

Full Length Articles

The LDA beamformer: Optimal estimation of ERP source time series using linear discriminant analysis



Matthias S. Treder ^{a,b,*}, Anne K. Porbadnigk ^c, Forooz Shahbazi Avarvand ^c, Klaus-Robert Müller ^{c,d,**}, Benjamin Blankertz ^{a,***}

^a Neurotechnology Group, Technische Universität Berlin, Germany

^b Behavioural & Clinical Neuroscience Institute, Department of Psychiatry, University of Cambridge, UK

^c Machine Learning Laboratory, Technische Universität Berlin, Germany

^d Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

ARTICLE INFO

Article history:

Received 12 May 2015

Accepted 9 January 2016

Available online 20 January 2016

ABSTRACT

We introduce a novel beamforming approach for estimating event-related potential (ERP) source time series based on regularized linear discriminant analysis (LDA). The optimization problems in LDA and linearly-constrained minimum-variance (LCMV) beamformers are formally equivalent. The approaches differ in that, in LCMV beamformers, the spatial patterns are derived from a source model, whereas in an LDA beamformer the spatial patterns are derived directly from the data (i.e., the ERP peak). Using a formal proof and MEG simulations, we show that the LDA beamformer is robust to correlated sources and offers a higher signal-to-noise ratio than the LCMV beamformer and PCA. As an application, we use EEG data from an oddball experiment to show how the LDA beamformer can be harnessed to detect single-trial ERP latencies and estimate connectivity between ERP sources. Concluding, the LDA beamformer optimally reconstructs ERP sources by maximizing the ERP signal-to-noise ratio. Hence, it is a highly suited tool for analyzing ERP source time series, particularly in EEG/MEG studies wherein a source model is not available.

© 2016 Elsevier Inc. All rights reserved.

Introduction

The overarching goal of electroencephalography (EEG) and magnetoencephalography (MEG) research is to infer the dynamics of brain activity on the basis of measurements from a sensor array. The inference problem is complicated by field spread that causes activity in a brain area to spread into many sensors (van Veen and Buckley, 1988; van Veen et al., 1997). This is captured in the linear model of EEG/MEG that is stated as follows (van Veen and Buckley, 1988; van Veen et al., 1997; Hauk and Stenroos, 2013; Hauk et al., 2011; Galka et al., 2004; Lopes da Silva, 2013):

$$\mathbf{x}(t) = \mathbf{P}\mathbf{s}(t) + \boldsymbol{\epsilon} = (\mathbf{p}_1 \ \mathbf{p}_2 \ \dots \ \mathbf{p}_M) \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_M(t) \end{pmatrix} + \boldsymbol{\epsilon} \quad (1)$$

Here, $\mathbf{x}(t)$ is a $(N \times 1)$ column vector representing the EEG/MEG amplitude in each of the N sensors at time t . The source activity in M brain regions is captured in the $(M \times 1)$ vector \mathbf{s} . Source activities and sensor measurements are related via a $(N \times M)$ matrix of spatial patterns \mathbf{P} (a.k.a. gain matrix or lead field matrix), whose columns \mathbf{p}_i specify how activity in source s_i propagates into the sensors. The error term $\boldsymbol{\epsilon}$, assumed to be normally distributed, models sensor noise and artifacts.

Due to field spread, univariate (i.e., single-sensor) analysis of the brain activity is severely limited. However, using a multivariate approach, one can exploit the fact that different sources produce characteristic spatial patterns in sensor space. By applying a spatial filter \mathbf{w} , estimates \mathbf{s} of cortical sources can be formed as a linear combination of sensor data (Parra et al., 2005; Tomioka and Müller, 2010; Müller et al., 2003; Blankertz et al., 2008; Kawahabe et al., 2014; Blankertz et al., 2011; Gross and Ioannides, 1999; Lemm et al., 2011):

$$\mathbf{s}(t) = \mathbf{w}^T \mathbf{x}(t). \quad (2)$$

A spatial pattern specifies the *forward model* of a source, that is, how activity at the source level propagates into the sensors. In contrast, a spatial filter specifies the *inverse model*, that is, how the activity of a source can be recovered as a linear combination of the sensor data (Haufe et al., 2014a; Blankertz et al., 2011; Parra et al., 2005).

* Correspondence to: M. S. Treder, Neurotechnology Group, Technische Universität Berlin, Germany.

** Correspondence to: K.-R. Müller, Machine Learning Laboratory, Technische Universität Berlin, Germany.

*** Corresponding author.

E-mail addresses: matthias.treder@gmail.com (M.S. Treder), klaus-robert.mueller@tu-berlin.de (K.-R. Müller), benjamin.blankertz@tu-berlin.de (B. Blankertz).

A popular technique to obtain the spatial filter is linearly constrained minimum-variance (LCMV) beamforming (van Veen and Buckley, 1988; van Veen et al., 1997; Sekihara et al., 2004; Gross et al., 2013; van Vliet et al., 2016). The spatial filter suppresses activity that does not originate from the desired location. To this end, LCMV beamformers minimize the projected power of the covariance matrix. In the presence of correlated sources, however, beamformers suffer from signal cancellation so that the desired signal cannot be recovered. Dual beamformers and multi-core beamformers have been developed to deal with the latter (Brookes et al., 2007; Diwakar et al., 2011), and the nulling beamformer has been proposed to deal with cross-talk (Hui et al., 2010). However, the approaches require either a priori information about the location of the correlated sources or an exhaustive search.

The purpose of this paper is to introduce a novel beamformer based on regularized linear discriminant analysis (LDA) that has two unique characteristics. First, unlike conventional beamformers, its performance improves rather than deteriorates in the presence of correlated sources. The correlated sources are estimated directly from the data, and hence a priori information or a search is not required. Second, it does not require any structural information about the brain (i.e., a source model) and hence it can be applied to a wider range of EEG/MEG studies than other beamformers. Note that since the LDA beamformer is independent of a source model, it cannot be used for localization. However, we show later that it offers an optimal signal-to-noise ratio for single-trial ERPs.

The paper is organized as follows. We introduce LDA and then expand on its use in source reconstruction. To dissociate our approach from the common use of LDA in classification problems, we use the term *LDA beamformer* (as opposed to *LDA classifier*). Subsequently, commonalities between LDA and LCMV beamformers are elucidated. We formally prove that the LDA beamformer does not suffer from signal cancellation in the presence of correlated sources. On the contrary, its performance even improves. In MEG simulations, the LDA beamformer is shown to outperform the LCMV beamformer and PCA. Finally, we illustrate how the LDA beamformer can be used on an EEG dataset (Porbadnigk et al., 2013) to recover the single-trial latencies of ERP components in an oddball task and to estimate connectivity.

Material and methods

Linear discriminant analysis (LDA)

In brain research, LDA has been applied in diverse fields such as brain-computer interfaces (e.g., (Höhne et al., 2012; Schreuder et al., 2010; Treder and Blankertz, 2010; Treder et al., 2014)), monitoring of mental states during driving (Haufe et al., 2011; Sonnleitner et al., 2014; Haufe et al., 2014b; Kohlmorgen et al., 2007) and perception (Porbadnigk et al., 2013; Scholler et al., 2012; Porbadnigk et al., 2011; Müller et al., 2008). It has been used in conjunction with beamformers (Grosse-Wentrup et al., 2009) and different neuroimaging modalities (Fazli et al., 2012; Kriegeskorte et al., 2006; Mur et al., 2009; Fazli et al., 2015; Dähne et al., 2015).

LDA can be shown to optimally separate classes if the following conditions are met:

- *The features are Gaussian distributed.* By the central limit theorem, the sum of independent signals with finite variance converges to a Gaussian distribution. This holds approximately for EEG/MEG data since the activity at each sensor is the sum of multiple generators.
- *All classes have the same covariance matrix.* Raw EEG/MEG is dominated by ongoing oscillations such as the alpha rhythm. However, the statistics of background oscillations are known to be non-stationary. We discuss the consequences of this further below.
- *The true mean and the true covariance are known.* This is obviously not true. Mean and covariance matrix have to be estimated from the data. The quality of the estimates deteriorates in face of few data points or

high levels of noise. We show that, for covariance estimation, regularization using shrinkage reduces the problem. If the optimality criteria are violated, LDA can still be a useful tool, but it is not anymore theoretically guaranteed to give the best class separation.

The LDA beamformer

Let the $(N \times 1)$ column vectors \mathbf{p}_1 and \mathbf{p}_2 be the spatial patterns of a particular component in two different experimental conditions, e.g. target stimuli versus non-target stimuli in a detection paradigm. We denote the difference pattern as $\mathbf{p} := \mathbf{p}_1 - \mathbf{p}_2$ and the $(N \times N)$ covariance matrix as \mathbf{C} . As we prove in Appendix A.2, we can then restate the LDA optimization problem as

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \mathbf{w}^T \mathbf{C} \mathbf{w} \\ & \text{subject to} && \mathbf{w}^T \mathbf{p} = 1 \end{aligned} \quad (3)$$

The term to be minimized represents the power of the source time series. The constraint $\mathbf{w}^T \mathbf{p} = 1$ assures that the signal of interest is preserved. Consequently, the solution to the optimization problem maximizes the signal-to-noise ratio (SNR) of the desired signal. The optimal spatial filter is given by

$$\mathbf{w} = \mathbf{C}^{-1} \mathbf{p} \left(\mathbf{p}^T \mathbf{C}^{-1} \mathbf{p} \right)^{-1} \quad (4)$$

As sketched in Fig. 1a the estimated source activity represented by a $(T \times 1)$ vector \mathbf{s} can then be obtained by applying Eq. (2). Since the spatial pattern is estimated from the data, the unit of the spatial filter \mathbf{w} is the reciprocal of the unit of the sensor activity (e.g., $1/\mu\text{V}$ in case of EEG). Consequently, the projected data is dimensionless and will be referred to as *LDA amplitude*. However, it is proportional to the activity of the underlying neural generators. In other words, while absolute LDA amplitude does not have a physical interpretation, relative comparisons of amplitudes across experimental conditions are informative.

Estimation of LDA parameters

The LDA beamformer depends on the spatial pattern \mathbf{p} and the covariance matrix \mathbf{C} . Both are estimated from the EEG/MEG data.¹

Spatial pattern

The spatial pattern can be obtained from the data as the mean amplitude in a time window centered on the designated ERP peak. The ERP peak mainly reflects the activity of the sources involved in its generation. In case of two experimental conditions, the spatial pattern can be derived from the difference ERP. To increase the quality of the estimate, it is recommended to perform an outlier rejection prior to estimating the LDA parameters.

Covariance matrix

If the data has been normalized to zero mean, the covariance matrix can be estimated using $\mathbf{C} = \frac{1}{T-1} \mathbf{X} \mathbf{X}^T$. If the data is assumed to be strongly non-stationary, it may be beneficial to restrict the covariance estimation to a time window around the ERP peak; note, however, that the covariance matrix should still be estimated using the raw trials, not the ERP. This relaxes the strict stationarity assumption to stationarity in a short time window. In many cases (e.g., after maxfiltering or ICA-based artifact rejection), the covariance matrix is singular and needs to be regularized in order to be invertible. This can be achieved with shrinkage (Blankertz et al., 2011; Blankertz et al., 2008; Ledoit

¹ A MATLAB function lda_beamformer.m implementing the LDA beamformer and the nulling LDA beamformer is included as supplemental material.

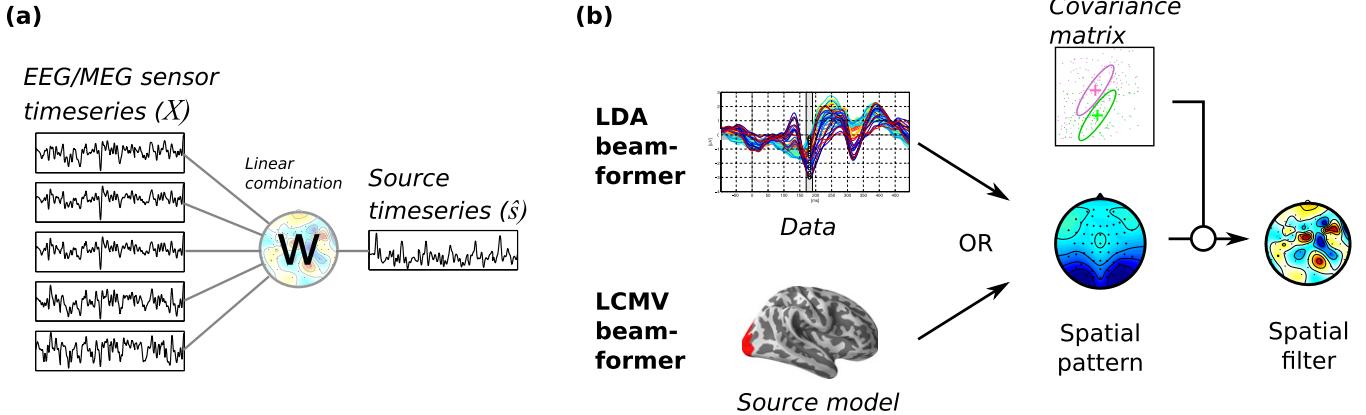


Fig. 1. (a) EEG/MEG sensors are linearly combined, using the coefficients of the spatial filter, to estimate the source time series. (b) Both LDA beamformer and LCMV beamformer require a spatial pattern and the covariance matrix in order to derive a spatial filter. In the LDA beamformer, the spatial pattern is derived from the data. In the LCMV beamformer, it corresponds to a voxel in the source model.

and Wolf, 2004; Schäfer and Strümmer, 2005; Bartz and Müller, 2013; Samek et al., 2014), wherein the regularized covariance matrix is a convex combination of the sample covariance matrix and an identity matrix:

$$\mathbf{C}_{\text{reg}} = (1 - \gamma)\mathbf{C} + \gamma \nu \mathbf{I} \quad (5)$$

where $\nu := \text{trace}(\mathbf{C})/N$ is a weighting factor that brings the shrinkage target on equal footing (i.e., same total variance) with the sample covariance matrix. The shrinkage parameter γ blends between unregularized covariance ($\gamma=0$) and spherical covariance ($\gamma=1$).

LDA beamformer versus LCMV beamformer

The optimization problems posed in LDA and LCMV beamformers are formally equivalent (see Appendix A.2). The methods differ only in how the spatial pattern \mathbf{p} is determined. This is illustrated in Fig. 1b. LCMV beamformers can be used both for source localization (estimating the spatial location of a source) and source reconstruction (recovering its time series). The spatial patterns correspond to voxels obtained from forward modeling. In contrast, the LDA beamformer requires no forward modeling because the spatial pattern is estimated directly from the data, e.g., from an ERP peak. Hence, it can only be used for source reconstruction. Both ways of deriving the spatial pattern are valid approaches, and in a noise-free situation they converge. In practice, neither method is error-free due to intrinsic noise in the data and simplifying model assumptions.

Robustness to correlated sources

LCMV beamforming implicitly assumes that sources are uncorrelated. This is known to be violated, e.g. following auditory stimulation. The consequences can be as severe as full signal cancellation (van Veen et al., 1997). In contrast, the LDA beamformer is robust to correlated sources. To formally prove this, we assume the error term ϵ in the estimation of the spatial pattern to be negligible since this relates to the issue of parameter estimation which is not specific to beamformers. Likewise, for the LCMV beamformer, we assume a localization error of zero. The full robustness proof is given in Appendix A.3, here we only discuss the results.

To understand the reason for signal cancellation, we consider the activity s_k of the k -th source voxel. An index set S denotes sources that are correlated with s_k , called “signal sources”. The set N indexes all other “noise sources”. We assume that signal sources are fully correlated. To this end, we split partially correlated sources in a signal part (fully correlated with s_k) and a noise part (uncorrelated with s_k). For

convenience, we treat the sources s_i and the sensor measurement \mathbf{X} as random variables with zero mean, so that the variance can be denoted as $\sigma_i^2 = \mathbb{E}\{s_i^2\}$. Using Eqs. (1) and (2), the variance of the k -th source estimate is

$$\begin{aligned} \text{var}(s_k) &= \text{var}(\mathbf{w}^\top \mathbf{P} \mathbf{s}) = \underbrace{\mathbb{E}\left\{\sum_{i \in S} \langle \mathbf{w}, \mathbf{p}_i \rangle^2 s_i^2\right\}}_{\text{variance of signal sources}} \\ &\quad + 2\mathbb{E}\left\{\sum_{i, j \in S; i < j} \langle \mathbf{w}, \mathbf{p}_i \rangle s_i s_j \langle \mathbf{w}, \mathbf{p}_j \rangle\right\} + \underbrace{\mathbb{E}\left\{\sum_{i \in N} \langle \mathbf{w}, \mathbf{p}_i \rangle^2 s_i^2\right\}}_{\text{variance of noise sources}} \\ &\quad + 2\mathbb{E}\left\{\sum_{i, j \in N; i < j} \langle \mathbf{w}, \mathbf{p}_i \rangle s_i s_j \langle \mathbf{w}, \mathbf{p}_j \rangle\right\} \end{aligned} \quad (6)$$

Here, $\langle \cdot, \cdot \rangle$ is the standard scalar product. If all sources are uncorrelated, the covariance terms vanish. However, in the case of correlated sources, the covariance terms $\langle \mathbf{w}, \mathbf{p}_i \rangle \langle \mathbf{w}, \mathbf{p}_j \rangle$ can be negative, leading to signal cancellation. Departing from Eq. (6), we can define a signal-to-noise ratio (SNR) as the fraction

$$\text{SNR} := \frac{\mathbb{E}\left\{\sum_{i \in S} \langle \mathbf{w}, \mathbf{p}_i \rangle^2 s_i^2\right\} + 2\mathbb{E}\left\{\sum_{i, j \in S; i < j} \langle \mathbf{w}, \mathbf{p}_i \rangle s_i s_j \langle \mathbf{w}, \mathbf{p}_j \rangle\right\}}{\mathbb{E}\left\{\sum_{i \in N} \langle \mathbf{w}, \mathbf{p}_i \rangle^2 s_i^2\right\} + 2\mathbb{E}\left\{\sum_{i, j \in N; i < j} \langle \mathbf{w}, \mathbf{p}_i \rangle s_i s_j \langle \mathbf{w}, \mathbf{p}_j \rangle\right\}} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{noise}}^2} \quad (7)$$

that contains the (co)variance of the signal sources as numerator and the (co)variance of noise sources as denominator. The LDA beamformer uses a linear combination of the spatial patterns of multiple correlated sources \mathbf{p}_i . Then the linear optimization constraint can be written as $\sum_{i \in S} \mathbf{w}^\top \mathbf{p}_i =: \alpha_1 + \alpha_2 + \dots + \alpha_{|S|} = 1$ with $\mathbf{w}^\top \mathbf{p}_i = \alpha_i \in \mathbb{R}$.

In other words, LDA assures that the joint gain of the correlated sources is 1, that is, it integrates information across all correlated sources. This property leads to the following theorem:

Theorem 1. (Robustness to correlated sources) Let $k \in \mathbb{N}$ be the number of correlated sources, \mathbf{X}_k the corresponding data and \mathbf{w}_k the corresponding LDA beamformer. Then

1. the signal-to-noise ratio (SNR) of the source estimate increases with k and is unbounded, i.e. $\text{SNR} \xrightarrow{k \rightarrow \infty} \infty$.

2. the correlation between the estimated source time series and the true source time series s_0 increases with k to limit 1, i.e.
 $\text{Corr}(s_0, \mathbf{w}_k^\top \mathbf{X}_k) \xrightarrow{k \rightarrow \infty} 1$.

Proof. See Appendix A.3. \square

According to Theorem 1, the LDA beamformer is not only robust to correlated sources: its performance even improves.

MEG simulation

To compare the efficacy of the LDA beamformer to the LCMV beamformer and another data-driven approach, principal component analysis (PCA), we performed a simulation of 273-channel gradiometers MEG wherein we created event-related fields (ERFs) with varying degrees of noise.

Simulation 1: Noise level and regularization

We simulated ERF data consisting of 100 epochs of 1000 ms length and a sampling frequency of 250 Hz. The ERF signal consisted of a sinusoidal peak with a peak frequency of 5 Hz, centered at 460 ms with a random trial-to-trial jitter in the range of ± 20 ms. An individual MRI volume with a volumetric grid consisting of 3000 locations confined to gray matter served as source model. Forward calculations were performed using a realistic single-shell model (Nolte, 2003). The ERF source was a single randomly selected voxel with a random moment. In addition, we randomly placed 300 noise sources featuring identically distributed, independent noise. We varied the noise level (noise-to-signal variance ratio) from 0.5 to 40 in 80 steps. To induce a temporal correlation to the noise (as in real data), each noise source was filtered in the band 1–40 Hz. Simulations were repeated 100 times with randomly selected sources. Fig. 3a shows the ERFs for two different noise levels.

For the LDA beamformer, the spatial pattern was estimated from the ERF peak as average amplitude in the interval 400–520 ms. The covariance matrix was calculated on the full dataset. We additionally varied the regularization parameter γ from 0 (unregularized) to 1 (spherical covariance). For the LCMV beamformer, we estimated the covariance matrix of the ERF in the interval 340–580 ms. For each voxel, the direction that maximized power was chosen using singular value decomposition. To counteract the depth bias, we normalized the variance in this interval by dividing through the projected noise covariance $\mathbf{w}^T \mathbf{C}_N \mathbf{w}$, where the noise covariance was calculated based on the 0–320 ms interval. All covariance matrices were regularized by the same γ . The voxel yielding maximum normalized power was selected as the ERF source.

For PCA, we extracted the first principal component of the unregularized ERF covariance matrix calculated for a time window centered on the ERF peak. To find the optimal window, we varied its width from 8 ms to 112 ms in 8 ms steps. The best overall SNR was obtained for a window width of 80 ms. This window width was subsequently used in both MEG simulations.

To quantify the quality of the reconstructed time series, we calculated SNR and correlation between the true sources and the estimated sources. To obtain SNR, we applied the spatial filter obtained with each method (LDA, LCMV, and PCA) separately to the signal part and to the noise part of the simulated MEG. In accordance with Eq. (7), SNR is then defined on the source level as signal variance divided by noise variance. To obtain correlation, we correlated the (noisy) source estimates with the true (noise-free) source time series.

Simulation 2: Correlated sources

To characterize the performance of the LDA beamformer in the presence of correlated sources, we performed another simulation. We used the same parameters as in Simulation 1, except that the noise level was fixed to 5.0 and the regularization parameter was fixed to 0.001. These

values yielded the lowest dipole localization error for the LCMV beamformer in Simulation 1. We varied the number of fully correlated sources from 1 to 10. Each correlated source corresponded to a randomly chosen voxel with a random moment. The ERF time series consisted of 100 epochs containing a sinusoidal peak, created in the same way as in Simulation 1. Since the sources were 100% correlated, they all had the same ERF time series. Simulations were repeated 100 times.

We determined the source estimates, SNR, and correlation as in Simulation 1. For LDA, we estimated the spatial pattern directly from the ERF peak. For LCMV, we used the voxel yielding maximum normalized power. For PCA, we used the first principal component of the covariance matrix. Note that all signal sources were expressed in a single principal component because they were fully correlated. As before, we calculated SNR on the source level as signal variance divided by noise variance, and we calculated correlation by correlating the source estimates with the true source time series which was identical for all correlated sources.

EEG experiment

To illustrate how the LDA beamformer can serve the analysis of single-trial ERF latencies and connectivity, we examined an EEG dataset on an auditory oddball experiment (Porbadnik et al., 2013). Previous ERP studies showed that N2 and P3 components have an increased amplitude for targets vs. non-targets in oddball paradigms (Folstein and van Petten, 2008). The N2 is a negative deflection in the EEG/MEG occurring about 200 ms after stimulus onset associated with stimulus novelty and response preparation (Folstein and van Petten, 2008; Ritter et al., 1979). The classical parietal P3 component (P3b), occurs 300–600 ms after a target stimulus (Bledowski et al., 2004; Katayama and Polich, 2009). It has been linked to context updating, context closure, and event-categorization, and it is mainly subserved by parietal and inferior temporal areas (Bledowski et al., 2004).

Task and setup

Eleven participants (mean age 25, 7 females) were asked to press a button when they noticed an aberrant stimulus in series of spoken syllables. Phoneme /i/ was a target stimulus (6% of trials), phoneme /a/ was a non-target (70%). The other conditions are not considered here (cf. (Porbadnik et al., 2013)).

Stimuli were presented with 1000 ms SOA in blocks of 300 (stimulus duration: 160 ms), with 8 to 12 blocks per participant. EEG was recorded using a Brain Products (Munich, Germany) 64-electrodes system and a BrainAmp EEG amplifier. Electrodes were placed according to the international 10–10 system with nose reference and a forehead ground electrode. Impedances were kept below 10 k Ω . Sampling rate was 1000 Hz with a hardware bandpass filter of 0.016–250 Hz. For offline analysis, the EEG was downsampled to 500 Hz and bandpass filtered at 1–9 Hz (4th order Butterworth). A narrow band was chosen in order to attenuate alpha activity which is often pronounced in ongoing EEG. Epochs were defined from –200 ms to 1000 ms with respect to stimulus onset. Artifacts were rejected using a min-max and a variance criterion.

Recovering N2 and P3 sources

For each subject, we devised LDA beamformers for N2 and P3 separately. To this end, we determined the difference ERP (target ERP minus non-target ERP) and identified N2 and P3 peak latencies by visual inspection. The spatial patterns were estimated as the amplitudes at these peak latencies (see Fig. 5a). The covariance matrix was calculated using the whole unepoched dataset. The resulting spatial filters were applied to the data, obtaining N2 and P3 time series.

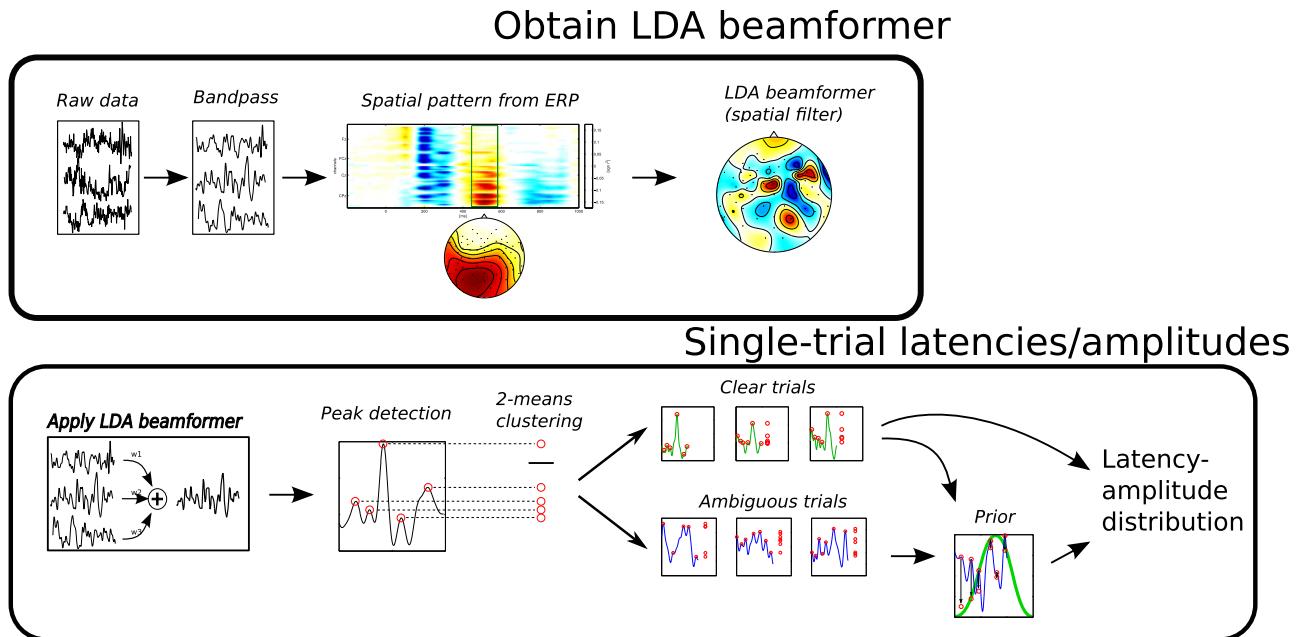


Fig. 2. Overview of the latency estimation method. First, the signal is band pass filtered, the spatial pattern is estimated from the data, and the spatial filter is devised. Then, the data is projected on the filter and the distribution of single-trial peak latencies and peak amplitudes is derived.

Latency estimation

Due to its low SNR, the estimation of single-trial ERP latencies is intricate (Aricò et al., 2014; Thompson et al., 2013; Weeda et al., 2012; Gaspar et al., 2011; D'Avanzo et al., 2011; Hu et al., 2010; Li et al., 2009; Wu et al., 2013). To illustrate that our approach recovers the N2 and P3 time series, we implemented a peak-picking approach for each participant, as illustrated in Fig. 2.

For each target epoch, all peaks (local maxima) were used as input in a two-step approach. First, for each epoch, a two-means clustering algorithm was applied to the peak amplitudes. Trials wherein the absolute maximum yielded a separate cluster showed a single clear peak and were labeled as “clear” trials. The other trials were labeled as “ambiguous” trials. After rejecting outliers, mean and standard deviation of the latency distribution in the “clear” trials were determined, defining a Gaussian prior that specifies where a peak is most likely to occur. After weighting the peaks in the “ambiguous” trials with the Gaussian prior, the resulting maximum peak was extracted. This yielded a distribution of single-trial peak latencies and peak amplitudes.

There are more advanced techniques for single-trial ERP estimation (Weeda et al., 2012; Gaspar et al., 2011; D'Avanzo et al., 2011; Hu et al., 2010; Li et al., 2009; Wu et al., 2013). We resort to the more simple peak-picking method because it allows us to refrain from additional assumptions on the ERP such as the existence of an ERP template. Furthermore, working on the raw time series illustrates the power of the LDA beamformer to increase SNR.

Connectivity analysis

As another application of the LDA beamformer, we investigated whether the connectivity between the N2 and P3 sources changes as a function of time, frequency, and stimulus type (target vs. non-target). To this end, we calculated time-frequency spectra (1–30 Hz) for N2 and P3 time series using Morlet wavelets for both target and non-target epochs. For the non-targets, we only considered those epochs that were both preceded and followed by a non-target stimuli in order to prevent temporal leakage of activity from previous/following trials. We then calculated two connectivity measures between N2 and P3 by averaging across trials, namely the phase-locking value (Lachaux et al.,

1999) and correlation of the signal envelope (Hipp et al., 2012). For the latter, zero-lag correlation between the sources was removed using the regression method outlined in (Hipp et al., 2012). We used cluster permutation tests in FieldTrip (Oostenveld et al., 2011) to establish whether there is a significant difference in connectivity between target and non-target epochs.

Results

MEG simulation 1

Fig. 3b depicts the dipole localization error for the LCMV beamformer as a function of noise level. The lowest mean localization error is 0.79 cm and it is found for $\gamma = 0.001$ and a noise level of 5.0. Fig. 3c depicts SNR. Irrespective of the regularization parameter, the LDA beamformer outperforms the other two methods. The difference is most prominent for small values of γ . For $\gamma \rightarrow 1$, LDA converges to the PCA solution. This is not surprising, since the information of the sample covariance matrix is discarded and the LDA spatial filter becomes equal to the spatial pattern, like in PCA. Fig. 3d depicts the correlation between the true source and the estimated source. As for the SNR, the LDA beamformer performs significantly better than the LCMV beamformers. PCA performs better than LCMV for low noise levels, and LCMV outperforms PCA for high noise levels.

To find statistical support for this observation, we performed an analysis of variance (ANOVA) with method, noise level, and regularization value as variables. For both SNR and correlation, there were main effects of method (SNR: $F = 620.65, p < .001$; correlation: $F = 3990.3, p < .001$), noise level (SNR: $F = 89.58, p < .001$; correlation: $F = 457.81, p < .001$), and regularization (SNR: $F = 171.32, p < .001$; correlation: $F = 455.48, p < .001$). Tukey-Kramer post hoc tests showed that LDA yields significantly higher scores than the other methods, both in terms SNR and correlation.

MEG simulation 2

Fig. 4a depicts SNR as a function of the number of correlated sources. While the SNR drops close to zero for the LCMV beamformer,

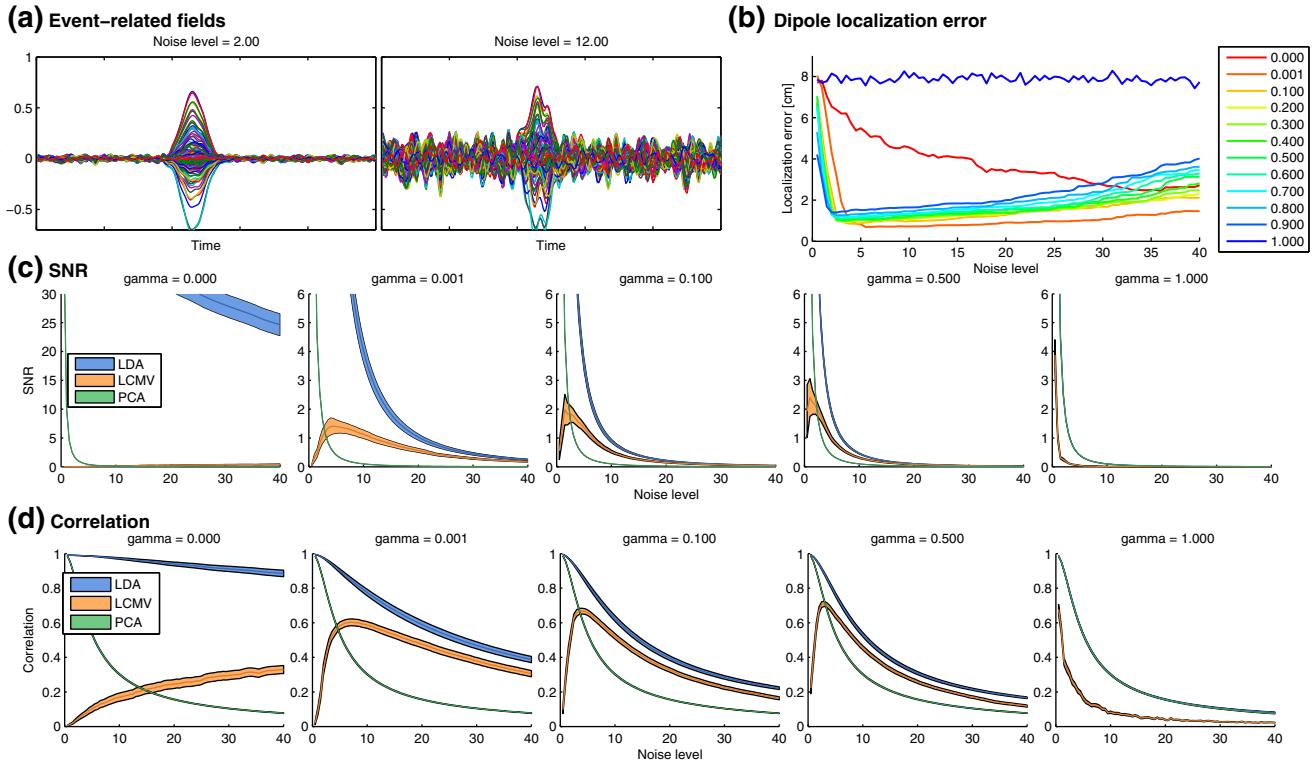


Fig. 3. Results of simulation 1. (a) Event-related fields for low, medium, and high noise levels. Each colored curve represents a different gradiometer. (b) Dipole localization error as a function of noise for the two LCMV beamformers. The shaded area represents 2 SEM. Different plots represent different values of the regularization parameter γ . (c) SNR of the LDA beamformer and the LCMV beamformers. (d) Correlation between the estimated source signal and the true source signal.

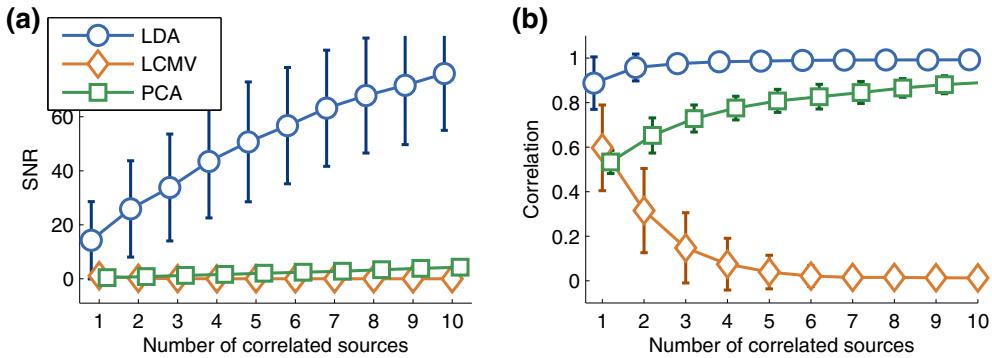


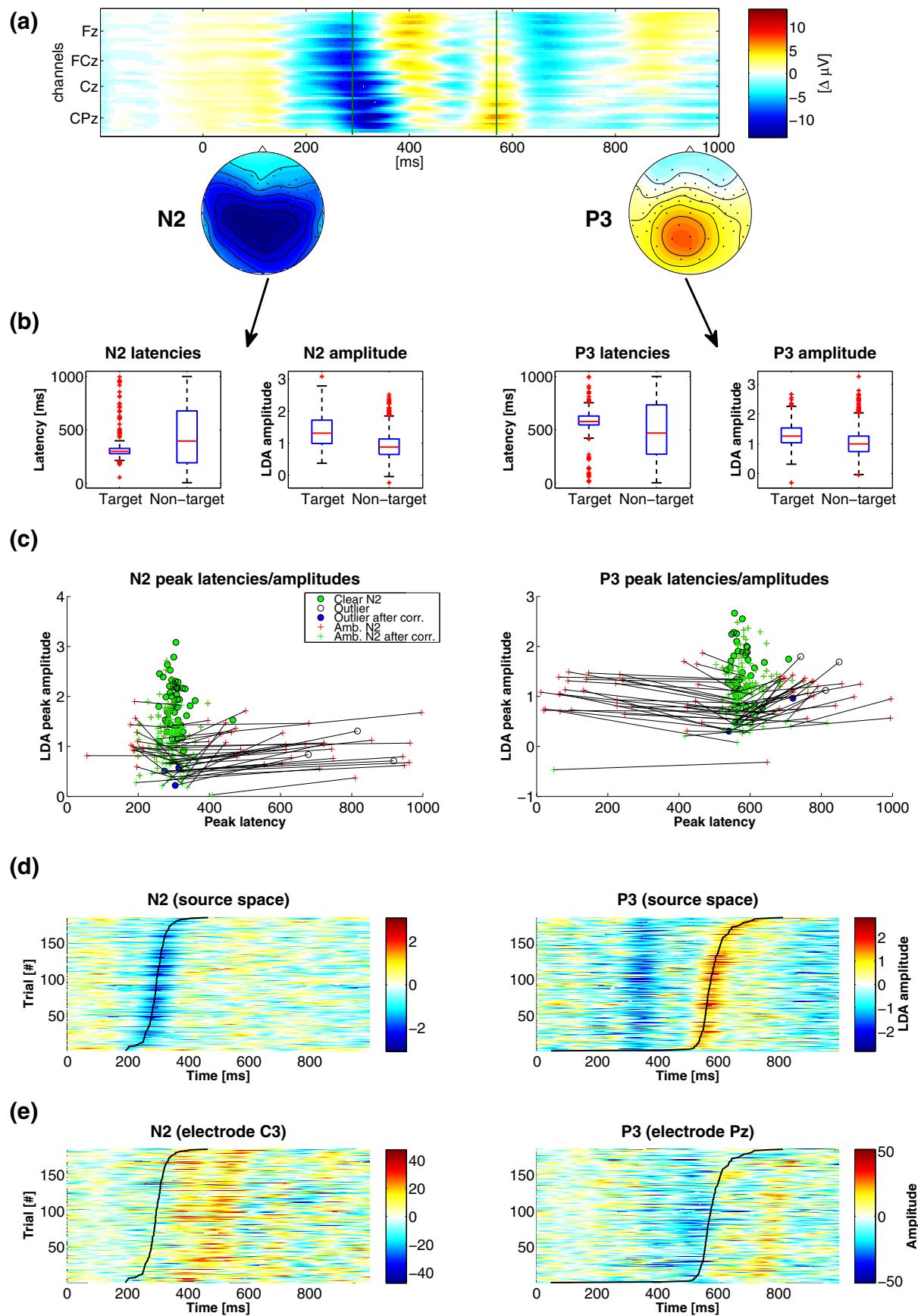
Fig. 4. Results of simulation 2 on correlated sources. Error bars indicate 1 standard deviation. (a) SNR as a function of the number of correlated sources. (b) Correlation between estimated source time series and true source time series.

it steadily increases for the LDA beamformer. It also increases for PCA, but at a clearly lower pace. In line with this, the LDA beamformer yields a significantly higher correlation than the other two methods. Most strikingly, for the LDA beamformer the correlation increases rapidly with the number of correlated sources. In contrast, the LCMV beamformer suffers from signal cancellation and correlation drops towards zero.

We performed an ANOVA with method and number of correlated sources as fixed effects variables and the simulation iteration number

as random effects variable. For SNR, there are main effects of method ($F=699.35, p<.001$) and number of correlated sources ($F=320.09, p<.001$). The interaction between these variables is significant ($F=289.31, p<.001$). For correlation, we found analogous main effects of method ($F=43920.93, p<.001$) and number of correlated sources ($F=26.07, p<.001$), with a significant interaction ($F=634.02, p<.001$). Tukey–Kramer post hoc tests showed that LDA yields significantly higher scores than the other methods, both in terms SNR and correlation.

Fig. 5. Single-trial EEG analysis shown exemplarily for the first subject (analyses for the other subjects are provided as supplemental material). (a) Difference ERP (target minus non-target). Each row represents an EEG channel and amplitudes are color-coded. N2 and P3 peak latencies are indicated by vertical lines. The corresponding spatial patterns are displayed underneath. (b) Boxplots showing the distribution of single-trial peak latencies and peak amplitudes for both target stimuli and non-target stimuli. (c) Latency–amplitude distribution for target trials. For ambiguous trials (crosses) or outliers (circles), peaks are shown before and after correction. (d) Single trials on the source level, sorted w.r.t. peak latency. Each row represents a single trial, with LDA amplitude being color coded. (e) Single trials on channel level. The channel with the largest N2/P3 amplitude was selected for each participant.



EEG experiment

Fig. 5a shows the difference ERP pattern exemplarily for the first participant. Analyses for all the other participants are provided as supplemental material. N2 latency analysis yielded results conforming with the hypothesis that single-trial latencies should be concentrated around 200 ms. However, for participants #4 and #10, a significant proportion of trials violates this rule, indicating a low SNR. P3 latency yielded satisfactory results for all participants.

ERP latencies and amplitudes

As indicated in **Fig. 5b** for one participant, we found mean ERP amplitudes to be higher for targets than for non-targets. This effect is significant on the group level for both N2 ($t = 7.62, p < .0001$) and P3 ($t = 8.51, p < .0001$). In line with this, the standard deviation of the latencies is smaller for target trials than for non-target trials. This effect is significant both for N2 latencies ($t = 13.77, p < .0001$) and P3 latencies ($t = 34.01, p < .0001$).

Fig. 5c depicts the latency–amplitude distribution for the same participant. Green circles represent “clear” trials. Red crosses show the peaks in the “ambiguous” trials, and the black lines indicate the location of the peak in the same trial after correction. The clear trials form a cluster with little variability in latency, showing a higher mean amplitude than ambiguous trials. To substantiate this observation on a group level, we subjected the mean amplitudes and standard deviations of the latencies to a paired-samples t-test. We found that peaks in clear trials have a significantly higher amplitude than peaks in ambiguous trials (N2: $t = 7.23, p < .0001$; P3: $t = 9.73, p < .0001$). Furthermore, the standard deviation of the latencies is significantly lower for clear trials than for ambiguous trials (N2: $t = 4.59, p < .01$; P3: $t = 9.06, p < .0001$).

Fig. 5d shows the LDA source time series for N2 and P3. Each row of the matrix represents a single trial, sorted by peak latency. LDA amplitude is color coded. Peak latencies are superimposed on the plot (black line). For comparison, we produced the same plots using the raw channel time series instead of the source time series (**Fig. 5e**). For each subject, we selected the EEG channel that showed the highest ERP amplitude for the respective component. Trials are ordered in the same way as for the LDA source time series, with superimposed peak latencies. While the ERP peaks are clearly visible in **Fig. 5d**, they cannot be easily spotted in **Fig. 5e**. This illustrates the power of the LDA beamformer to amplify the component of interest and attenuate noise.

Comparing LDA against LCMV and PCA

We repeated the latency analyses using the LCMV beamformer and PCA. For PCA, the ERP covariance matrix was calculated in a 200 ms interval centered on the ERP peak. Since the ground truth is unknown, we devised the standard deviation of the latencies in the “clear” trials as a quality measure for each method. Assuming that signal and noise independently contribute to single-trial ERP variability, the rationale is that the method with the best SNR should result in less variability of the latencies. The standard deviations for each participant, method, and component are depicted in **Fig. 6**. For N2, mean standard deviations across participants are 171 ms (LDA), 225 ms (LCMV), and 206 ms (PCA). For P3, we find 53 ms (LDA), 200 ms (LCMV), and 221 ms (PCA). Two-way ANOVAs with method as fixed effect and participant as random effect revealed main effects of method (N2: $F = 3.56, p < 0.05$; P3: $F = 48.22, p < .0001$). Tukey–Kramer post-hoc tests showed that, for N2, LDA had a significantly smaller standard deviation than LCMV. For P3, LDA had a significantly smaller standard deviation than either other method. No other contrasts were significant.

Connectivity analysis

Fig. 7 shows the results of the connectivity analysis. Both envelope correlation and phase-locking value yield significantly higher values for targets than for non-targets in the 1–3 Hz range (PLV: $p = .002$; envelope correlation: $p = .0479$). We found no significant positive or negative clusters for the LCMV beamformer (PLV: $p = .31$; envelope correlation: $p = .08$) and for PCA (PLV: $p = .06$; envelope correlation: $p = .0679$). At this frequency range, it is likely that the higher SNR ERPs obtained with the LDA beamformer contributed to the connectivity estimates.

Discussion

We identified and elucidated an intimate relationship between linear discriminant analysis (LDA) and LCMV beamforming. The underlying optimization problems can be shown to be equivalent. Consequently, LDA can be considered as a scalar minimum-variance beamformer. The main difference between the two methods is that in LCMV beamforming the spatial pattern is derived from forward calculations, whereas for LDA beamforming the spatial pattern is estimated from the data.

MEG simulations

In MEG simulations, the LDA beamformer consistently outperformed the LCMV beamformer and PCA in terms of SNR and correlation between estimated source time series and real source time series. The LDA beamformer is not only robust to correlated sources, both SNR and correlation actually increase with the number of correlated sources. This contrasts with the LCMV beamformer, which suffers from increasing signal cancellation. The robustness of the LDA beamformer to correlated sources, formally proven in [Appendix A.3](#), stems from the fact that the spatial pattern is estimated from the data. Hence, it is a mixture of the spatial patterns of the correlated sources. In contrast, the standard LCMV beamformer considers one source at a time, permitting signal cancellation from other correlated sources.

EEG experiment

We illustrated two possible applications of the LDA beamformer. First, we extracted single-trial latencies and amplitudes of N2 and P3 components. Latencies exhibited a significantly lower variance for the LDA beamformer than for LCMV beamformer and PCA. Second, we performed connectivity analysis on the virtual N2 and P3 time courses. Using cluster permutation statistics, we found a significant cluster in the 1–3 Hz range for LDA but not for LCMV or PCA. Both analyses corroborate the result of the MEG simulations, namely that LDA exposes the signal of interest at a higher SNR than its rival approaches. In our EEG dataset, obtaining the source time series for a single subject took 2 s for the LDA beamformer and 5.7 s for the LCMV beamformer on average (see supplemental data for a more thorough analysis), showing that the LDA beamformer is also computationally efficient.

The LDA beamformer maximizes ERP signal-to-noise ratio

Let the covariance matrix \mathbf{C} of the raw data around an ERP peak be given as a sum of the covariance of ERP sources \mathbf{C}_s and noise covariance \mathbf{C}_n :

$$\mathbf{C} = \mathbf{C}_s + \mathbf{C}_n. \quad (8)$$

After averaging across E epochs, the covariance of the ERP is

$$\mathbf{C}_{\text{ERP}} = \mathbf{C}_s + \frac{1}{E} \mathbf{C}_n. \quad (9)$$

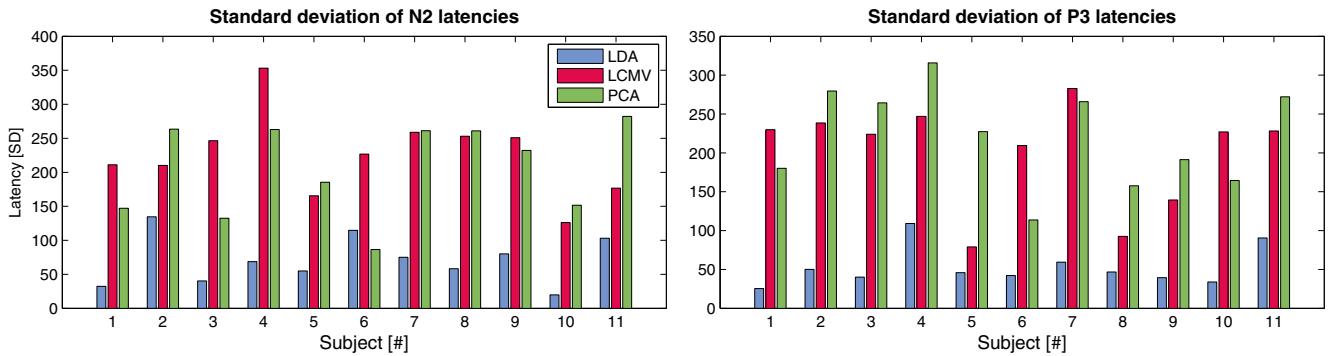


Fig. 6. Standard deviations of latencies in the “clear” trials, for all participants and for N2 and P3 components separately. Bar colors indicate the method (LDA, LCMV, PCA).

Since the ERP source covariance is proportional to the spatial pattern of the ERP source $\mathbf{C}_s \propto \mathbf{p}\mathbf{p}^T$, and using Eqs. (11) and (16), we can restate the LDA optimization as

$$J(\mathbf{w}) \propto \frac{\mathbf{w}^T \mathbf{C}_{\text{ERP}} \mathbf{w}}{\mathbf{w}^T (\mathbf{C}_s + \mathbf{C}_n) \mathbf{w}} \approx \frac{\mathbf{w}^T \mathbf{C}_s \mathbf{w}}{\mathbf{w}^T (\mathbf{C}_s + \mathbf{C}_n) \mathbf{w}} \quad (10)$$

The quantity $J(\mathbf{w})$ is thus an estimate of SNR, and the spatial filter obtained with LDA maximizes this quantity. Interestingly, the numerator of this quantity $\mathbf{w}^T \mathbf{C}_{\text{ERP}} \mathbf{w}$ is what is maximized by PCA. In other words, PCA maximizes the variance of the ERP, but LDA maximizes signal-to-noise ratio. This is the theoretical explanation why, in the MEG simulations, the LDA beamformer usually outperforms PCA.

Relationship to methods based on the generalized Rayleigh quotient

We already elucidated the relationship between LDA beamformers, LCMV beamformers, and PCA. In Eq. (10), we show that the LDA beamformer maximizes ERP covariance (\mathbf{C}_{ERP}) and at the same time minimizes overall covariance ($\mathbf{C}_s + \mathbf{C}_n$). Such a fraction using two different covariance matrices is also known as generalized Rayleigh quotient, and there is a whole family of methods based on this term. A well-known approach is Common Spatial Patterns (CSP), wherein the covariance matrices are taken from two different experimental conditions, and the goal is to find sources that are maximally active in one condition and minimally active in the other condition (Blankertz et al., 2008). Spatio-spectral decomposition capitalizes on the same idea in the frequency domain, with covariance matrices corresponding to different

frequency bands (Nikulin et al., 2011). Similarly, the xDAWN method is used as a preprocessing tool for enhancing ERP responses prior to classification in brain-computer interfaces (Rivet et al., 2009).

Conclusion

It is both a limit and a strength that the LDA beamformer estimates the spatial pattern directly from the data. Not relying on a source model implies that the LDA beamformer cannot be used for source localization. At the same time, one can capitalize on this fact by applying it to datasets wherein a source model is not available, such as many EEG studies. Furthermore, robustness to correlated sources is an implicit property of the LDA beamformer that does not require a priori knowledge or a dipole search. We presented two possible applications of the LDA beamformer: single-trial ERP analysis and connectivity analysis. Concluding, LDA beamforming works out-of-the-box by avoiding source modeling. Due to its simplicity of use, robustness to correlated sources, and optimal signal-to-noise ratio, it is an efficient tool for the single-trial analysis of event-related activity.

Acknowledgements

We would like to thank Marieke Mur and Taylor Schmitz for insightful comments on the manuscript. The research leading to this results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement number 611570, the BK21 program of NRF, BMBF grants 01GQ0850, 01IS14013A-E and

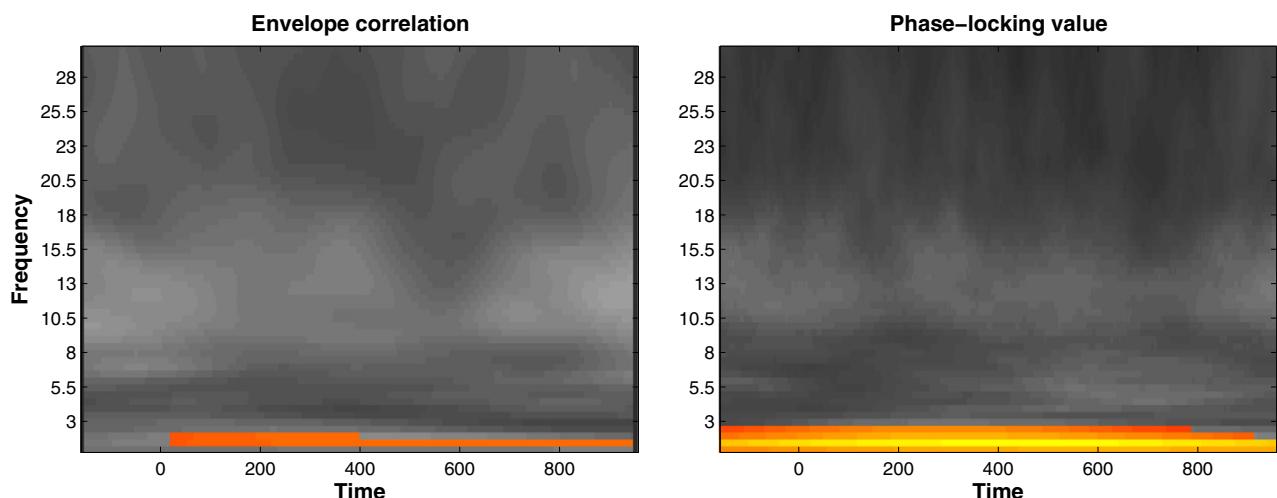


Fig. 7. Results of the cluster permutation test on the connectivity between N2 and P3 time series, contrasting target vs. non-target epochs. The significant cluster in the 1–3 Hz range is shown in color for each connectivity method.

01GQ1115, and DFG grant MU 987/19-1. Correspondence to MST, KRM, and BB.

Appendix A

A.1. Linear discriminant analysis (LDA)

LDA and Fisher's discriminant analysis (FDA) are formally equivalent. As sole difference, LDA makes the additional assumption that each class is Gaussian distributed with equal covariance matrices for both classes. In FDA, a projection \mathbf{w} is sought that maximizes the so-called Fisher criterion defined as

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top S_B \mathbf{w}}{\mathbf{w}^\top S_W \mathbf{w}} \quad (11)$$

where S_B is the $(N \times N)$ between-class scatter matrix and S_W is the $(N \times N)$ within-class scatter matrix for N features (Bishop, 2006). Let \mathbf{p}_1 and \mathbf{p}_2 be the class means, then between-class and within-class scatter can be defined as

$$S_B = (\mathbf{p}_2 - \mathbf{p}_1)(\mathbf{p}_2 - \mathbf{p}_1)^\top,$$

$$S_W = \sum_{n \in \mathcal{C}_1} (\mathbf{x}_n - \mathbf{p}_1)(\mathbf{x}_n - \mathbf{p}_1)^\top + \sum_{n \in \mathcal{C}_2} (\mathbf{x}_n - \mathbf{p}_2)(\mathbf{x}_n - \mathbf{p}_2)^\top \quad (12)$$

where \mathbf{x}_n is the data vector in the n -th trial. \mathcal{C}_1 and \mathcal{C}_2 are two sets that indicate the trials belonging to each of the two classes. Between-class scatter measures the separation of the means. This becomes clear when having a closer look at the numerator of $J(\mathbf{w})$:

$$\mathbf{w}^\top S_B \mathbf{w} = \mathbf{w}^\top (\mathbf{p}_2 - \mathbf{p}_1)(\mathbf{p}_2 - \mathbf{p}_1)^\top \mathbf{w} = (p_1 - p_2)^2 \quad (13)$$

Here, $p_1 = \mathbf{w}^\top \mathbf{p}_1$ and $p_2 = \mathbf{w}^\top \mathbf{p}_2$ are the projections of the two class means on \mathbf{w} . Hence, the numerator gives the squared Euclidean distance between the projected class means. Within-class scatter represents the spread within each class. We can make this explicit by looking at the denominator of $J(\mathbf{w})$:

$$\begin{aligned} \mathbf{w}^\top S_W \mathbf{w} &= \mathbf{w}^\top \left(\sum_{n \in \mathcal{C}_1} (\mathbf{x}_n - \mathbf{p}_1)(\mathbf{x}_n - \mathbf{p}_1)^\top + \sum_{n \in \mathcal{C}_2} (\mathbf{x}_n - \mathbf{p}_2)(\mathbf{x}_n - \mathbf{p}_2)^\top \right) \mathbf{w} \\ &= \sum_{n \in \mathcal{C}_1} (\mathbf{x}_n - \mathbf{p}_1)^2 + \sum_{n \in \mathcal{C}_2} (\mathbf{x}_n - \mathbf{p}_2)^2 \end{aligned} \quad (14)$$

Here, $\mathbf{x}_n = \mathbf{w}^\top \mathbf{x}_n$ is the projected data. This shows that within-class scatter is an unnormalized estimator of the variances within each class. Plugging Eqs. (13) and (14) into Eq. (11), we can rewrite the Fisher criterion as

$$J(\mathbf{w}) = \frac{(p_1 - p_2)^2}{\sum_{n \in \mathcal{C}_1} (\mathbf{x}_n - \mathbf{p}_1)^2 + \sum_{n \in \mathcal{C}_2} (\mathbf{x}_n - \mathbf{p}_2)^2} \quad (15)$$

We thus see that maximizing the Fisher criterion is equivalent to finding a projection \mathbf{w} that maximizes the separation of the projected means (numerator), and at the same time minimizes the sum of the spreads within the projected class distributions (denominator). In fact, if the two class distributions are equal and follow a Gaussian distribution, FDA yields the optimal projection, i.e. a projection that maximizes the correct classification rate.

A.2. Restating the LDA optimization problem

The LDA optimization problem can be rewritten into a form that is analogous to the optimization problem posed in the LCMV beamformer. First, the Fisher criterion is invariant to the scaling of \mathbf{w} so that for any $\alpha \neq 0: J(\alpha \mathbf{w}) = J(\mathbf{w})$.

Hence, the absolute values in the numerator and the denominator are arbitrary and it is valid to fix the numerator to a constant value, say 1. Defining $\mathbf{p} := \mathbf{p}_1 - \mathbf{p}_2$, we can set the numerator to

$$\mathbf{w}^\top S_B \mathbf{w} = (\mathbf{w}^\top (\mathbf{p}_1 - \mathbf{p}_2))^\top = (\mathbf{w}^\top \mathbf{p})^2 = 1. \quad (16)$$

Second, the between-class scatter matrix is proportional to the covariance matrix. Let \mathbf{C}_1 and \mathbf{C}_2 be the covariance matrices in class 1 and class 2. We then get

$$\begin{aligned} S_W &= \sum_{n \in \mathcal{C}_1} (\mathbf{x}_n - \mathbf{p}_1)(\mathbf{x}_n - \mathbf{p}_1)^\top + \sum_{n \in \mathcal{C}_2} (\mathbf{x}_n - \mathbf{p}_2)(\mathbf{x}_n - \mathbf{p}_2)^\top \\ &= |\mathcal{C}_1| \mathbf{C}_1 + |\mathcal{C}_2| \mathbf{C}_2 = (|\mathcal{C}_1| + |\mathcal{C}_2|) \mathbf{C} \end{aligned} \quad (17)$$

where $|\mathcal{C}_1|$ and $|\mathcal{C}_2|$ are the number of trials in each class. The last equality follows from the fact that in LDA we assume the two covariance matrices to be equal. The factor $(|\mathcal{C}_1| + |\mathcal{C}_2|)$ is insignificant to our optimization problem and can be dropped. The optimization problem now reduces to minimizing the projected covariance under the constraint in Eq. (16):

$$\begin{array}{ll} \text{minimize}_{\mathbf{w}} & \mathbf{w}^\top \mathbf{C} \mathbf{w} \\ \text{subject to} & \mathbf{w}^\top \mathbf{p} = 1 \end{array} \quad (18)$$

This is a quadratic programming problem with a linear constraint. It can be translated into the unconstrained minimization of the Lagrangian function by introducing a Lagrangian multiplier λ and incorporating the equality constraint:

$$\mathcal{L}(\mathbf{w}, \lambda) = \mathbf{w}^\top \mathbf{C} \mathbf{w} - \lambda (\mathbf{w}^\top \mathbf{p} - 1). \quad (19)$$

A necessary condition for an extremum is the following partial derivatives to be zero.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 2 \mathbf{C} \mathbf{w} - \lambda \mathbf{p} \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{w}^\top \mathbf{p} - 1 \quad (20)$$

We then set the derivative for \mathbf{w} to zero which yields $\mathbf{w} = \frac{\lambda}{2} \mathbf{C}^{-1} \mathbf{p}$. Inserting this into the derivative for λ and setting the derivative to zero, we obtain

$$\left(\frac{\lambda}{2} \mathbf{C}^{-1} \mathbf{p} \right)^\top \mathbf{p} = 1 \Leftrightarrow \frac{\lambda}{2} = \left(\mathbf{p}^\top \mathbf{C}^{-1} \mathbf{p} \right)^{-1} \quad (21)$$

Inserting Eq. (21) back into Eq. (20) yields the solution

$$\mathbf{w} = \mathbf{C}^{-1} \mathbf{p} \left(\mathbf{p}^\top \mathbf{C}^{-1} \mathbf{p} \right)^{-1} \quad (22)$$

This solution exists if \mathbf{C} is invertible. Moreover, we have $\frac{\partial^2 \mathcal{L}}{\partial \mathbf{w}^2} = \mathbf{C}$, which is positive-definite. This implies that the solution is a local minimum. Given the fact that the optimization problem is convex, Eq. (22) is indeed the global optimum. Regularization can be used to assure that a solution exists.

A.3. Proof of Theorem 1

Our simulations showed that the SNR of the LDA beamformer and the correlation with the true source time series increases with the number correlated sources. Here, we formally prove that these observations hold asymptotically under plausible assumptions on the data. To this end, we first develop a number of definitions and lemmas.

Definition 1. A source is a tuple (s, \mathbf{p}) consisting of a random variable s representing the activity of the source and a corresponding spatial pattern $\mathbf{p} \in \mathbb{R}^n$, where n is the number of sensors.

There are situations wherein adding correlated sources leads to a physical cancellation of the signal in the sensors. As an extreme example, consider a source with a spatial pattern \mathbf{p} and another fully correlated source of equal variance with a spatial pattern $-\mathbf{p}$. On the sensor level, activity in one source is canceled out perfectly by the other source. Obviously, such pairs of correlated sources are not visible in the data and they hence cannot be recovered by any method. We thus have to require that correlated sources do not physically cancel each other out. We call such sources additive. This leads to the following definition:

Definition 2. Let $I \subset \mathbb{N}$ be an index set and $\{(s_i, \mathbf{p}_i) \mid i \in I\}$ be a set of sources. The set of sources is called **additive** if $\|\mathbf{p}_i + \mathbf{p}_j\|_2 > \max\{\|\mathbf{p}_i\|_2, \|\mathbf{p}_j\|_2\}$ for all $i, j \in I$.

Note that $\|\cdot\|_2$ refers to the Euclidean norm. Conceiving of \mathbf{p}_i and \mathbf{p}_j as vectors in \mathbb{R}^n , the definition states that adding two vectors must increase the length of the resulting vector. If the sum would have a smaller norm than either of the constituting vectors, the spatial patterns would partially cancel each other, leading to physical signal loss. For the remainder of this section, we thus assume the following assumption to hold:

Assumption 1. The set of correlated sources is additive. Moreover, the sum of the spatial patterns does not converge, i.e., $\|\sum_{i=1}^{\infty} \mathbf{p}_i\|_2 = \infty$.

In the brain, spatial patterns are not randomly aligned. To empirically investigate whether spatial patterns are usually additive, we investigated the individual head model consisting of 5003 source locations used in the MEG simulations. To obtain a spatial pattern for each voxel, SVD was used on the three orthogonal spatial patterns characterizing each source. We then normalized the patterns to norm 1. We randomly selected two, three, or four sources. In each iteration, the respective number of sources was randomly drawn and the norm of the sum of the spatial patterns was recorded. For each number of sources, 50,000 iterations were run. The results are shown in Fig. 8. Although there are cases of signal cancellation, the norm is likely to increase with the number of sources. This suggests that the non-random nature of the sources facilitates additivity.

Assumption 2. The norm of a spatial pattern \mathbf{p} that is estimated from the data is determined by the amplitude of the source. Hence, if $\tilde{\mathbf{p}}$ represents the spatial pattern of a source at unit activity and α is the amplitude of the source, we have $\mathbf{p} = \alpha \tilde{\mathbf{p}}$.

If a source time series has a large amplitude, the corresponding spatial pattern estimated from the data has a large norm. This means that the norm of the vector \mathbf{p} (but not its direction) varies with the amplitude of the signal. This assumption is important in the proof of Lemma 1 because it implies that the source amplitude is contained in the estimate of the spatial pattern.

Assumption 3. The covariance matrix \mathbf{C} is well-conditioned. In particular, there is a (potentially large) positive number M such that

$$\kappa(\mathbf{C}) = \frac{\lambda_{\max}}{\lambda_{\min}} \leq M$$

where λ_{\max} and λ_{\min} are the maximum and minimum eigenvalues of \mathbf{C} .

Assumption 3 assures that the covariance matrix is not singular. Since we use regularization, this assumption is always met.

Assumption 4. The correlated sources are fully correlated.

This assumption eases the proof, but it is not very restrictive. Every source can be conceived of as the sum of two sources: one source that is fully correlated with a reference source, and another source that is uncorrelated. The structure of the proof is as follows:

1. We show that the numerator of SNR is a constant, being equal to the variance of a reference source s_0 (Lemma 1).
2. We show that the filter norm $\|\mathbf{w}\|_2$ tends to zero as the number of correlated sources tends to infinity (Lemma 2).
3. Both lemmas in conjunction imply that the SNR tends to infinity as the number of correlated sources tends to infinity.

Lemma 1. Let $k \in \mathbb{N}$ be the number of correlated sources with $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k \in \mathbb{R}^n$ being the corresponding spatial patterns of the sources that can be observed in the data. Let s_0 be a reference source representing the common activity of the correlated sources. If $\mathbf{w} \in \mathbb{R}^n$ is the corresponding LDA beamformer solution, then $\sigma_{\text{signal}}^2 = \mathbb{E}\{s_0^2\}$.

Proof. Since the signal sources are fully correlated, we can rewrite each source in terms of a reference source s_0 , so that $s_i = \beta_i s_0$ for a $\beta_i > 0$. Then, β_i is a scaling factor that represents the amplitude of source i .

Furthermore, according to Assumption 2, the norm of the pattern reflects the amplitude of the source. Let $\tilde{\mathbf{p}}_i$ be the spatial pattern of the source at unit activity and $\mathbf{p}_i = \beta_i \tilde{\mathbf{p}}_i$. Then we have

$$\langle \mathbf{w}, \tilde{\mathbf{p}}_i \rangle \beta_i s_0 = \langle \mathbf{w}, \mathbf{p}_i \rangle s_0. \quad (23)$$

In other words, the β_i 's are absorbed in the pattern estimates. This simplifies the signal variance given in Eq. (7):

$$\sigma_{\text{signal}}^2 = \mathbb{E}\{s_0^2\} \left(\sum_{i=1}^k \langle \mathbf{w}, \mathbf{p}_i \rangle^2 + 2 \sum_{i,j \in \{1, \dots, k\}; i < j} \langle \mathbf{w}, \mathbf{p}_i \rangle \langle \mathbf{w}, \mathbf{p}_j \rangle \right) \quad (24)$$

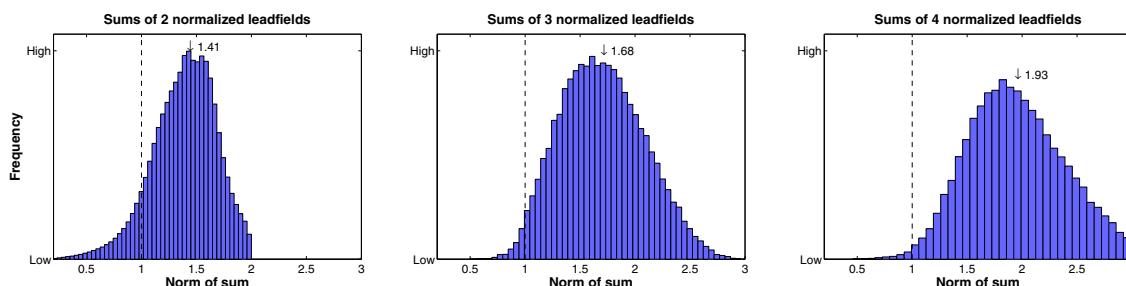


Fig. 8. The effect of summing normalized spatial patterns on the resulting norm of the sum. Histograms are shown for two, three, and four sources. The arrows point at the respective medians. The dashed line indicates the point of norm 1. Points left of the dashed line indicate spatial pattern cancellation.

where $\mathbb{E}\{s_0^2\}$ is the variance of the reference source which is independent of \mathbf{w} and hence a constant. Consequently, we can consider the term in the brackets. Solving the linear constraint of the LDA beamformer for the k -th summand, we obtain

$$\langle \mathbf{w}, \mathbf{p}_k \rangle = 1 - \sum_{i=1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle.$$

This yields for the term in the brackets:

$$\begin{aligned} & \sum_{i=1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle^2 + \left(1 - \sum_{i=1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle \right)^2 + 2 \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle \langle \mathbf{w}, \mathbf{p}_j \rangle \\ & + 2 \sum_{i=1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle \left(1 - \sum_{i=1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle \right) = 2 \sum_{i=1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle^2 \\ & + 1 + 4 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle \langle \mathbf{w}, \mathbf{p}_j \rangle - 2 \sum_{i=1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle + 2 \sum_{i=1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle \\ & - 2 \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \langle \mathbf{w}, \mathbf{p}_i \rangle \langle \mathbf{w}, \mathbf{p}_j \rangle = 1 \end{aligned} \quad (25)$$

and hence the result is $\sigma_{\text{signal}}^2 = \mathbb{E}\{s_0^2\}$.

Lemma 2. (Norm of spatial filter) Let $S_k = \{(s_i, \mathbf{p}_i) | i \in I\}$ be a set of noise and signal sources indexed by I and containing k correlated signal sources. S_{k+1} is obtained from S_k by adding a correlated source $S_{k+1} = S_k \cup (s_{k+1}, \mathbf{p}_{k+1})$. Let $\mathbf{w}_k \in \mathbb{R}^n$ be the LDA beamformer corresponding to a model with k correlated sources.

Then $\|\mathbf{w}_k\|_2 \xrightarrow{k \rightarrow \infty} 0$.

Proof. By the cosine law, we have $\cos \varphi_k = (\|\mathbf{w}_k\|_2 \|\mathbf{p}_k\|_2)^{-1}$. Defining $\tilde{\mathbf{p}}_k = \mathbf{p}_k / \|\mathbf{p}_k\|_2$ and inserting the definition of \mathbf{w} into the cosine formula, we obtain

$$\cos \varphi_k = \frac{\|\mathbf{p}_k\|_2^2 |\tilde{\mathbf{p}}_k^T \mathbf{C}_k^{-1} \tilde{\mathbf{p}}_k|}{\|\mathbf{p}_k\|_2^2 \|_2 \mathbf{C}_k^{-1} \tilde{\mathbf{p}}_k\|_2} \geq \frac{|\lambda_{\min}|}{|\lambda_{\max}|} = \kappa(\mathbf{C}_k)^{-1} \geq M^{-1} > 0 \quad (26)$$

Solving the cosine formula for $\|\mathbf{w}_k\|_2$ we obtain

$$\|\mathbf{w}_k\|_2 = \frac{1}{\cos \varphi_k \|\mathbf{p}_k\|_2} \leq \frac{M}{\|\mathbf{p}_k\|_2} \quad (27)$$

Since M is constant and $\|\mathbf{p}_k\|_2 \xrightarrow{k \rightarrow \infty} \infty$ we have $\|\mathbf{w}_k\|_2 \xrightarrow{k \rightarrow \infty} 0$.

Proof of Theorem 1.

Let $\mathbf{w}_{\max} := \arg \max_{\|\mathbf{w}\|_2=1} \sigma_{\text{noise}}^2$ and let $\tilde{\mathbf{w}}_k := \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|_2}$. Then

$$\begin{aligned} \sigma_{\text{noise}}^2 &= \mathbb{E} \left\{ \sum_{i \in N} \langle \mathbf{w}_k, \mathbf{p}_i \rangle^2 s_i^2 \right\} + 2 \mathbb{E} \left\{ \sum_{i,j \in N; i < j} \langle \mathbf{w}_k, \mathbf{p}_i \rangle s_i s_j \langle \mathbf{w}_k, \mathbf{p}_j \rangle \right\} \\ &= \|\mathbf{w}_k\|_2^2 \left(\mathbb{E} \left\{ \sum_{i \in N} \langle \tilde{\mathbf{w}}_k, \mathbf{p}_i \rangle^2 s_i^2 \right\} + 2 \mathbb{E} \left\{ \sum_{i,j \in N; i < j} \langle \tilde{\mathbf{w}}_k, \mathbf{p}_i \rangle s_i s_j \langle \tilde{\mathbf{w}}_k, \mathbf{p}_j \rangle \right\} \right) \\ &\leq \|\mathbf{w}_k\|_2^2 \left(\mathbb{E} \left\{ \sum_{i \in N} \langle \mathbf{w}_{\max}, \mathbf{p}_i \rangle^2 s_i^2 \right\} + 2 \mathbb{E} \left\{ \sum_{i,j \in N; i < j} \langle \mathbf{w}_{\max}, \mathbf{p}_i \rangle s_i s_j \langle \mathbf{w}_{\max}, \mathbf{p}_j \rangle \right\} \right) \\ &\xrightarrow{k \rightarrow \infty} 0 \end{aligned} \quad (28)$$

We thus have $\frac{\sigma_{\text{signal}}^2}{\sigma_{\text{noise}}^2} = \text{SNR} \xrightarrow{k \rightarrow \infty} \infty$. For the correlation between the true source time series and time series recovered by the beamformer we get the following result:

$$\text{Corr}(s_0, \mathbf{w}_k^T \mathbf{x}_k) = \frac{\mathbb{E}\{s_0 \mathbf{w}_k^T \mathbf{x}_k\}}{\sqrt{\mathbb{E}\{s_0^2\} \mathbb{E}\{(\mathbf{w}_k^T \mathbf{x}_k)^2\}}} \quad (29)$$

For the numerator we get the result

$$\begin{aligned} & \mathbb{E} \left\{ s_0 \mathbf{w}_k^T \left(\sum_{i \in S} \mathbf{p}_i s_0 + \sum_{j \in N} \mathbf{p}_j s_j \right) \right\} \\ &= \mathbb{E}\{s_0^2\} \underbrace{\mathbf{w}_k^T \sum_{i \in S} \mathbf{p}_i}_{=1} + \sum_{j \in N} \mathbf{w}_k^T \mathbf{p}_j \underbrace{\mathbb{E}\{s_0 s_j\}}_{=0} = \mathbb{E}\{s_0^2\} \end{aligned} \quad (30)$$

For the denominator we get the result

$$\sqrt{\mathbb{E}\{s_0^2\} (\sigma_{\text{signal}}^2 + \sigma_{\text{noise}}^2)} = \sqrt{\mathbb{E}\{s_0^2\}^2 + \mathbb{E}\{s_0^2\} \sigma_{\text{noise}}^2} \xrightarrow{k \rightarrow \infty} \mathbb{E}\{s_0^2\} \quad (31)$$

Hence $\text{Corr}(s_0, \mathbf{w}_k^T \mathbf{x}_k) \xrightarrow{k \rightarrow \infty} 1$.

References

- Aricò, P., Aloise, F., Schettini, F., Salinari, S., Mattia, D., Cincotti, F., 2014. Influence of P300 latency jitter on event related potential-based brain-computer interface performance. *J. Neural Eng.* 11 (3), 035008. <http://dx.doi.org/10.1088/1741-2560/11/3/035008>
- Bartz, D., Müller, K.-R., 2013. Generalizing analytic shrinkage for arbitrary covariance structures. In: Burges, C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K. (Eds.), *Advances in Neural Information Processing Systems 26*, pp. 1869–1877.
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. Springer.
- Blankertz, B., Tomioka, R., Lemm, S., Kawabata, M., Müller, K.-R., 2008. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process. Mag.* 25 (1), 41–56.
- Blankertz, B., Lemm, S., Treder, M.S., Haufe, S., Müller, K.-R., 2011. Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56, 814–825. <http://dx.doi.org/10.1016/j.neuroimage.2010.06.048>
- Bledowski, C., Prvulovic, D., Scherg, K.H.M., Wibral, M., Goebel, R., Linden, D.E.J., 2004. Localizing P300 generators in visual target and distractor processing: a combined event-related potential and functional magnetic resonance imaging study. *J. Neurosci.* 24 (42), 9353–9360.
- Brookes, M.J., Stevenson, C.M., Barnes, G.R., Hillebrand, A., Simpson, M.I., Francis, S.T., Morris, P.G., 2007. Beamformer reconstruction of correlated sources using a modified source model. *NeuroImage* 34 (4), 1454–1465. <http://dx.doi.org/10.1016/j.neuroimage.2006.11.012> (URL <http://www.sciencedirect.com/science/article/pii/S1053811906011062>)
- Dähne, S., Bießmann, F., Samek, W., Haufe, S., Goltz, D., Gundlach, C., Villringer, A., Fazli, S., Müller, K.-R., 2015. Multivariate machine learning methods for fusing functional multimodal neuroimaging data. *Proc. IEEE* 103 (9), 1507–1530. <http://dx.doi.org/10.1109/JPROC.2015.2425807>
- D'Avanzo, C., Schiff, S., Amadio, P., Sparacino, G., 2011. A bayesian method to estimate single-trial event-related potentials with application to the study of the P300 variability. *J. Neurosci. Methods* 198 (1), 114–124.
- Diwakar, M., Huang, M.-X., Srinivasan, R., Harrington, D.L., Robb, A., Angeles, A., Muzzatti, L., Pakdaman, R., Song, T., Theilmann, R.J., Lee, R.R., 2011. Dual-core beamformer for obtaining highly correlated neuronal networks in MEG. *NeuroImage* 54 (1), 253–263. <http://dx.doi.org/10.1016/j.neuroimage.2010.07.023> (URL <http://www.sciencedirect.com/science/article/pii/S1053811910009754>)
- Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K.-R., Blankertz, B., 2012. Enhanced performance by a hybrid NIRS-EEG brain computer interface. *NeuroImage* 59 (1), 519–529. <http://dx.doi.org/10.1016/j.neuroimage.2011.07.084> (open Access).
- Fazli, S., Dähne, S., Samek, W., Bießmann, F., Müller, K.-R., 2015. Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based brain-computer interfaces. *Proc. IEEE* 103 (6), 891–906. <http://dx.doi.org/10.1109/JPROC.2015.2413993>
- Folstein, J.R., van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170. <http://dx.doi.org/10.1111/j.1469-8986.2007.00602.x>
- Galka, A., Yamashita, O., Ozaki, T., Biscay, R., Valdés-Sosa, P., 2004. A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. *NeuroImage* 23 (2), 435–453. <http://dx.doi.org/10.1016/j.neuroimage.2004.02.022> (URL <http://www.sciencedirect.com/science/article/pii/S1053811904001144>)
- Gaspar, C.M., Rousselet, G.A., Pernet, C.R., 2011. Reliability of ERP and single-trial analyses. *NeuroImage* 58 (2), 620–629.

- Gross, J., Ioannides, A.A., 1999. Linear transformations of data space in MEG. *Phys. Med. Biol.* 44 (8), 2081–2097.
- Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J.R., van Wassenhove, V., Wibral, M., Schöffelen, J.-M., 2013. Good practice for conducting and reporting MEG research. *NeuroImage* 65 (0), 349–363. <http://dx.doi.org/10.1016/j.neuroimage.2012.10.001> (URL <http://www.sciencedirect.com/science/article/pii/S1053811912009895>).
- Grosse-Wentrup, M., Liefhold, C., Gramann, K., Buss, M., 2009. Beamforming in noninvasive brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 56, 1209–1219.
- Hauße, S., Treder, M.S., Gugler, M.F., Sagebaum, M., Curio, G., Blankertz, B., 2011. EEG potentials predict upcoming emergency brakings during simulated driving. *J. Neural Eng.* 8, 056001 (URL <http://iopscience.iop.org/1741-2552/8/5/056001>).
- Hauße, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., Bießmann, F., 2014a. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* 87, 96–110.
- Hauße, S., Kim, J.-W., Kim, I.-H., Sonnleitner, A., Schrauf, M., Curio, G., Blankertz, B., 2014b. Electrophysiology-based detection of emergency braking intention in real-world driving. *J. Neural Eng.* 11 (5), 056011 (URL <http://stacks.iop.org/1741-2552/11/056011>).
- Hauße, O., Stenroos, M., 2013. A framework for the design of flexible cross-talk functions for spatial filtering of EEG/MEG data: DeFlect. *Hum. Brain Mapp.* 35 (4), 1642–1653.
- Hauße, O., Wakeman, D.G., Henson, R., 2011. Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *NeuroImage* 54 (3), 1966–1974.
- Hipp, J.F., Hawellek, D.J., Corbetta, M., Siegel, M., Engel, A.K., 2012. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat. Neurosci.* 15 (6), 884–890. <http://dx.doi.org/10.1038/nn.3101> (URL <http://www.ncbi.nlm.nih.gov/pubmed/22561454>).
- Höhne, J., Krenzlin, K., Dähne, S., Tangermann, M., 2012. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *J. Neural Eng.* 9 (4), 045003 (URL <http://stacks.iop.org/1741-2552/9/i=4/a=045003>).
- Hu, L., Mouraux, A., Hu, Y., Iannetti, G.D., 2010. A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials. *NeuroImage* 50 (1), 99–111.
- Hui, H.B., Pantazis, D., Bressler, S.L., Leahy, R.M., 2010. Identifying true cortical interactions in MEG using the nulling beamformer. *NeuroImage* 49 (4), 3161.
- Katayama, J., Polich, J., 2009. Auditory and visual P300 topography from a 3 stimulus paradigm. *Clin. Neurophysiol.* 110 (3), 463–468.
- Kawanabe, M., Samek, W., Müller, K.-R., Vidaurri, C., 2014. Robust common spatial filters with a maxmin approach. *Neural Comput.* 26 (2), 349–376.
- Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.-R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kincses, W., 2007. Improving human performance in a real operating environment through real-time mental workload detection. In: Dornhege, G., Millán, J. del R., Hinterberger, T., McFarland, D., Müller, K.-R. (Eds.), *Toward Brain-Computer Interfacing*. MIT press, Cambridge, MA, pp. 409–422.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* 103 (10), 3863–3868. <http://dx.doi.org/10.1073/pnas.0600244103> (URL <http://www.ncbi.nlm.nih.gov/pubmed/16537458>).
- Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8 (4), 194–208 (URL <http://www.ncbi.nlm.nih.gov/pubmed/10619414>).
- Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88, 365–411.
- Lemm, S., Blankertz, B., Dickhaus, T., Mäkläler, K.-R., 2011. Introduction to machine learning for brain imaging. *NeuroImage* 56 (2), 387–399. <http://dx.doi.org/10.1016/j.neuroimage.2010.11.004> (multivariate Decoding and Brain Reading URL <http://www.sciencedirect.com/science/article/pii/S1053811910014163>).
- Li, R., Keil, A., Principe, J.C., 2009. Single-trial P300 estimation with a spatiotemporal filtering method. *J. Neurosci. Methods* 177, 488–496.
- Lopes da Silva, F., 2013. EEG and MEG: relevance to neuroscience. *Neuron* 80 (5), 1112–1128. <http://dx.doi.org/10.1016/j.neuron.2013.10.017> (URL <http://www.ncbi.nlm.nih.gov/pubmed/24314724>).
- Müller, K.-R., Anderson, C.W., Birch, G.E., 2003. Linear and non-linear methods for brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11 (2), 165–169.
- Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B., 2008. Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *J. Neurosci. Methods* 167 (1), 82–90. <http://dx.doi.org/10.1016/j.jneumeth.2007.09.022>.
- Mur, M., Bandettini, P.A., Kriegeskorte, N., 2009. Revealing representational content with pattern-information fMRI—an introductory guide. *Soc. Cogn. Affect. Neurosci.* 4 (1), 101–109. <http://dx.doi.org/10.1093/scan/nsn044> (URL <http://www.ncbi.nlm.nih.gov/pubmed/19151374>).
- Nikulin, V.V., Nolte, G., Curio, G., 2011. A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition. *NeuroImage* 55, 1528–1535. <http://dx.doi.org/10.1016/j.neuroimage.2011.01.057>.
- Nolte, G., 2003. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys. Med. Biol.* 48 (22), 3637–3652.
- Oostenveld, R., Fries, P., Maris, E., Schöffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869. <http://dx.doi.org/10.1155/2011/156869>.
- Parra, L.C., Spence, C.D., Gerson, A.D., Sajda, P., 2005. Recipes for the linear analysis of EEG. *NeuroImage* 28 (2), 326–341.
- Porbadnik, A.K., Scholler, S., Blankertz, B., Ritz, A., Born, M., Scholl, R., Müller, K.-R., Curio, G., Treder, M.S., 2011. Revealing the neural response to imperceptible peripheral flicker with machine learning, in: Conf Proc IEEE Eng. Med. Biol. Soc. 2011, 3692–3695.
- Porbadnik, A.K., Treder, M.S., Blankertz, B., Antons, J.-N., Schleicher, R., Möller, S., Curio, G., Müller, K.-R., 2013. Single-trial analysis of the neural correlates of speech quality perception. *J. Neural Eng.* 10 (5), 056003.
- Ritter, W., Simson, R., Friedman, H.G.V.D., 1979. A brain event related to the making of a sensory discrimination. *Science* 203 (4387), 1358–1361.
- Rivet, B., Souloumiac, A., Attina, V., Gibert, G., 2009. Xdawn algorithm to enhance evoked potentials: application to brain-computer interface. *IEEE Trans. Biomed. Eng.* 56 (8), 2035–2043. <http://dx.doi.org/10.1109/TBME.2009.2012869> (URL <http://www.ncbi.nlm.nih.gov/pubmed/19174332>).
- Samek, W., Kawanabe, M., Müller, K.-R., 2014. Divergence-based framework for common spatial patterns algorithms. *IEEE Rev. Biomed. Eng.* 7, 50–72. <http://dx.doi.org/10.1109/RBME.2013.2290621>.
- Schäfer, J., Strimmer, K., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4 (Article32).
- Schöller, S., Bosse, S., Treder, M.S., Blankertz, B., Curio, G., Müller, K.-R., Wiegand, T., 2012. Towards a direct measure of video quality perception using EEG. *IEEE Trans. Image Process.* 21 (5), 2619–2629. <http://dx.doi.org/10.1109/TIP.2012.2187672>.
- Schreuder, M., Blankertz, B., Tangermann, M., 2010. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. *PLoS One* 5 (4), e9813.
- Sekihara, K., Nagarajan, S.S., Poeppel, D., Marantz, A., 2004. Asymptotic SNR of scalar and vector minimum-variance beamformers for neuromagnetic source reconstruction. *IEEE Trans. Biomed. Eng.* 51 (10), 1726–1734.
- Sonnleitner, A., Treder, M.S., Simon, M., Willmann, S., Ewald, A., Buchner, A., Schrauf, M., 2014. Analysis and single-trial classification of EEG alpha spindles on prolonged brake reaction times during auditory distraction in a real road-driving study. *Accid. Anal. Prev.* 62, 110–118.
- Thompson, D.E., Warschausky, S., Huggins, J.E., 2013. Classifier-based latency estimation: a novel way to estimate and predict BCI accuracy. *J. Neural Eng.* 10 (1), 016006. <http://dx.doi.org/10.1088/1741-2560/10/1/016006>.
- Tomioka, R., Müller, K.R., 2010. A regularized discriminative framework for EEG analysis with application to brain-computer interface. *NeuroImage* 49, 415–432 (URL <http://dx.doi.org/10.1016/j.neuroimage.2009.07.045>).
- Treder, M.S., Blankertz, B., 2010. (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behav. Brain Funct.* 6, 28 (URL <http://www.behavioralandbrainfunctions.com/content/6/1/28>).
- Treder, M.S., Purwins, H., Miklody, D., Sturm, I., Blankertz, B., 2014. Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *J. Neural Eng.* 11, 026009 (URL <http://iopscience.iop.org/1741-2552/11/2/026009>).
- van Veen, B.D., Buckley, K.M., 1988. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Mag.* 5, 4–24.
- van Veen, B.D., van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* 44, 867–880.
- van Vliet, M., Chumerin, N., De Deyne, S., Wiersema, J., Fias, W., Storms, G., Van Hulle, M., 2016. Single-trial ERP component analysis using a spatiotemporal LCMV beamformer. *IEEE Trans. Biomed. Eng.* 63 (1), 55–66 (URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arNumber=7202846>).
- Weeda, W.D., Grasman, R.P., Waldorp, L.J., van de Laar, M.C., van der Molen, M.W., Huizinga, M.W., 2012. A fast and reliable method for simultaneous waveform, amplitude and latency estimation of single-trial EEG/MEG data. *PLoS One* 7 (6), e38292.
- Wu, W., Wu, C., Gao, S., Liu, B., Li, Y., Gao, X., 2013. Bayesian estimation of ERP components from multicondition and multichannel EEG. *NeuroImage* 88, 319–339. <http://dx.doi.org/10.1016/j.neuroimage.2013.11.028>.