$. This means that $\underline{Q}_{vl}(a1)$ and $\underline{Q}_{vl}(a2)$ vanish, leading thus to infinite SIR and SAR.

Appendix 2: Theoretical mathematical expectation and variance

The Eq. 18 can be developed:

$$\begin{aligned} \hat{\underline{a}}_{reg}(\lambda) &= (\mathbf{D}^\dagger \mathbf{D} + \lambda \mathbf{N} \mathbf{I})^{-1} \mathbf{D}^\dagger (\mathbf{D} \underline{a} + \underline{n}) \\ \hat{\underline{a}}_{reg}(\lambda) &= \Delta \underline{a} + \Omega \underline{n} \end{aligned} \tag{22}$$

with $\Delta = (\mathbf{D}^\dagger \mathbf{D} + \lambda \mathbf{N} \mathbf{I})^{-1} \mathbf{D}^\dagger \mathbf{D}$ and $\Omega = (\mathbf{D}^\dagger \mathbf{D} + \lambda \mathbf{N} \mathbf{I})^{-1} \mathbf{D}^\dagger$.

In the simulated situations, \underline{a} and \underline{n} are known. So, the mathematical expectation $E[\hat{\underline{a}}]$, plotted in Figs. 6 and 10, is calculated as:

$$\begin{aligned} E[\hat{\underline{a}}] &= \Delta \underline{a} + \Omega E[\underline{n}] \\ E[\hat{\underline{a}}] &= \Delta \underline{a} \end{aligned} \tag{23}$$

\underline{n} is not zero-centered so $E[\underline{n}] = 0$.

In the same way, the variance $\text{Var}[\hat{\underline{a}}]$ is equal to:

$$\begin{aligned} \text{Var}[\hat{\underline{a}}] &= E[(\hat{\underline{a}} - E[\hat{\underline{a}}])(\hat{\underline{a}} - E[\hat{\underline{a}}])^\dagger] \\ \text{Var}[\hat{\underline{a}}] &= \Omega E[\underline{n} \underline{n}^\dagger] \Omega^\dagger \end{aligned} \tag{24}$$

Appendix 3: Results concerning the P300Speller dataset: Estimation of potential $a_2(t)$

The potential $a_2(t)$ can be explained by the P300 Speller paradigm. In this latter experiment, target and non-target stimuli were presented at a regular rate (here equal to 133 ms). Consequently, these regular intensifications elicited steady-state visual evoked potentials (SSVEP) over parietal and occipital areas. These SSVEP corresponded in the GLM to the term $\mathbf{D}^{(2)} a_2(t)$. From a numerical point of view, the estimation of the potential $a_2(t)$ is inaccurate because of the regularity of the ISI. Indeed, the matrix $\mathbf{D}^{(2)\dagger} \mathbf{D}^{(2)}$ is closed to be singular. The contribution $\mathbf{D}^{(2)} a_2(t)$ in the GLM (12) is well estimated. See Fig. 20 for this estimate on one epoch. This figure shows the SSVEP signal for an epoch. As expected, oscillations were observed with a periodicity of around 133 ms.

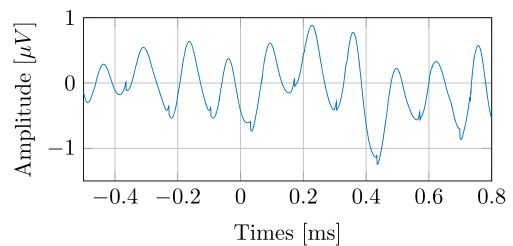


Fig. 20 SSVEP estimate on one epoch

References

- Bardy, F., Dillon, H., & Van Dun, B. (2014). Least-squares deconvolution of evoked potentials and sequence optimization for multiple stimuli under low-jitter conditions. *Clinical Neurophysiology*, 125(4), 727–737.
- Bardy, F., Van Dun, B., Dillon, H., & McMahon, C.M. (2014). Deconvolution of overlapping cortical auditory evoked potentials recorded using short stimulus onset-asynchrony ranges. *Clinical Neurophysiology*, 125(4), 814–826.
- Bekker, E.M., Kenemans, J.L., Hoeksma, M.R., Talsma, D., & Verbaten, M.N. (2005). The pure electrophysiology of stopping. *International Journal of Psychophysiology*, 55(2), 191–198.
- Bell, A.J., & Sejnowski, T.J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1129–1159.
- Brannon, E.M., Roussel, L.W., Meck, W.H., & Woldorff, M. (2004). Timing in the baby brain. *Cognitive Brain Research*, 21(2), 227–233.
- Brannon, E.M., Libertus, M.E., Meck, W.H., & Woldorff, M.G. (2008). Electrophysiological measures of time processing in infant and adult brains: Weber's law holds. *Journal of Cognitive Neuroscience*, 20(2), 193–203.
- Burns, M.D., Bigdely-Shamlo, N., Smith, N.J., Kreutz-Delgado, K., & Makeig, S. (2013). Comparison of averaging and regression techniques for estimating event related potentials. In: Engineering in medicine and biology society (EMBC), 2013 35th annual international conference of the IEEE (pp. 1680–1683).
- Congedo, M., Korczowski, L., Delorme, A., & other (2016). Spatio-temporal common pattern: a companion method for ERP analysis in the time domain. *Journal of Neuroscience Methods*, 267, 74–88.
- Crosse, M.J., Butler, J.S., & Lalor, E.C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *The Journal of Neuroscience*, 35(42), 14195–14204.
- Dale, A.M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8(2–3), 109–114.
- Dandekar, S., Privitera, C., Carney, T., & Klein, S.A. (2012). Neural saccadic response estimation during natural viewing. *Journal of Neurophysiology*, 107(6), 1776–1790.
- Delgado, R.E., & Ozdamar, O. (2004). Deconvolution of evoked responses obtained at high stimulus rates. *The Journal of the Acoustical Society of America*, 115(3), 1242–1251.
- Devillez, H., Guyader, N., & Guérin-Dugué, A. (2015). An eye fixation related potentials analysis of the P300 potential for fixations onto a target object when exploring natural scenes. *Journal of Vision*, 15(13), 20–20.
- Devillez, H., Kristensen, E., Guyader, N., Rivet, B., & Guérin-Dugué, A. (2015). The P300 potential for fixations onto target object when exploring natural scenes during a visual task after denoising overlapped EFRP. In *2015 7th international IEEE/EMBS conference on neural engineering (NER)* (pp. 1024–1027).
- Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A.M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, 140(4), 552.
- Engl, H.W., & Grever, W. (1994). Using the L-curve for determining optimal regularization parameters. *Numerische Mathematik*, 69(1), 25–31.
- Eysholdt, U., & Schreiner, C. (1982). Maximum length sequences: a fast method for measuring brain-stem-evoked responses. *Audiology*, 21(3), 242–250.
- Farwell, L.A., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6), 510–523.
- Fiebelkorn, I.C., Foxe, J.J., McCourt, M.E., Dumas, K.N., & Molholm, S. (2013). Atypical category processing and hemispheric asymmetries in high-functioning children with autism: Revealed through high-density EEG mapping. *Cortex*, 49(5), 1259–1267.
- Frey, A., Ionescu, G., Lemaire, B., López-orocho, F., Baccino, T., & Guérin-Dugué, A. (2013). Decision-making in information seeking on texts: an eye-fixation-related potentials investigation. *Frontiers in Systems Neuroscience*, 7.
- Golub, G.H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2), 215–223.
- Hansen, J.C. (1983). Separation of overlapping waveforms having known temporal distributions. *Journal of Neuroscience Methods*, 9(2), 127–139.
- Hansen, P.C. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4), 561–580.
- Hopfinger, J.B., & Mangun, G.R. (1998). Reflexive attention modulates processing of visual stimuli in human extrastriate cortex. *Psychological Science*, 9(6), 441–447.
- Jewett, D.L., Caplovitz, G., Baird, B., Trumpis, M., Olson, M.P., & Larson-Prior, L.J. (2004). The use of QSD (q-sequence deconvolution) to recover superposed, transient evoked-responses. *Clinical Neurophysiology*, 115(12), 2754–2775.
- Kaunitz, L.N., Kamienskowski, J.E., Varatharajah, A., Sigman, M., Quiroga, R.Q., & Ison, M.J. (2014). Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *NeuroImage*, 89, 297–305.
- Kenemans, J., Bekker, E., Lijffijt, M., Overtom, C., Jonkman, L., & Verbaten, M. (2005). Attention deficit and impulsivity: selecting, shifting, and stopping. *International Journal of Psychophysiology*, 58(1), 59–70.
- Key, A.P.F., Dove, G.O., & Maguire, M.J. (2005). Linking brainwaves to the brain: an ERP primer. *Developmental Neuropsychology*, 27(2), 183–215.
- Kiebel, S., & Holmes, P. (2003). *The general linear model*. Academic.
- Kristensen, E., Guerin-Dugué, A., & Rivet, B. (2015). Comparison between adjar and xdown algorithms to estimate eye-fixation related potentials distorted by overlapping. In *2015 7th international IEEE/EMBS conference on neural engineering (NER)* (pp. 976–979).
- Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., McDarby, G., & Foxe, J.J. (2006). The VESPA: a method for the rapid estimation of a visual evoked potential. *NeuroImage*, 32(4), 1549–1561.
- Lalor, E.C., Power, A.J., Reilly, R.B., & Foxe, J.J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Journal of Neurophysiology*, 102(1), 349–359.
- Luck, S.J. (2014). *An introduction to the event-related potential technique*. MIT Press.
- Middendorf, M., McMillan, G., Calhoun, G., Jones, K.S., & other (2000). Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Transactions on Rehabilitation Engineering*, 8(2), 211–214.
- Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, 69(6), 523–531.
- Niedermeyer, E., & da Silva, F.L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams Wilkins.

- Nikolaev, A.R., Meghanathan, R.N., & van Leeuwen, C. (2016). Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and Cognition*, *107*, 55–83.
- Parra, L.C., Spence, C.D., Gerson, A.D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, *28*(2), 326–341.
- Rivet, B., & Souloumiac, A. (2007). Subspace estimation approach to p300 detection and application to brain-computer interface. In *EMBS 2007. 29th annual international conference of the IEEE engineering in medicine and biology society, 2007* (pp. 5071–5074).
- Rivet, B., & Souloumiac, A. (2013). Optimal linear spatial filters for event-related potentials based on a spatio-temporal model: Asymptotical performance analysis. *Signal Processing*, *93*(2), 387–398.
- Rivet, B., Souloumiac, A., Gibert, G., & Attina, V. (2008). P300 speller brain-computer interface: Enhancement of p300 evoked potential by spatial filters. In *2008 16th European signal processing conference* (pp. 1–5).
- Rivet, B., Souloumiac, A., Attina, V., & Gibert, G. (2009). x{{DAWN}} algorithm to enhance evoked potentials: application to brain-computer interface. *IEEE Transactions on Biomedical Engineering*, *56*(8), 2035–2043.
- Schmajuk, M., Liotti, M., Busse, L., & Woldorff, M.G. (2006). Electrophysiological activity underlying inhibitory control processes in normal adults. *Neuropsychologia*, *44*(3), 384–395.
- Subramaniyam, N.P., Väisänen, O.R., Wendel, K.E., & Malmivuo, J.A. (2010). Cortical potential imaging using L-curve and GCV method to choose the regularisation parameter. *Nonlinear Biomedical Physics*, *4*(Suppl 1), S4.
- Talsma, D., Doty, T.J., & Woldorff, M.G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*(3), 679–690.
- Tanaka, H., Komatsuzaki, A., & Hentona, H. (1996). Usefulness of auditory brainstem responses at high stimulus rates in the diagnosis of acoustic neuroma. *ORL*, *58*(4), 224–228.
- Tikhonov, A. (1963). Solution of incorrectly formulated problems and the regularization method. In *Soviet math. dokl.* (vol. 5, pp. 1035–1038).
- Tremblay, A., & Newman, A.J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*(1), 124–139.
- Vincent, E., Gribonval, R., & Févotte, C. (2006). Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(4), 1462–1469.
- Wang, T., Özdamar, Ö., Bohórquez, J., Shen, Q., & Cheour, M. (2006). Wiener filter deconvolution of overlapping evoked potentials. *Journal of Neuroscience Methods*, *158*(2), 260–270.
- Woldorff, M.G. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs: analysis and correction. *Psychophysiology*, *30*(1), 98–119.
- Yagi, A. (1981). Visual signal detection and lambda responses. *Electroencephalography and Clinical Neurophysiology*, *52*(6), 604–610.