Van Vliet, Marijn; Chumerin, Nikolay; De Deyne, Simon; Wiersema, Jan Roelf; Fias, Wim; Storms, Gerrit; Van Hulle, Marc M.

Single-trial ERP component analysis using a spatiotemporal LCMV beamformer

Published in:
IEEE Transactions on Biomedical Engineering

DOI:
10.1109/TBME.2015.2468588

Published: 01/01/2016

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

April 10, 2017
Author copy. Original published in IEEE Transactions on Biomedical Engineering. DOI: 10.1109/TBME.2015.2468588

Single-trial ERP component analysis using a spatio-temporal LCMV beamformer

Marijn van Vliet[1,2], Nikolay Chumerin[2], Simon De Deyne[3], Jan Roel Wiersema[4], Wim Fias[5], Gerrit Storms[3], and Marc M. Van Hulle[2]

[1]Department of Neuroscience and Biomedical Engineering, Aalto University
[2]Department of Neurosciences, KU Leuven
[3]Department of Experimental Psychology, KU Leuven
[4]Department of Experimental Clinical and Health Psychology, Ghent University
[5]Department of Experimental Psychology, Ghent University

Abstract

Goal: For statistical analysis of event-related potentials (ERPs), there are convincing arguments against averaging across stimuli or subjects. Multivariate filters can be used to isolate an ERP component of interest without the averaging procedure. However, we would like to have certainty that the output of the filter accurately represents the component. *Methods:* We extended the linearly constrained minimum variance (LCMV) beamformer, which is traditionally used as a spatial filter for source localization, to be a flexible spatio-temporal filter for estimating the amplitude of ERP components in sensor space. In a comparison study on both simulated and real data, we demonstrated the strengths and weaknesses of the beamformer as well as a range of supervised learning approaches. *Results:* In the context of measuring the amplitude of a specific ERP component on a single trial basis, we found that the spatio-temporal LCMV beamformer is a filter that accurately captures the component of interest, even in the presence of both structured noise (e.g., other overlapping ERP components) and unstructured noise (e.g., ongoing brain activity and sensor noise). *Conclusion:* The spatio-temporal LCMV beamformer method provides an accurate and intuitive way to conduct analysis of a known ERP component, without averaging across trials or subjects. *Significance:* Eliminating averaging allows us to test more detailed hypotheses and apply more powerful statistical models. For example, it allows the usage of multi-level regression models that can incorporate between subject/stimulus variation as random effects, test multiple effects simultaneously and control confounding effects by partial regression.

1 Introduction

Electroencephalography (EEG) measures the electrical activity that spreads outwards from its origin source through the various tissues and fluids in the head, until it is registered by the electrodes on the scalp.[1] This means that an EEG electrode typically picks up a mixture of signals originating from many different sources in the brain. Likewise, signals originating from a single source are typically picked up by multiple electrodes. In many cases, researchers are interested in the behavior of a specific signal, which is but a single voice within the chatter of all the various processes going on in the brain. The focus in this paper is on isolating signal components that are part of the ERP.[2] These components are time-locked to perceptual, cognitive or motor events and correspond to specific peaks and valleys in the ERP waveform. They are named by their positive/negative deflection and time offset relative to the onset of the event in milliseconds (e.g., P300, N400). Such a component will

[1] Nunez et al., 1997

[2] Münte, Urbach, Düzel, and Kutas, 2000

be referred to here as an ERP component of interest (COI). Of particular interest is the amplitude of a COI. In this study, we assume that increased activity at the neural generator responsible for the COI, translates into a uniform, linear scaling of the COI shape. Whenever we speak of the amplitude of the COI, we refer to this scaling.

1.1 Limitations of the averaging technique

A widely used technique to isolate the ERP is to extract trials (also referred to as epochs) from the ongoing EEG, based on the onsets of event markers, and compute their average. Through this operation, signal components that are time-locked to the event markers are retained while unrelated components are suppressed.[3] Statistical analysis of COI amplitude typically proceeds by taking some measurement on the resulting waveform in a certain region of interest (ROI), i.e.some specific electrodes and time range), usually the mean voltage.[4] We will refer to this method as the ROI-mean measure. A downside of this intuitive approach is that, due to the averaging procedure, it yields only a few data points per subject; usually one for each experimental condition.

[3] Handy, 2004; Luck, 2005

[4] Hoormann, Falkenstein, Schwarzenau, and Hohnsbein, 1998

Since so few data points are produced, studies that employ averaging across trials traditionally follow a design that manipulates a single property of the stimuli or task per experimental condition. Such an experimental design is limiting, as not only does it take time to test different manipulations one by one, but manipulating only a single property of a stimulus can be very difficult. For example, in semantic studies, constructing two word lists where the words differ in only one relevant property (e.g., length, frequency of use, age of acquisition, etcifnextchar...) is almost impossible.[5]

[5] Cutler, 1981

By increasing the number of subjects, experimental designs become possible that enable the use of regression techniques when dealing with ERPs.[6] This opens up the possibility to test the effect of multiple manipulations simultaneously and allows correction for unwanted effects through partial regression. As hypotheses become more intricate and effect sizes become smaller, these designs require ever increasing amounts of subjects.

[6] Boudewyn, Long, and Swaab, 2012; Dambacher, Kliegl, Hofmann, and Jacobs, 2006; Federmeier, Kutas, and Schul, 2010; M. Kutas and Iragui, 1998; Laszlo and Federmeier, 2010; Petten and Van Petten, 2014

An additional disadvantage of averaging across stimuli or subjects is that statistical models are no longer able to incorporate either between-stimulus or between-subject variability. This is referred to as the language-as-a-fixed-effect fallacy and cannot be simply ignored.[7] To address this fallacy, multi-level models, such as linear mixed effects (LME) models are becoming increasingly popular in linguistic studies.[8] While these models can theoretically account for between-stimulus and between-subject effects,[9] they must operate on unaveraged data to do so. Without averaging across trials, the mean or peak voltage in a ROI measure can only reveal very strong effects, given a large amount of subjects. For example, Vossen *et al.*[10] present a study where multi-level models are used for statistical analysis of the ERP on 85 subjects that were administered electrical pain stimuli.

[7] Aarts, Verhage, Veenvliet, Dolan, and van der Sluis, 2014; Clark, 1973

[8] Faraway, 2006; West, Welch, and Galecki, 2007

[9] Baayen, Davidson, and M., 2008

[10] Vossen, van Breukelen, Hermens, van Os, and Lousberg, 2011

1.2 Performance criteria for multivariate techniques

There is clearly an opportunity for methods that do not rely on averaging to isolate a COI. A filter that separates overlapping ERP components should boost signal-to-noise substantially. The application of such filters,[11] falls under the category of techniques known as multivariate analysis. The usefulness of multivariate methods, especially linear ones, for EEG analysis has been acknowledged for a long time[12] and single-trial analysis has been growing in popularity.[13]

[11] e.g. Comon and Jutten, 2010; Niedermeyer and da Silva, 2005; Parra et al., 2003

[12] Donchin, 1969

[13] Pernet, Sajda, and Rousselet, 2011

Multivariate methods can be applied to produce filters that combine the EEG signal from multiple electrodes (i.e.a spatial filter), multiple time samples (referred to in this paper as a temporal filter) or both (a spatio-temporal filter) into one representative value. Since we are interested in estimating the amplitude of a COI in a trial, thus reducing the samples from all electrodes and all time points to a single value, we will be looking at spatio-temporal filters.

If the output of such a filter is used as estimation for the amplitude of a COI, two important performance criteria are that it:

1. correlates well with the actual amplitude of the COI (sensitivity)
2. does *not* correlate with any structured interfering signals, such as other ERP components or eye movements (specificity)

A filter that scores reasonably well on the sensitivity criterion does not necessarily score well on the specificity criterion. The specificity criterion states that it is preferable for any variation of the filter output that is not explained by the amplitude of the COI to be due to unstructured (e.g., zero-mean Gaussian) noise. Take for example a filter which output correlates with the amplitudes of both the N/P150 and N400 components. If a researcher uses this output as estimation for N400 amplitude, he will mistakingly find that the N400 is modulated both by word-frequency (that actually does modulate the N400[14]) and font size (which modulates the N/P150, but not the N400[15]).

[14] Marta Kutas and Federmeier, 2011

[15] Grainger and Holcomb, 2009

1.3 The beamformer technique

In order to create a multivariate filter that both performs well on the criteria above and can be interpreted intuitively, we examined beamformer techniques. Beamformers were originally formulated for processing sonar, radar, and seismic data[16] and have since been applied to EEG as a spatial filter that isolates the signal originating from a specific point on the cortex.[17] They have also been applied to brain–computer interfaces (BCIs) to isolate activity from specific regions of the motor cortex,[18] which allows a user to send commands to a computer system by imagining movement that activates these regions. In this paper, we bring a beamformer algorithm into the context of isolating a COI, even when its source location in the brain is unknown, by extending the original formulation to a spatio-temporal filter. A feature that makes beamformers compelling is that a template of the COI is given as explicit input to the algorithm. It then proceeds to construct a filter that isolates the COI by combining the template with the inversed covariance matrix of the entire signal. The method is very transparent, because the user is in full control of constructing the template of

[16] Van Veen and Buckley, 1988

[17] Van Veen, van Drongelen, Yuchtman, and Suzuki, 1997

[18] Grosse-Wentrup, Liefhold, Gramann, and Buss, 2009

the COI and robust estimation of the covariance matrix for EEG signals is a thoroughly studied subject.[19]

[19] Chen, Wiesel, Eldar, and Hero, 2010; Engemann and Gramfort, 2015; Schäfer and Strimmer, 2005

1.4 Assessment of various multivariate techniques

The appropriateness of a signal processing technique depends on the question the investigator would like to answer and the underlying assumptions placed on the data. In order to demonstrate the circumstances when a beamformer is suitable and when another multivariate method is preferable, a simulation study was performed. Simulated EEG recordings were generated with varying parameters, including the level of structured and unstructured noise, variation of the COI shape between subjects, etcifnextchar... The model was used to analyze the performance (in terms of the two criteria listed above) of the beamformer as well as the traditional ROI-mean measure and a variety of supervised learning approaches. The results demonstrate the strengths and weaknesses of each method when it comes to accurately isolating a COI from the ongoing EEG signal.

Finally, each method was applied to EEG data recorded during a semantic priming experiment.[20] Semantic priming is a commonly used method to study the operation and organization of the semantic processes in the brain (for a review, see[21]). In such an experiment, the subject is given a task that involves responding to a target semantic stimulus, which in this study was a single word. When the target is preceded by a semantically and/or associatively related prime word, it allows the subject to respond more efficiently, lowering the response time (RT) of the subject in a decision task that requires reading the words.[22] The semantic priming effect has also been successfully measured using electroencephalography (EEG), where it manifests itself in the ERP, mainly through a component called the N400.[23] A regression study was performed to analyze the relation between N400 amplitude and several properties of the stimuli. To avoid the language-as-a-fixed-effect fallacy, a linear mixed-effects (LME) model was employed, where both subjects and stimuli were entered as random effects, following the recommendations of Baayen *et al.*.[24]

[20] van Vliet et al., 2014

[21] Marta Kutas and Federmeier, 2011; Neely, 1991

[22] Hutchison, 2003; McNamara, 2005

[23] Marta Kutas and Federmeier, 2011; Marta Kutas and Hillyard, 1980

[24] Baayen et al., 2008

2 Methods

2.1 Linear model of EEG

In this section, we introduce the mathematical model used to generate the artificial EEG data and to discuss the merits of the various multivariate methods. The mathematical notation adopts the convention of denoting variables that represent a scalar by cursive letters, those that represent vectors with bold lowercase letters and those that represent matrices by bold uppercase letters. See table 1 for a summary of all the variables and their meaning.

During a single trial, EEG signals are sampled at n time points at m electrodes, yielding an $m \times n$ matrix $\mathbf{S}$. An entire recording contains r trials, with corresponding matrices $\mathbf{S}_i,\ i = 1, \ldots, r$.

Activity in the brain is modeled as the summed activation of various source equiv-

m	number of EEG electrodes
n	number of time samples in a trial
r	number of trials in a recording
r_{train}	number of trials used for training
s	number of subjects that were recorded
y	the true COI amplitude during a trial
$\hat{y}$	unitless estimation of the COI amplitude
$\mathbf{S}$	$m \times n$ matrix containing the samples at all EEG electrodes and time points during a single trial
$\mathbf{A}$	$m \times n$ matrix containing the shape of the COI during a single trial
$\mathbf{N}$	$m \times n$ matrix containing the summation of all noise sources during a single trial
$\mathbf{x}$	$(mn) \times 1$ vector containing a "column-wise flattened" version of $\mathbf{S}$, constructed by stacking the columns of $\mathbf{S}$
$\mathbf{X}$	$(mn) \times r$ matrix composed by concatenation of r vectors $\mathbf{x}$
$\mathbf{w}$	$(mn) \times 1$ vector representing a spatio-temporal filter

Table 1: Mathematical notation

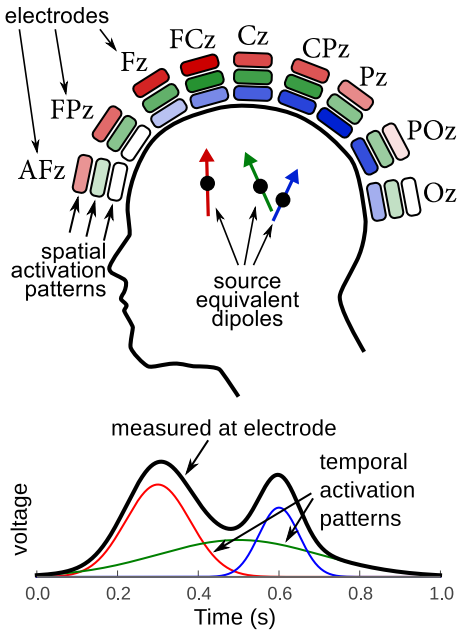


Figure 1: Sketch showing how signals generated at different dipole generators are captured by the EEG electrodes. Activity at the three generators is drawn in blue, red and green. Each electrode records a mixture of the three generators. (top) The spatial activation patterns produced by the generators are shown as rectangles with different color intensities. (bottom) The temporal activation patterns of the generators are shown as curves in different colors. The black curve represents the summation of these patterns as captured by a single EEG electrode.

alent dipole generators in the brain (figure 1). The model assumes that activity at a generator is linearly transferred through volume conduction in the head to the EEG electrodes. When a generator is active, it produces an *activation pattern* on the sensors, as electrodes close to the generator will pick up more activity than electrodes further away (depicted as rectangles in figure 1, top), and also across time as the activity of the generator rises and decays (depicted as curves in figure 1, bottom). One of these generators is used to model the COI and will be referred to as the generator of interest (GOI). Other generators which activity is time-locked to the onset of the trial are regarded as structured noise. They are used to model ERP components that are not the COI. Generators which activity is not time-locked to the onset of the trial are regarded as unstructured noise. They are used to model task irrelevant EEG that would be suppressed by averaging over trials instead of showing as ERP components. The last noise source in the model is sensor noise, which is modeled as zero-mean Gaussian noise.

Let y denote the activity at the GOI during a trial. The activation pattern $\mathbf{A}$, an $m \times n$ matrix, maps activity at y to activity recorded both at the EEG electrodes (spatially) and to time samples (temporally). All other EEG activity during the trial (other ERP components, ongoing brain activity, sensor noise, etc.) is modeled as $m \times n$ matrix $\mathbf{N}$, which can be decomposed into structured noise $\mathbf{N}_{\text{struct}}$, unstructured noise $\mathbf{N}_{\text{unstruct}}$ and sensor noise ϵ:

$$\mathbf{S} = y\mathbf{A} + \mathbf{N}, \tag{1}$$

$$\mathbf{S} = y\mathbf{A} + \mathbf{N}_{\text{struct}} + \mathbf{N}_{\text{unstruct}} + \epsilon. \tag{2}$$

The activation pattern $\mathbf{A}$ can be interpreted as the shape of the COI and y can be interpreted as its amplitude. Each of the methods explored in this study takes a different approach to deduce the amplitude y of the COI, given the signal $\mathbf{S}$ recorded during a trial.

2.2 Multivariate filters

Linear multivariate filters aim to capture activity generated at the GOI by linearly combining EEG samples. These filters aim to eliminate activity generated by noise generators to isolate the GOI, which succeeds if the generators' dipoles differ in location and/or orientation and/or when the dipole's activity differs in timing. A spatial filter combines samples recorded at the same time, but at different electrodes, into one virtual component that optimizes a particular property of the signal, for example correlation of the output with activity of a single generator over time. In the example sketched in figure 1, a possible spatial filter would be the sample at Pz minus the sample at FPz. The Pz channel records the desired activity stronger than FPz and both electrodes capture the unwanted dipoles approximately equal, so the Pz – FPz combination would contain the desired signal with much less noise. A temporal filter combines samples recorded at the same electrode or virtual component, but at different time points, into one representative value that optimizes a particular property of the signal, for example correlation of the output with the amplitude of a COI. In the example, if we would be interested in obtaining an amplitude measurement of

the blue component, the traditional approach would be to take the sample at 600 ms, or the mean of the samples from 500 ms to 700 ms. However, due to overlapping components, it would in this case be more accurate to take the sample at 600 ms, which measures both the blue and green components, and subtract the sample at 500 ms, which measures mostly the green component.

In this study, we are concerned with linear spatio-temporal filters that combine samples from both different electrodes and different time points. A linear spatio-temporal filter can be represented by a vector $\mathbf{w} \in \mathbb{R}^{(mn)\times 1}$ that operates on a data vector $\mathbf{x} \in \mathbb{R}^{(mn)\times 1}$ constructed by stacking all the columns of $\mathbf{S}$. The result of such filtering is a scalar value $\hat{y}$:

$$\hat{y} = \mathbf{w}^\top \mathbf{x}, \tag{3}$$

which can be used as estimation for the amplitude of the COI during the trial (y). The optimal $\mathbf{w}$ depends on both the activation pattern of the COI $\mathbf{A}$ and the noise $\mathbf{N}$. Even if $\mathbf{A}$ is known in advance due to prior studies, it is difficult to know the various noise sources in advance. The most successful multivariate filters therefore contain data-driven elements to estimate $\mathbf{A}$ and/or $\mathbf{N}$ from the recording(s) currently under consideration.

2.3 Supervised learning approach

A popular way to estimate $\mathbf{w}$ employs supervised training. Linear regression models in particular have long been used for effective spatio-temporal filtering.[25] Since implementations of such algorithms are readily available, we skip the implementation details and describe how they can be applied to the problem at hand.

[25] E. Donchin and Heffley, 1978; Blankertz, Benjamin and Lemm, Steven and Treder, Matthias and Haufe, Stefan and Müller, Klaus Robert, 2011; Lemm, Steven and Blankertz, Benjamin and Dickhaus, Thorsten and Müller, Klaus-Robert, 2011

In the context of this study, the learning algorithms operate on a training set which consists of a set training pairs $\{(\mathbf{S}_i, y_i) : i = 1, \ldots, r_{\text{train}}\}$, where $\mathbf{S}_i$ contains the data for a single training trial, y_i is the true (or a good estimate of the) amplitude of the COI in trial $\mathbf{S}_i$, and r_{train} is the number of trials used for training. Given the training set, the learning algorithm will produce a weight vector $\mathbf{w}$, which is interpreted in this paper as a spatio-temporal filter, to estimate the amplitude $\hat{y}$ of the COI given trial data $\mathbf{S}$ in the form of vector $\mathbf{x}$ using (3).

In many cases, it is difficult to obtain reliable estimates of the exact amplitude of the COI to use as training labels y_i. A useful approximation in this case is to limit the training labels to $\{-1, 1\}$, encoding "small" and "large" COI amplitudes respectively. The training data is in this case limited to the trials expected to have an exceptionally small or large COI amplitude. During training, the original regression problem is now substituted by a classification problem, which aims to distinguish between two classes using a decision boundary.

Suitable training data can be obtained by creating a contrast study where the experimental manipulation is designed to only modify the amplitude of the COI. The quality of the training labels and therefore the quality of the filter depends on the suitability of the contrast used (see section 2.7 for our case).

In this study, we evaluated the linear support vector machine (linear support vector

machine (lSVM))[26] as representative supervised learning algorithm. In addition to the straightforward approach of using all samples as features, three variations were evaluated. The first variation was to restrict the features to an ROI subset of the data (restricted-lSVM), where the ROI was the same as the one used for the ROI-mean measurement. The second variation was to train the lSVM on the training data, pooled from all recordings (group-lSVM). The final variation was to first pass the data through a spatial filter, using the xDAWN algorithm,[27] keeping the 4 most descriptive spatial components, and then using an lSVM to perform the final filtering. The covariance matrix for the xDAWN algorithm was estimated using oracle approximating shrinkage (OAS).[28] In all cases, the penalty parameter (C) for the lSVM was optimized using five-fold cross-validation on the training set. A further assortment of alternative supervised learning approaches are evaluated in the supplementary information section.

[26] Burges, 1998

[27] Rivet, Souloumiac, Attina, and Gibert, 2009

[28] Chen et al., 2010

2.4 Beamformer approach

In contrast to the supervised learning approach, beamformers take the shape of the COI (matrix $\mathbf{A}$) as explicit input. Of the various beamformer approaches, the LCMV beamformer[29] seems suitable for our purposes as it was shown to accurately recover the activity at the GOI over time.[30]

[29] Van Veen et al., 1997

[30] Jonmohamadi, Poudel, Innes, and Weiss, 2014

The LCMV beamformer was originally formulated as a spatial filter $\mathbf{w}_{\text{sp}} \in \mathbb{R}^{m\times 1}$ that, when applied to the centered EEG signal $\mathbf{S}$, minimizes the variance of the result $\mathbf{w}_{\text{sp}}^{\mathsf{T}}\mathbf{S}$:

$$\mathbf{w}_{\text{sp}} = \arg\min_{\mathbf{w}_{\text{sp}}} \mathbf{w}_{\text{sp}}^{\mathsf{T}}\mathbf{S}(\mathbf{w}_{\text{sp}}^{\mathsf{T}}\mathbf{S})^{\mathsf{T}} = \arg\min_{\mathbf{w}_{\text{sp}}} \mathbf{w}_{\text{sp}}^{\mathsf{T}}\Sigma_{\text{sp}}\mathbf{w}_{\text{sp}}, \tag{4}$$

where $\Sigma_{\text{sp}} \in \mathbb{R}^{m\times m}$ is the spatial covariance matrix of the signal $\mathbf{S}$, which can be replaced by $\sum_{i=1}^{r}\text{cov}\mathbf{S}_i$. To estimate this covariance matrix, one might consider using robust estimation techniques that employ shrinkage.[31] In this study, as said above, all covariance matrices were estimated using OAS.[32]

[31] Chen et al., 2010; Engemann and Gramfort, 2015; Schäfer and Strimmer, 2005

[32] Chen et al., 2010

To avoid trivial solutions of (4), vector $\mathbf{w}_{\text{sp}}$ must be constrained appropriately, for example, linearly:

$$\mathbf{a}_{\text{sp}}^{\mathsf{T}}\mathbf{w}_{\text{sp}} = 1, \tag{5}$$

where $\mathbf{a}_{\text{sp}} \in \mathbb{R}^{m\times 1}$ is the spatial activation pattern of the GOI (depicted by the colored rectangles in figure 1, top).

Following,[33] the solution of (4) can be found using for example the method of Lagrange multipliers:

[33] Van Veen et al., 1997

$$\mathbf{w}_{\text{sp}} = \frac{\Sigma_{\text{sp}}^{-1}\mathbf{a}_{\text{sp}}}{\mathbf{a}_{\text{sp}}^{\mathsf{T}}\Sigma_{\text{sp}}^{-1}\mathbf{a}_{\text{sp}}}. \tag{6}$$

The formulation of the LCMV beamformer can be expanded to a spatio-temporal filter. Let $\mathbf{X} \in \mathbb{R}^{(mn)\times r}$ be a matrix consisting of r columns $\mathbf{x}_i$, which are the column-wise flattened versions of the corresponding EEG trials $\mathbf{S}_i$ ($i = 1,\ldots,r$), $\Sigma \in \mathbb{R}^{(mn)\times(mn)}$

be the covariance matrix of $\mathbf{X}$, and $\mathbf{a} \in \mathbb{R}^{(mn)\times 1}$ be a vector containing the column-wise flattened version of $\mathbf{A}$. Similarly to the spatial case, the spatio-temporal linearly constrained minimum variance (stLCMV) $\mathbf{w} \in \mathbb{R}^{(mn)\times 1}$ is the result of minimization of the variance of $\mathbf{w}^\intercal \mathbf{X}$ constrained by $\mathbf{a}^\intercal \mathbf{w} = 1$:

$$\mathbf{w} = \frac{\Sigma^{-1}\mathbf{a}}{\mathbf{a}^\intercal \Sigma^{-1}\mathbf{a}}. \tag{7}$$

Alternatively, a simpler spatio-temporal filter can be obtained by sequentially applying to the trial data $\mathbf{S}$ a spatial beamformer $\mathbf{w}_{\text{sp}}$ and then a temporal beamformer $\mathbf{w}_{\text{tmp}}$. To define $\mathbf{w}_{\text{tmp}}$, we define $\mathbf{B} \in \mathbb{R}^{r\times n}$ as a matrix containing the results of applying the spatial beamformer $\mathbf{w}_{\text{sp}}$ to the EEG trials $\mathbf{S}_i$:

$$\mathbf{B} = \begin{pmatrix} \mathbf{w}_{\text{sp}}^\intercal \mathbf{S}_1 \\ \mathbf{w}_{\text{sp}}^\intercal \mathbf{S}_2 \\ \vdots \\ \mathbf{w}_{\text{sp}}^\intercal \mathbf{S}_r \end{pmatrix}, \tag{8}$$

the covariance matrix of which we denote by Σ_{tmp}. Considering $\mathbf{a}_{\text{tmp}} \in \mathbb{R}^{n\times 1}$ as a vector containing the temporal activation pattern of the GOI (e.g., one of the curves in figure 1, bottom), the temporal LCMV beamformer can be expressed as:

$$\mathbf{w}_{\text{tmp}} = \frac{\Sigma_{\text{tmp}}^{-1}\mathbf{a}_{\text{tmp}}}{\mathbf{a}_{\text{tmp}}^\intercal \Sigma_{\text{tmp}}^{-1}\mathbf{a}_{\text{tmp}}}. \tag{9}$$

The spatial and temporal beamformers may be chained together to perform spatio-temporal filtering:

$$\hat{y} = \mathbf{w}_{\text{sp}}^\intercal \mathbf{S}\mathbf{w}_{\text{tmp}}, \tag{10}$$

to obtain a single scalar value $\hat{y}$ from an EEG trial $\mathbf{S}$.

The resulting filter looses the advantage of being able to take interactions between electrodes over time (e.g., a traveling wave) into account, but the number of free parameters is greatly reduced. Where the original spatio-temporal filter had mn (in this study $32 \times 54 = 1728$) free parameters, chaining a separate spatial and temporal filter has $m + n$ (in this study $32 + 54 = 86$). This beamformer will be referred to as the chained LCMV beamformer (chained-LCMV).

2.5 Modeling the activation pattern

While the covariance matrix can be readily computed from the data, it is up to the researcher to provide the beamformer filter with the activation pattern of the GOI: the shape of the COI. In the traditional application of beamformers, the problem of source localization, the activation pattern is computed through a realistic anatomical model of the subject's brain and head.[34] However, since uncovering the location of the GOI in the brain is not our purpose in this study, an anatomical model is not required.

[34] Van Veen et al., 1997

We propose to estimate the activation pattern through the traditional manner of averaging across trials, using data from all available recordings. First, a training set is selected analogous to the one used in section 2.3. The training pairs in the training set were divided into two classes: those known to contain the COI with a large amplitude and those with a small COI amplitude (see section 2.7 for the contrast we used for the evaluation on real EEG data). Grand average ERPs are constructed for both classes and the difference wave is taken as first estimate of the activation pattern. Let matrix $\hat{\mathbf{A}} \in \mathbb{R}^{m \times n}$ denote this activation pattern, which can be interpreted as a template for the COI.

This template can be refined at will by the researcher. In this study, we opted to do this refinement by approximating the activation pattern $\hat{\mathbf{A}}$ as a product of the separate spatial $\hat{\mathbf{a}}_{\text{sp}}$ and temporal $\hat{\mathbf{a}}_{\text{tmp}}$ activation patterns. For the spatial pattern, we used the column of $\hat{\mathbf{A}}$ corresponding to the time point at which the COI reaches its maximum amplitude. To find it, a suitable ROI was first defined by visual inspection of $\hat{\mathbf{A}}$. Let $\mathbf{c} \subset \{1, \ldots, m\}$ denote a set of row indices of $\hat{\mathbf{A}}$ corresponding to the electrodes of interest. Likewise, let $\mathbf{t} \subset \{1, \ldots, n\}$ denote a set of column indices of $\hat{\mathbf{A}}$ corresponding to the samples from the time window of interest. Then:

$$t_{\text{peak}} = \arg\max_{t \in \mathbf{t}} \left| \sum_{c \in \mathbf{c}} \hat{\mathbf{A}}(c, t) \right|, \tag{11}$$

$$\hat{\mathbf{a}}_{\text{sp}} = [\hat{\mathbf{A}}(1, t_{\text{peak}}), \ldots, \hat{\mathbf{A}}(m, t_{\text{peak}})]^{\mathsf{T}} \in \mathbb{R}^{m \times 1}, \tag{12}$$

To construct the temporal activation pattern $\hat{\mathbf{a}}_{\text{tmp}} \in \mathbb{R}^{n \times 1}$, for each recording, a spatial beamformer $\mathbf{w}_{\text{sp}}$ was constructed from $\hat{\mathbf{a}}_{\text{sp}}$ using (6) and applied to the data:

$$\hat{\mathbf{a}}_{\text{tmp}} = \frac{1}{r} \sum_{i=1}^{r} \mathbf{S}_i^{\mathsf{T}} \mathbf{w}_{\text{sp}}. \tag{13}$$

The resulting temporal activation patterns $\hat{\mathbf{a}}_{\text{tmp}}$ were averaged across recordings. Finally, all values outside the range defined by the first and last zero-crossings inside the temporal ROI (from 300 ms to 600 ms) were set to zero to eliminate small deviations from zero at irrelevant time points. The refined spatio-temporal template for the COI is then a product of the spatial and temporal activation patterns:

$$\hat{\mathbf{A}}_{\text{refined}} = \hat{\mathbf{a}}_{\text{sp}} \hat{\mathbf{a}}_{\text{tmp}}^{\mathsf{T}}. \tag{14}$$

2.6 Evaluation on simulated data

In order to evaluate the different methods of measuring COI amplitude, simulated EEG recordings were generated. The software model allows control over the different signal components listed in (2): y, $\mathbf{A}$, $\mathbf{N}_{\text{struct}}$, $\mathbf{N}_{\text{unstruct}}$ and ϵ.

For a pool of 10 subjects, 400 trials were simulated that consisted of virtual recordings at 32 electrodes, distributed over the scalp using the extended 10–20 system. Each simulated trial lasted 1 s and was sampled at 50 Hz, mimicking the properties of the real EEG data discussed later on.

The data were generated by simulating dipoles in a spherical head model. The

activity at the dipoles was modeled as Gaussian curves:

$$f(t) = y\, e^{-\frac{1}{2}\left(\frac{t-t_{\text{peak}}}{\sigma}\right)^2}, \tag{15}$$

where $f(t)$ is the activity (in μV) at the dipole at time t, y is the amplitude of the dipole, the peak activity occurs at t_{peak} and σ determines the speed at which the activity reaches its peak and decays back to zero. For each time point, the spatial activation pattern of each dipole was computed using a three layer bounded element method (BEM) model, where the layers represented the inner skull, outer skull and outer skin boundaries.[35] Each dipole has a location $\mathbf{l} = [l_x, l_y, l_z]$ defined as a x-, y- and z-coordinate, restricted to be within the inner skull layer of the BEM model, and an orientation $\mathbf{o} = [r_x, r_y, r_z]$ defined by Euler rotations around the x-, y- and z-axes.

[35] Gramfort, Alexandre and Papadopoulo, Théodore and Olivi, Emmanuel and Clerc, Maureen, 2011

The COI was simulated as a single dipole (the GOI) at a fixed location $\mathbf{l}_{\text{GOI}} = [0, 0, 0.5]$. Coordinates are normalized so the inner skull sphere has a radius of 1. This places this dipole centered between the auricular points (x-axis), centered between the nasion and inion (y-axis) and raised towards the top of the head (z-axis). Its orientation was $\mathbf{o}_{\text{GOI}} = [\pi, 0, 0]$ (pointing straight up), yielding a spatial activation pattern that centers on the Cz electrode, spreading radially to all other electrodes. The COI peaked at $t_{\text{peak}} = 0.4 + j$, where j is randomly drawn from a uniform distribution defined over the range $[-J_{\text{COI}}, J_{\text{COI}}]$ for each recording, but held constant between trials. The parameter J_{COI} is the first model parameter: the amount of temporal jitter of the COI between subjects. The width of the COI was held constant at $\sigma = 0.05$. For half of the trials, the amplitude y of the COI was 1 μV and for the remaining trials it was 0 μV.

The sensor noise ϵ was modeled for each sample individually, by randomly drawing from a zero-mean, Gaussian distribution with a standard deviation of 0.1.

The structured noise $\mathbf{N}_{\text{struct}}$ was modeled using 20 dipoles with initially random parameters: each parameter value was drawn from a uniform distribution, defined over the range of possible valid values for the parameter. Between trials, all parameters were held constant except for the amplitude of the dipole activity. For each trial, the amplitude parameter y was randomly drawn from a uniform distribution over the range $[0, S_{\text{struct}}]$, where S_{struct} is the second model parameter: the scale of the structured noise.

Finally, the unstructured noise $\mathbf{N}_{\text{unstruct}}$ was modeled using 20 dipoles in the same manner as the structured noise dipoles, except that parameters for the unstructured noise dipoles were randomly drawn for each individual trial: no parameter was held constant. The uniform distribution used for the amplitude parameter was defined over the range $[0, S_{\text{unstruct}}]$, where S_{unstruct} is the third model parameter: the scale of the unstructured noise.

The task for each of the filters was to estimate the amplitude of the COI (y) for each trial, which corresponds to the peak amplitude of the GOI dipole in the model and to y in (2). The total set of trials was split 50-50 into a training and test set. Each set contained 200 trials for each of the 10 subjects, 100 trials containing the COI with an amplitude of 1 μV and 100 trials not containing the COI (its amplitude was 0 μV).

Training labels l were produced using a mixture of the true amplitude y of the COI and the amplitude y_{struct} of one of the structured noise dipoles:

$$l = R_{\text{labels}}\, y + (1 - R_{\text{labels}})\, y_{\text{struct}}, \tag{16}$$

where R_{labels} is the fourth model parameter: the accuracy of the training labels, ranging from 0 (labels follow a noise component) to 1 (labels follow the COI).

Each of the multivariate filters was trained on the training set, plus the corresponding training labels $\mathbf{l}$. Each method then produced an estimate of COI amplitude ($\hat{y}$) for each of the trials in the test set, for which the training labels were withheld.

The entire procedure was ran 10 times in order to assess the variation between simulation runs, producing 100 data sets (10 subjects × 10 runs). The performance of each filter was observed for different values of the four model parameters J_{COI}, S_{struct}, S_{unstruct} and R_{train}. The base model settings were $J_{\text{COI}} = 0$, $S_{\text{unstruct}} = 1$, $S_{\text{struct}} = 3$, $R_{\text{labels}} = 1$. During the simulation, each parameter was changed in isolation, leaving the others at their base values.

2.7 Evaluation on real EEG data

The analysis on real EEG data was conducted on the dataset recorded in,[36] where the COI is the N400 potential. Subjects read a series of sequentially presented words, organized in pairs, and pressed one of two buttons to indicate whether the two words of a word-pair were related or not. The prime word was presented for 200 ms and the target-word for 1000 ms with a stimulus onset asynchrony (SOA) of 500 ms. Since a speeded button response task will generate ERP components that can mask N400 modulations,[37] the subjects performed delayed their button response until a cue was given.

[36] van Vliet et al., 2014

[37] van Vliet et al., 2014

The experiment was performed with 10 university students (3 female, aged 19–27 years), all right-handed and native speakers of Flemish-Dutch. Ethical approval of this study has been granted by an independent ethical committee ("Commissie voor Medische Ethiek" of the UZ Leuven). This study was conducted according to the most recent version of the declaration of Helsinki.

The word-pairs used were a total of 800 Flemish-Dutch word-pairs, selected with varying forward association strength (FAS), as determined from an association norm dataset.[38] In this norm dataset, FAS between a prime and target word is defined as the number of respondents, out of 100, that wrote down the target as first response to the prime word in a free association task. The list of word-pairs consisted of the top 100 strongest related word-pairs (FAS ranged 69–95, mean FAS = 75.62) and 100 word-pairs where the prime and target words were randomly chosen and no record of the word-pair existed in the association norm data, therefore having an assumed FAS of zero. The remaining 600 word-pairs were chosen such that the logarithm of their FAS score is uniformly distributed using a log scale. The log scale was chosen because when the association norm data were analyzed, some properties of the word-pairs that co-vary with the FAS, correlate better with its logarithm than the raw values.[39] All selected words for the stimulus list have a length of 4–6 letters,

[38] De Deyne and Storms, 2008

[39] van Vliet et al., 2014

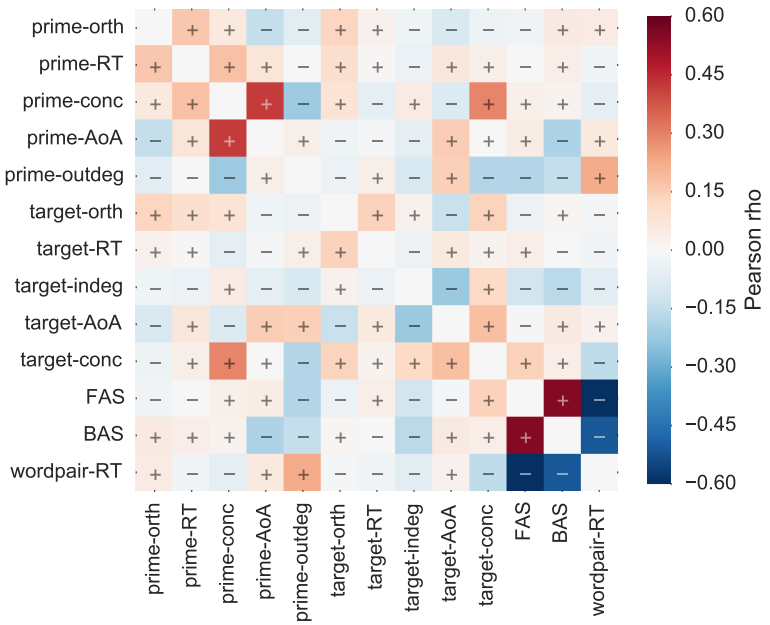


Figure 2: Pearson's correlation coefficients between dependent variables used in this study. See table 2 for a description of each one. The sign of coefficients is indicated with '+' and '-' symbols.

a minimum word frequency of 2 occurrences per 10^6 words in the SUBTLEX-NL corpus[40] and a minimum in-degree of 5 in the association norm dataset.

[40] Keuleers, Brysbaert, and New, 2010

In addition to capturing the button response of the participant, EEG was recorded continuously using 32 active electrodes (extended 10–20 system) with a BioSemi Active II System, having a 5th order frequency filter with a pass band from 0.16 Hz to 100 Hz, and sampled at 2048 Hz. An electrooculogram (EOG) was recorded simultaneously and used to reduce eye artifacts in the EEG using the procedure outlined in.[41] Two electrodes were placed on both mastoids and their average was used as a reference for the EEG.

[41] Croft and Barry, 2000

2.7.1 Stimulus properties

Since the true amplitude of the N400 is unknown, performance was based on a regression analysis with a selection of stimulus properties that have been shown to correlate with semantic priming in earlier regression studies.[42] See table 2 for a complete list and explanation of the stimulus properties used in the regression analysis. The aggregated set covers some strong and weaker correlates with the N400. Pearson's correlation coefficients between all stimulus properties are presented in figure 2.

[42] Hutchison, Balota, Cortese, and Watson, 2008; Laszlo and Federmeier, 2010; Petten and Van Petten, 2014

2.7.2 Data preprocessing

The EEG was bandpass filtered offline between 0.1 Hz and 50 Hz by a 4th order zero-phase Butterworth filter to attenuate large drifts and irrelevant high frequency noise, but retain eye movement artifacts. The EOG was used to attenuate eye artifacts from the EEG signal using the regression method outlined in.[43] After the EOG correction procedure, the signal was band pass filtered again using a tight passband around the frequency range in which the N400 component was found, namely between 0.5 Hz and 15 Hz, by a 4th order zero-phase Butterworth filter. Individual trials were

[43] Croft and Barry, 2000

Property	Description	Ref.
prime-orth	The orthographic neighborhood size of the prime word, i.e.the number of valid Dutch words with a Levenshtein distance of 1 from the prime word. Calculated using the SUBTLEX-NL corpus.	(Keuleers, Brysbaert, and New, 2010)
prime-RT	Reaction time to the prime word in a lexical decision task.	(Keuleers, Diependaele, and Brysbaert, 2010)
prime-outdeg	Number of outgoing links of the prime word, with an association strenght of ≥ 2 in the association norm data.	(De Deyne and Storms, 2008)
prime-age of acquisition (AOA)	Age of acquisition rating for the prime word.	(Brysbaert, Stevens, De Deyne, Voorspoels, and Storms, 2014)
prime-conc	Concreteness rating for the prime word.	(Brysbaert et al., 2014)
target-orth	The orthographic neighborhood size of the target word, i.e.the number of valid Dutch words with a Levenshtein distance of 1 from the target word. Calculated using the SUBTLEX-NL corpus.	(Keuleers, Brysbaert, and New, 2010)
target-RT	Reaction time to the target word in a lexical decision task.	(Keuleers, Diependaele, and Brysbaert, 2010)
target-outdeg	Number of outgoing links of the target word, with an association strenght of ≥ 2 in the association norm data.	(De Deyne and Storms, 2008)
target-AOA	Age of acquisition rating for the target word.	(Brysbaert et al., 2014)
target-conc	Concreteness rating for the target word.	(Brysbaert et al., 2014)
FAS	The logarithm of the forward association strength between the prime and target words.	(De Deyne and Storms, 2008)
BAS	The logarithm of the backward association strength between the prime and target words.	(De Deyne and Storms, 2008)
wordpair-RT	The mean response time of the subjects to the word-pair in a speeded button response task, obtained during a separate recording session several months prior to the current experiment.	(van Vliet et al., 2014)

Table 2: Stimulus properties used in the regression study

obtained by cutting the continuous signal from 0.1 s before the onset of each target stimulus to 1.0 s after. Baseline correction was performed using the average voltage in the interval before the stimulus onset (−100 ms to 0 ms) as baseline value. Before applying any multivariate analysis methods, the signal was further downsampled to 50 Hz to reduce the dimensionality.

2.7.3 Training data

To construct the training set for the supervised learning algorithms and model the activation pattern for the beamformer, a contrast has to be created that will produce both trials with a low and trials with a high N400 amplitude. For the training set, the 100 word-pairs with an FAS of 0 were chosen as the low N400 amplitude condition, and 100 word-pairs with the highest FAS for the high N400 amplitude condition. This contrast in FAS is well known to produce clear differences in N400 amplitude.[44] The trained filters were then applied to the remaining trials.

[44] Bajo, 1988; M. Kutas, 1993; Marta Kutas and Federmeier, 2011

Discarding the training set would mean a limited range of FAS for the trials in the test set, which would potentially eliminate a large portion of the N400 effect. Therefore, COI amplitudes were estimated for the training set as well, by using leave-one-out cross-validation scheme.

2.7.4 Statistics

Regression analysis between stimulus properties and N400 amplitude was done by means of a linear mixed-effects (LME) model. Since the stimulus properties used as independent variables are intercorrelated (figure 2) we have chosen for a univariate approach. Each independent variable is regressed onto the dependent variable in a separate model. All variables were z-transformed, so regression weights represent estimates of the Pearson correlation between the independent and dependent variable. Each regression model was fitted two times, one with both subjects (modeling slopes only) and word-pairs (modeling intercepts only) as random effects, and one with only word-pairs as random effects. If the first model did not achieve a significantly better fit than the latter, as measured using the area under curve (AUC) metric, the latter model was used. Models were fitted using maximum likelihood (ML) for computing the AUC metric, and using restricted maximum likelihood (REML) for computing t-values. This design follows the recommendations of Baayen *et al.*.[45] Degrees of freedom are hard to compute for mixed models and are often in the order of several thousands. Satterthwaite's method[46] was used to estimate them. Degrees of freedom are not provided in this text, as the relationship between t-scores and p-values converges at this number of degrees of freedom.

[45] Baayen et al., 2008

[46] Satterthwaite, 1946

2.8 Software

A full description of the various software packages used in this study and an implementation of the beamformer methods can be found in the supplementary information.

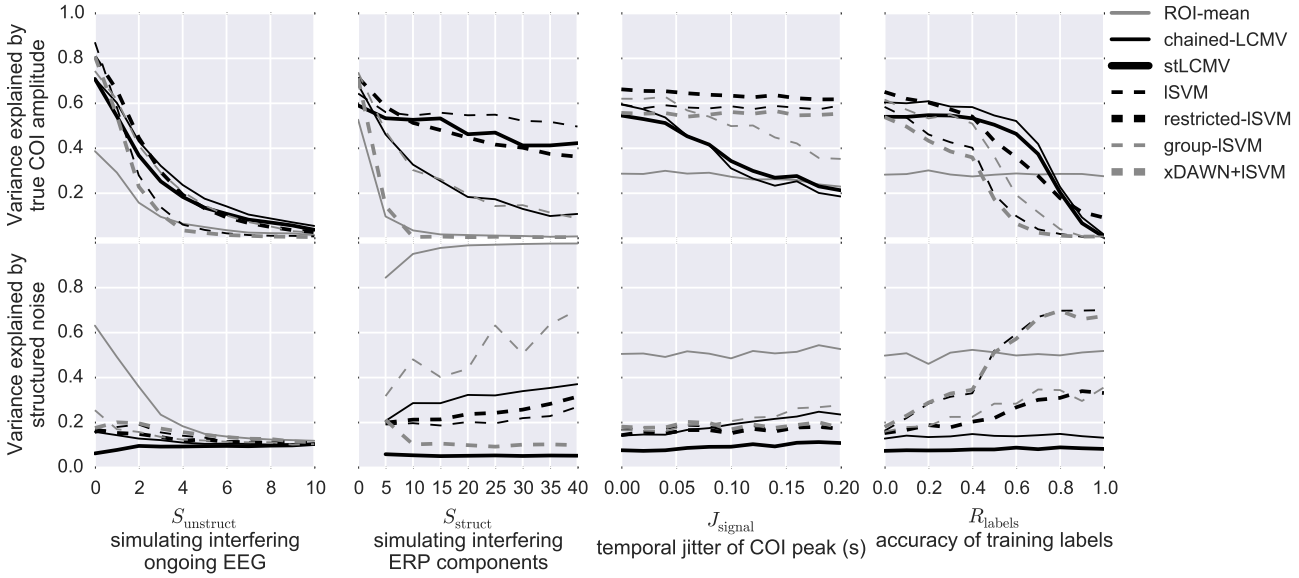


Figure 3: Performance of various techniques to estimate COI amplitude under various simulated conditions. The performance is quantified by two metrics: the variance of the output explained (R^2 stat) by the true amplitude of the COI (top, higher is better) and the variance explained by least squares regression with the structured noise sources (bottom, lower is better). Each curve represents an average of the analysis of 100 data sets (10 subjects × 10 runs). For clarity, confidence intervals of the curves are omitted in this figure. They can be found in the supplementary information.

3 Results

The performance of each multivariate filter on the datasets was assessed, as well as the traditional ROI-mean method.

Different ROI's were tried, and the best performing time range and electrode selections for the ROI-mean and restricted-lSVM methods was selected, namely for the simulation study electrodes FC2, FC5, Cz, C3, C4, CP5, CP6, time window 0.3 s to 0.5 s) and for the real EEG data electrodes CP5, CP6, P3, Pz, P4, PO3 and P04, time range 0.3 s to 0.5 s.

3.1 Simulated data

By modifying the four model parameters, artificial EEG datasets were generated with different properties. The performance of each method was assessed based on the two criteria listed in section 1.2 (figure 3). The first measure was the amount of variance of its output explained by the true amplitude of the COI, measured by regressing the estimated amplitude onto the true amplitude and computing the R^2 metric. Adhering to the sensitivity performance criterion, higher is better for this metric. The second measure was the amount of variance in the filter's output that could be explained by activity of the structured noise dipoles. This was quantified by multivariate-regression of the activity of the 20 noise dipoles onto the method's estimation of the COI amplitude and computing the R^2 metric. This measures how well the method manages to reduce the influence of nearby noise sources on its COI amplitude estimate. Adhering to the specificity criterion, lower is better for this metric. In the simulation, any variance in the method's output not explained

by the true amplitude of the COI or structured noise must necessarily be due to unstructured noise.

The traditional ROI-mean is generally the worst performing method, as the method does not actively counteract noise sources. Without averaging, the ROI-mean method does not measure solely the amplitude of the COI, but a mixture of the COI and the surrounding structured noise sources.

As the unstructured noise amplitude (S_{unstruct}) increases, all methods start failing to properly isolate the COI (figure 3, first column). This is to be expected, as this type of noise is spherical, so there is no possible orthogonal linear projection. The lSVM (with and without the xDAWN spatial filtering step) fails somewhat faster than the other multivariate methods, because the quality of the training data diminishes and it starts overfitting on the unstructured noise. Limiting the amount of features (restricted-lSVM) or adding more training data from other recordings (group-lSVM) effectively counters this behavior. The beamformer methods perform on par with the restricted-lSVM and the group-lSVM. The ROI-mean method initially correlates highly with the unstructured noise, because it doesn't actively counteract it. As the unstructured noise increases, this high correlation disappears as the output now correlates more with the unstructured noise.

A filter's ability to isolate the COI from structured noise sources is dependent on its ability to successfully model both of them (figure 3, second column). The lSVM, restricted-lSVM and stLCMV methods are remarkably capable of countering structured noise sources, even when the stimulated structured noise (S_{struct}) is raised to 40 times the amplitude of the COI. The estimation technique for the template, used by the beamformer approaches, suffered in these extreme conditions, causing the performance of the stLCMV beamformer to drop slightly. As the structured noise increases, the group-lSVM starts failing, because it cannot adapt to the different structured noise dipoles in each recording. Instead, it has to find a solution that isolates the COI from all structured noise dipoles on all 10 recordings that where pooled together. The chained-LCMV is able to adapt to each individual recording, but still fails, probably due its trade-off between model simplicity and power. This is also the case for the xDAWN spatial filter.

As the structured noise increases, its influence on the filters' output is more severe, causing most of the methods to fail the specificity criterion: fluctuations in one of the other ERP components is influencing the estimation of COI amplitude. The stLCMV beamformer is the notable exception, as the correlation between its output and the noise components is unaffected by their amplitude.

For the methods that pool together multiple recordings (i.e. chained-LCMV, stLCMV and group-lSVM), differences in the shape of the COI across recordings is problematic (figure 3, third column). This is especially true for the beamformer methods, as the method we used for constructing the template did not take between-subject variability into account. The beamformers are designed to fail when the given template of the COI doesn't match its actual shape.

Since the output of supervised learning algorithms is closely tied to the training labels, it is sensitive to their accuracy (figure 3, last column). When other noise

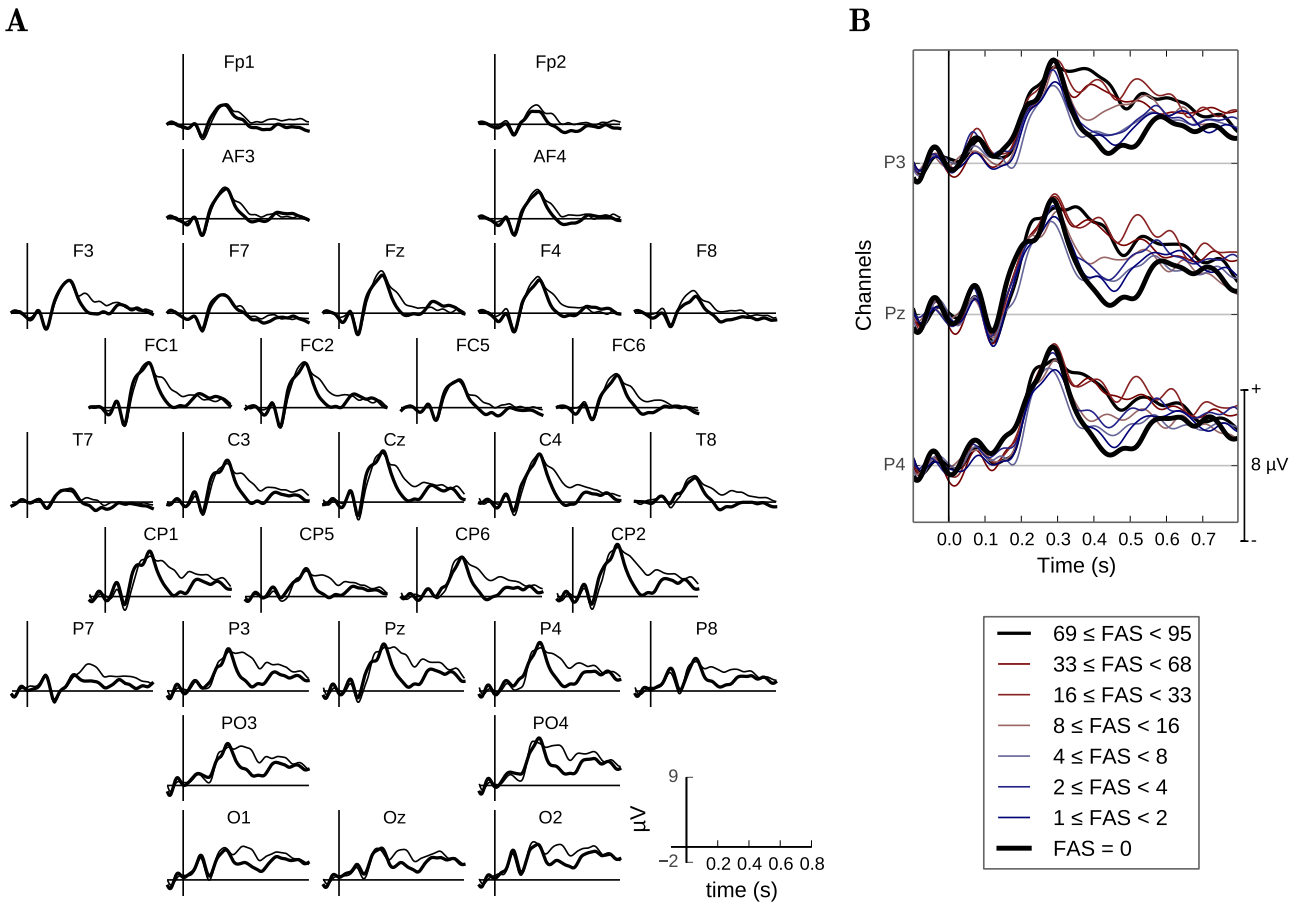


Figure 4: Grand average ERPs in response to word-pairs with increasing FAS during the delayed condition. Word-pairs were sorted by their FAS and grouped into 8 bins of 100 pairs and the average response is shown aligned to the onset of the target word (black vertical line). (a) ERPs of the first and last bins for each channel. This shows the most extreme case of unrelated word-pairs versus strongly related ones. Intermediate bins have been omitted here for clarity. (b) Blowup of three electrodes showing the ERP of each bin.

components correlate with the training labels in addition to the COI, the lSVM is unable to separate them, resulting in a decreasing correlation with the actual COI and an increasing correlation with the noise components. This can be alleviated by restricting the features to exclude some noise components (restricted-lSVM) or by pooling together more data, which increased the overall reliability of the training labels in this case (grouped-lSVM). The method used in this study to design the template for the COI used by the beamformers, also relies on the training labels. However, the beamformer methods are quite robust to inaccuracies in the training labels, as data are pooled across recordings and values outside the ROI were explicitly set to zero in the template. Of particular note is the fact that even when the training labels follow one of the structured noise sources instead of the COI, the correlation between beamformers' output and the structured noise sources remains low. Finally, since the ROI-mean method does not rely on training labels, its output is unaffected.

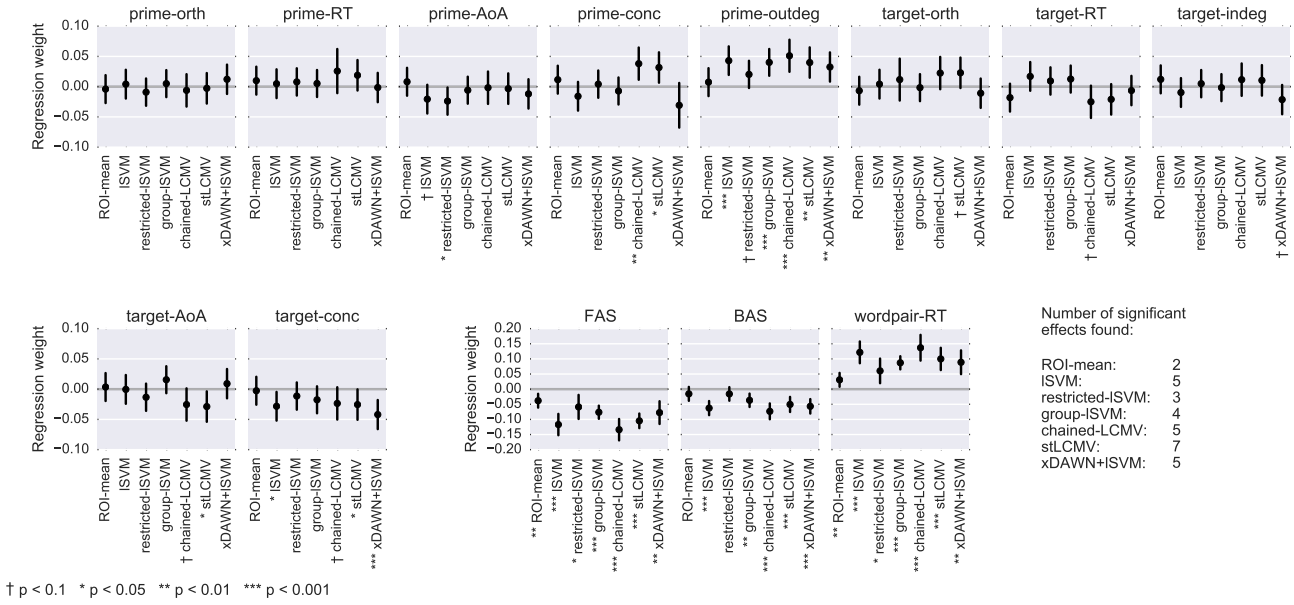


Figure 5: Univariate regression between each analysis method and each stimulus property. Dots indicate the regression weights obtained through the LME model and vertical lines indicate the 95% confidence interval. *P*-values are given for the null-hypothesis that the regression coefficient is zero. All variables were *z*-transformed before being entered into the model, hence regression weights can be interpreted as Pearson's correlation coefficients.

3.2 Real EEG data

The grand average ERPs (figure 4) show clear N2 and P2 components, associated with the response to a visual stimulus. The N400 components is modulated by the FAS of the word-pairs as expected, growing in amplitude as the FAS decreases, peaking around 400 ms and strongest at the central-parietal electrodes, which is consistent with the literature.[47]

[47] Marta Kutas and Federmeier, 2011

To test the performance of the multivariate filters in a more realistic setting, they were compared on a real EEG dataset, recorded in a semantic priming study.[48] A univariate regression study was performed using the amplitude of the N400, as quantified by the various analysis methods, as dependent variable and several stimulus properties (table 2), known to correlate with N400 amplitude, as independent variables (figure 5). To asses the performance of each method, we regarded the number of effects that were successfully found and the relative size of the regression weights. As all variables were *z*-transformed before entering them into the model, the regression weights can be interpreted like one would a Pearson's correlation coefficient. A comprehensive table of the exact regression weights, confidence intervals and *p*-values can be found in the supplementary information. Note that we have chosen not to correct the *p*-values for family-wise error rate. The purpose of this regression study is to compare the different analysis methods and due to the large number of tests, no significant effects would remain. Each stimulus property has already been shown in independent studies to correlate with the N400, so the occurrence of false-positives is unlikely. Also note that the various stimulus properties are correlated with each-other (figure 2), so the regression weights do not necessarily indicate the unique contribution of each stimulus property.

[48] van Vliet et al., 2014

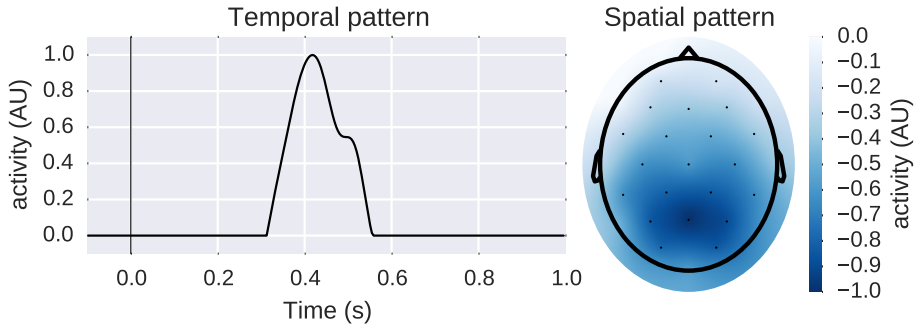


Figure 6: Temporal (left) and spatial (right) activation patterns used as template for the COI during beamformer analysis. In the visualization of the spatial pattern, dots represent the position of the electrodes and splines were used to interpolate the values between electrodes.

The spatial and temporal activation patterns used by the beamformer methods (created using the procedure described in section 2.5) are shown in figure 6.

The mean-ROI method only managed to find the two strongest effects on the N400: FAS and RT to the word-pair, but failed to find effects that strongly covariate with them (figure 2), namely BAS and the out-degree of the prime stimulus. Of the multivariate methods, only the restricted-lSVM failed to identify the latter effect. Where the chained-LCMV beamformer managed to find the effect of prime concreteness, the lSVM (with and without the xDAWN spatial filter) finds the effect of target concreteness instead. Finally, the stLCMV beamformer identified all of the effects the other methods found, as well as the effect of age-of-acquisition of the target word.

In terms of the size of the regression weights, the chained-LCMV beamformer performs best, followed by the lSVM, stLCMV beamformer, xDAWN+lSVM, restricted-lSVM, group-lSVM, and finally ROI-mean.

4 Discussion

The traditional method of measuring COI amplitude, namely taking the mean over a suitable ROI, is extremely sensitive to fluctuations of nearby (and not so nearby) ERP components (figure 3, bottom row). Averaging over many trials or subjects reduces these fluctuations and can vastly improve the result. However, when no such averaging is performed, it would be false to claim that limiting the analysis to a few selected electrodes and time points provides a measurement of the amplitude of the COI and the COI alone. Due to volume conduction, any electrode picks up signals from almost any part of the brain (figure 1, top) and the temporal dynamics of ERP components usually overlap as well (figure 1, bottom). The presence of other ERP components introduces a large amount of structured noise that becomes a problematically large confounding factor. When recordings are available from multiple electrodes over multiple time points, a linear multivariate filter can combine the signal from those electrodes and time points in order to actively counter interfering ERP components and other structured sources of noise.

The defining characteristic of the beamformer approach is the use of a template of the COI, which is both its biggest strength and weakness. The template allows a researcher to exert strict control over the signal that is isolated by the filter, which is a desirable property then the focus of the study is a specific, well defined COI. By verifying that the template of the COI is not tainted by other ERP components and noise sources such as eye artifacts, the filter can effectively counter such noise

sources. However, similar to the dependence of supervised learning techniques on the accuracy of the training labels, the performance of the beamformer is restricted by the method used to construct the template of the COI. For example, if the shape of the COI differs greatly between subjects and the method of estimating the template does not capture this (as is the case in the method we used), the beamformer will not perform optimally (figure 3, third column). Care must be taken to verify the validity of the template before drawing any conclusions about the output of the filter. For example, our method of constructing the template is only appropriate for isolating components that are known to have a relatively stable timing and scalp distribution, such as the N400 component. Note that the temporal pattern of the COI depends on the reference used in the recording. Therefore, the reference assumed by the template should match the reference used by the recording to which the beamformer filter is applied.

To be able to isolate components such as the N/P150, which exhibit polarity inversion,[49] other approaches must be explored, such as using the summation of two beamformers, constructed using separate templates for the N150 and P150. Further data-driven elements could be added to the process, such as shifting the template in time to optimally fit the ERP, although the more data-driven elements are added, the more the method will behave like a supervised learning approach. In order to distinguish between the behavior of the beamformer approach in general and the behavior of the particular method we used to estimate the template, simulation results are provided in the supplementary information for a beamformer that uses the true activation pattern of the COI as a template.

[49] Grainger and Holcomb, 2009

When we wish to study the effect of some experimental manipulation on a specific COI, the results favor the beamformer over the supervised learning approaches. In this case, the specificity criterion mentioned in section 1.2 comes into play and it is not only important to have a good representation of the COI, but also to properly counter the influence of other ERP components. The simulation study shows that while the chained-LCMV beamformer is more robust against unstructured noise, the stLCMV beamformer was better able to counteract structured noise, which is more important in this case.

5 Conclusion

We have demonstrated the need for multivariate analysis, when ERP component amplitude measurements are desired on a single-trial basis.

The LCMV beamformer can be expanded from its traditional formulation as a spatial filter, to be a spatio-temporal filter which workings can be easily understood. The behavior of the beamformer, along with various supervised learning approaches, has been evaluated on simulated and real EEG data. Supervised learning approaches, such as the lSVM, extract any information that aids in reproducing the labels of a training set, regardless of whether this information comes from the ERP component the researcher wishes to study (the COI), or from some other ERP component or from structured noise sources, such as EOG artifacts.

In contrast, the beamformer filter only isolates signals that conform to a given

template. By crafting a template for the ERP component, which amplitude the researches wishes to measure, the beamformer effectively suppresses structured noise sources, such as other ERP components, eye-artifacts, and so forth. A simple method of crafting such a template, based on the grand average ERP of multiple subjects, is demonstrated to perform well. Of the two approaches used to extend the LCMV beamformer to the spatio-temporal domain, the stLCMV approach scored best during the evaluation. During the simulation study, the stLCMV beamformer output was mostly unaffected by the presence of interfering ERP components under all tested circumstances.

This means that if a researcher can construct a template that is a good approximation of the shape of the COI, the stLCMV beamformer output is mostly attributable to changes in its amplitude and is mostly unaffected by structured noise, such as other ERP components.

Supplementary information

Complete simulation study

The complete list of methods of estimating COI amplitude that were evaluated is:

ROI-mean Traditional method of taking the mean voltage over a few selected channels and time points (channels FC2, FC5, Cz, C3, C4, CP5, CP6, time window 0.3 s to 0.5 s).

perfect-chained-LCMV A chained-LCMV beamformer filter, which uses a separate spatial and temporal beamformer. The spatial and temporal templates were set to the true shape of the COI. Estimation of the covariance matrix was performed using OAS.[50]

[50] Chen et al., 2010

chained-LCMV A chained-LCMV beamformer filter, which uses a separate spatial and temporal beamformer. The spatial and temporal templates were constructed using the procedure outlined in section 2.5. Estimation of the covariance matrix was performed using OAS.[51]

[51] Chen et al., 2010

perfect-stLCMV An stLCMV beamformer filter, where the template of the COI was set to the true shape of the COI. Estimation of the covariance matrix was performed using OAS.[52]

[52] Chen et al., 2010

stLCMV An stLCMV beamformer filter, where the template of the COI was constructed using the procedure outlined in section 2.5. Estimation of the covariance matrix was performed using OAS.[53]

[53] Chen et al., 2010

ss-stLCMV An stLCMV beamformer filter, where the template of the COI was constructed using the procedure outlined in section 2.5, but limited to the training data of a single recording (the one currently being processed). Estimation of the covariance matrix was performed using OAS.[54]

[54] Chen et al., 2010

lSVM A linear support vector machine.

shrinkage-LDA Linear discriminant analysis, with shrinkage of the covariance matrix. Estimation of the shrunk covariance matrix was performed using OAS.[55]

[55] Chen et al., 2010

LR Logistic regression. Estimation for the amount of $\ell 2$ regularization was performed using OAS.[56]

[56] Chen et al., 2010

restricted-lSVM An lSVM that does not use all the features. Instead, it only uses a few selected channels and time points (channels FC2, FC5, Cz, C3, C4, CP5, CP6, time window 0.3 s to 0.5 s).

group-lSVM An lSVM that was trained on the pooled training data of all 10 subjects.

xDAWN+lSVM An xDAWN spatial filter,[57] combined with an lSVM. The 4 most descriptive spatial components computed by the xDAWN algorithm were retained. Estimation of the covariance matrix was performed using OAS .[58]

[57] Rivet et al., 2009

[58] Chen et al., 2010

xDAWN+Riemann An xDAWN spatial filter,[59] combined with classification of the covariance matrix on a Riemannian manifold.[60] The 4 most descriptive spatial components computed by the xDAWN algorithm were retained. Estimation

[59] Rivet et al., 2009

[60] Barachant and Congedo, 2014

of the covariance matrix was performed using OAS .[61] This is based on the approach that won the Kaggle DecMeg 2014 competition (https://www.kaggle.com/c/decoding-the-human-brain).

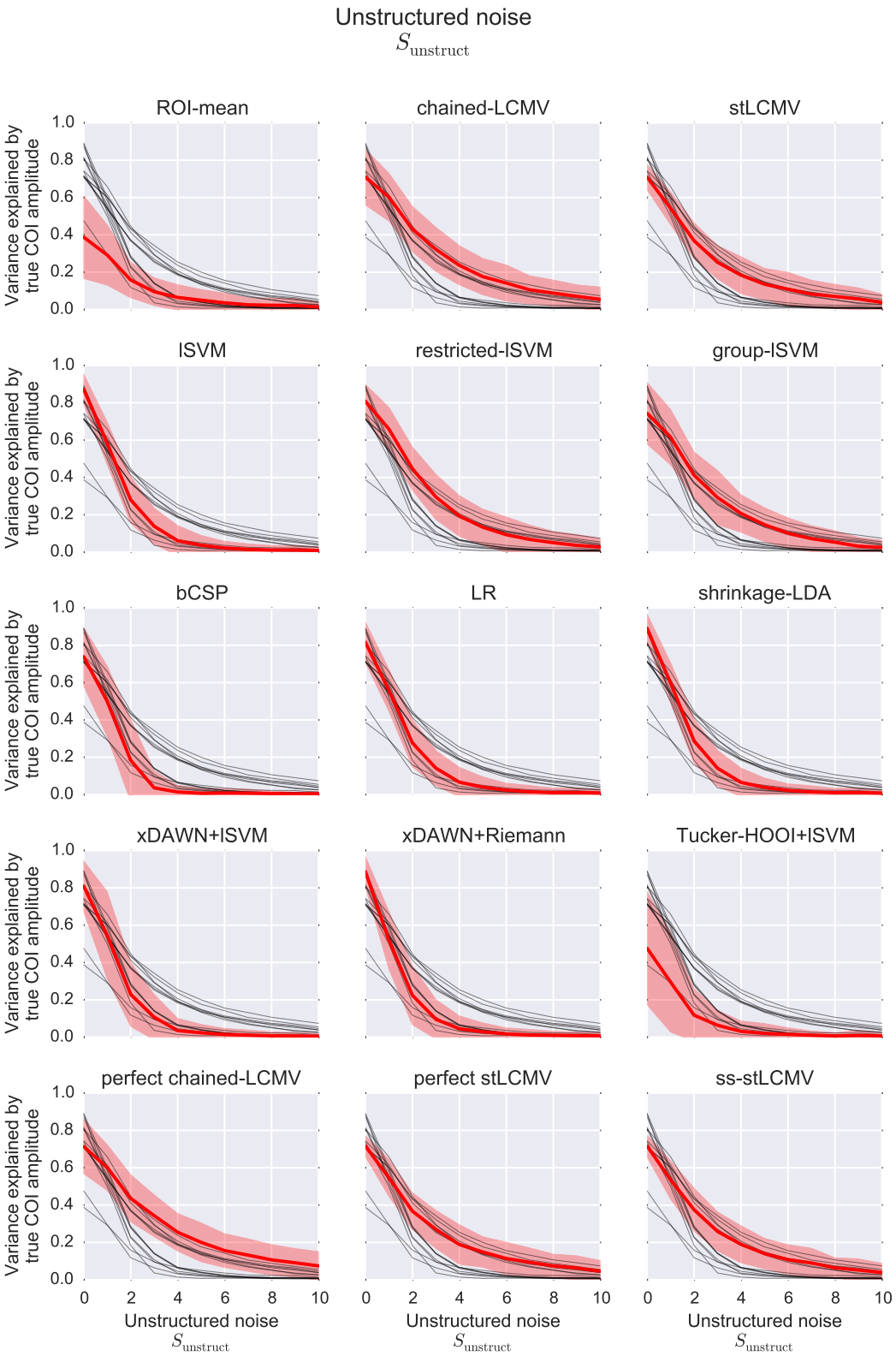
[61] Chen et al., 2010

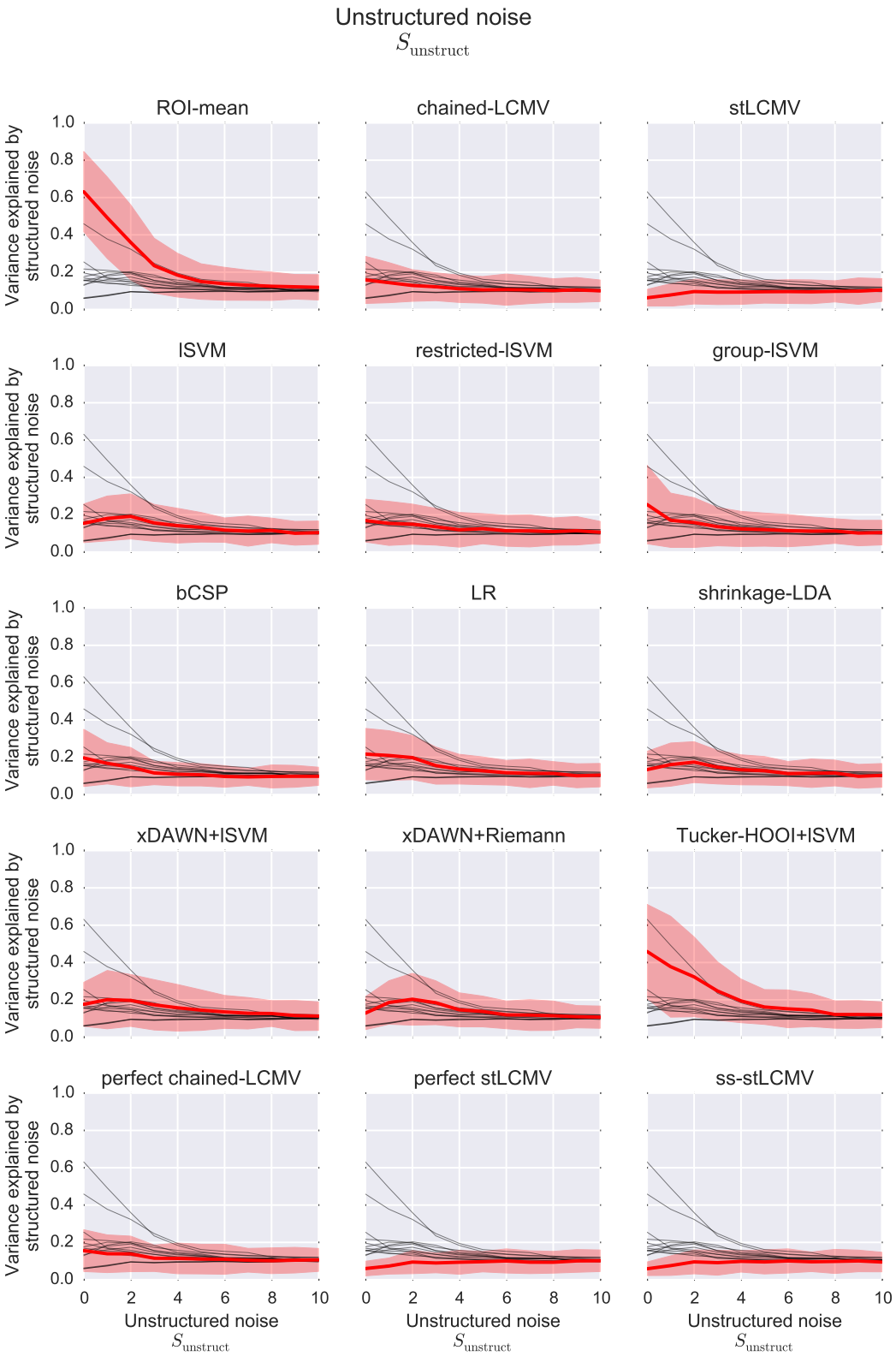
bCSP The bi-linear common spatial patterns algorithm.[62]

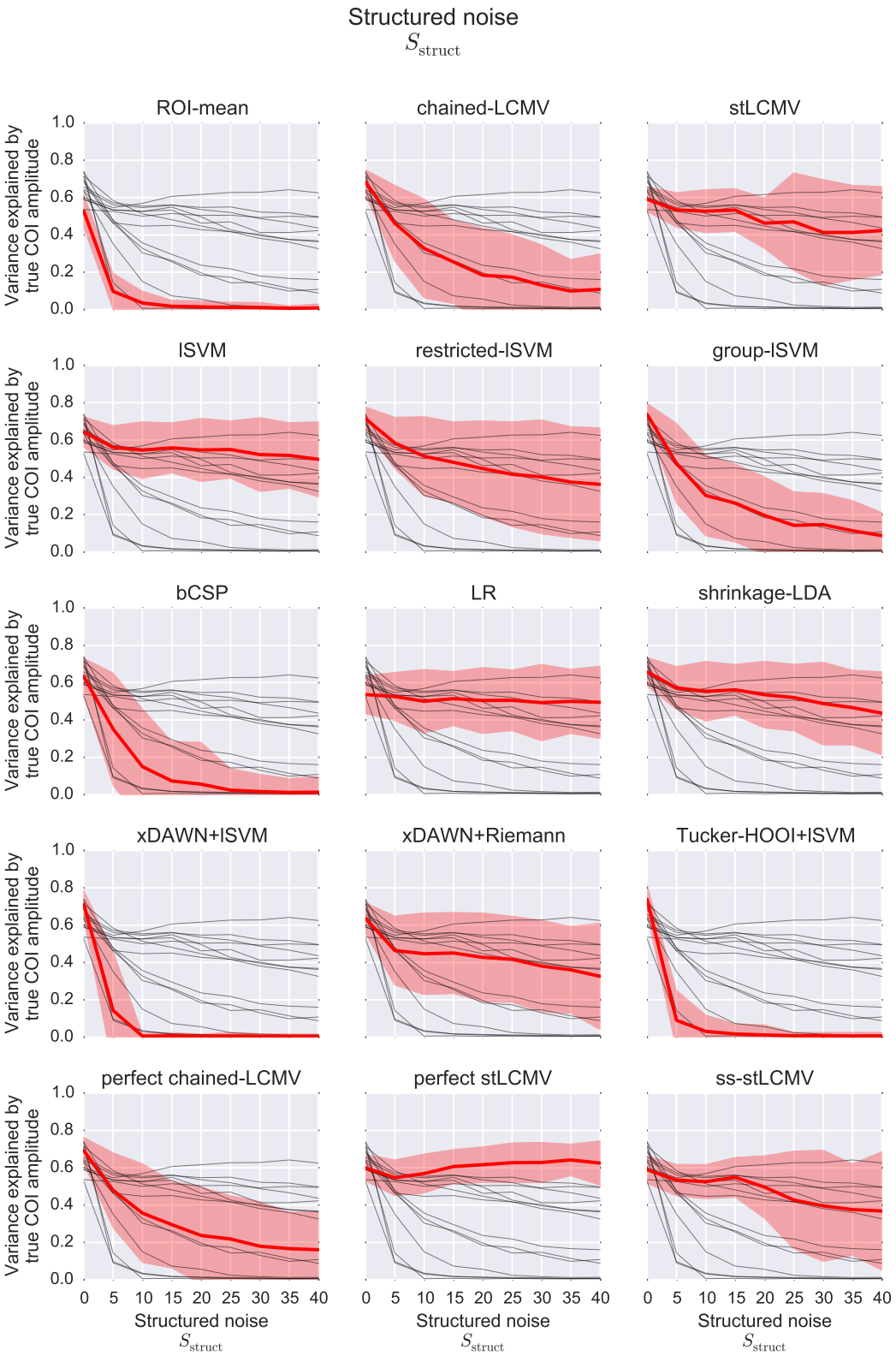
[62] Yu et al., 2012

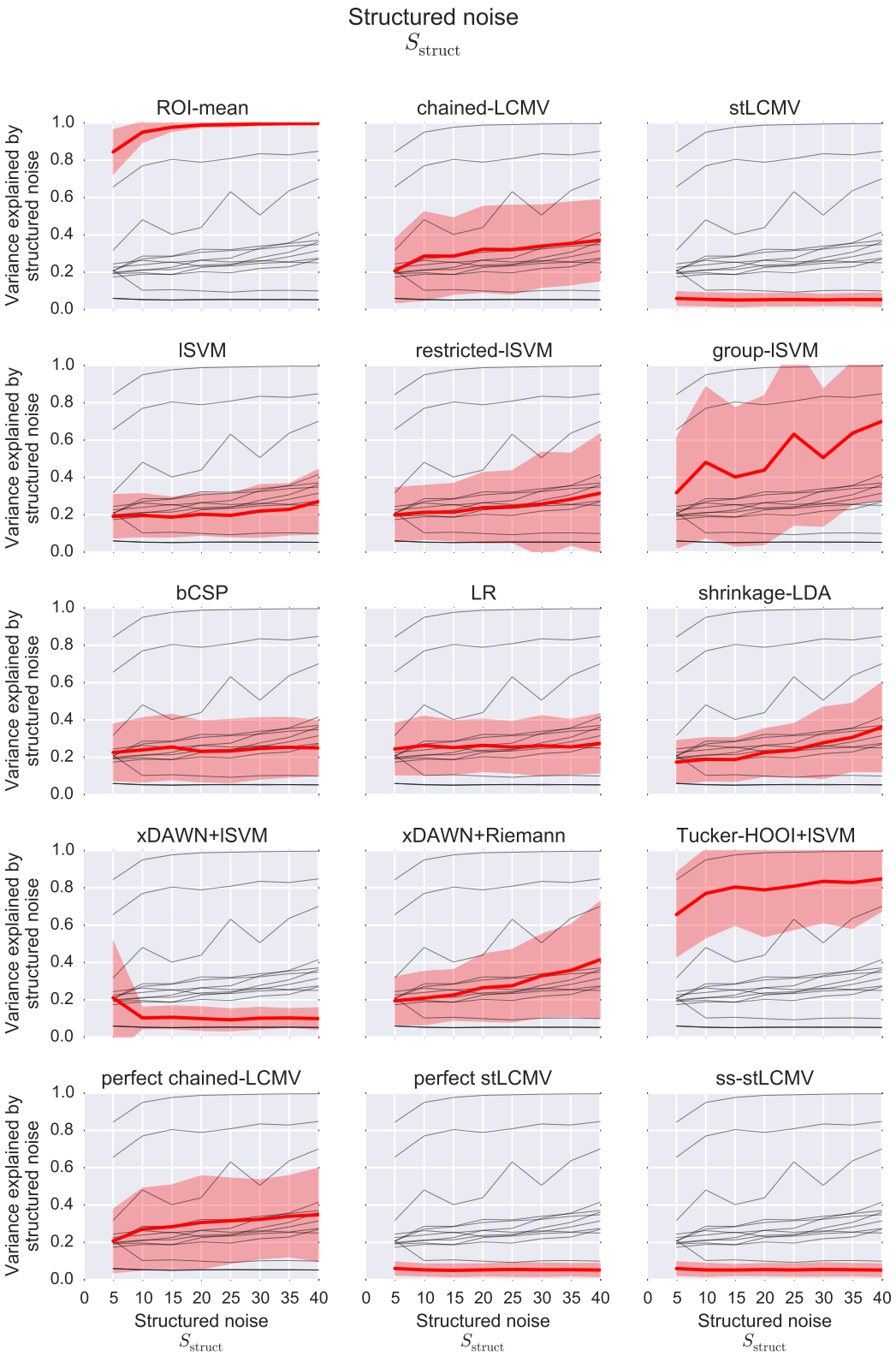
Tucker-HOOI+lSVM A combination of unsupervised tensor decomposition, using the Tucker-higher-order orthogonal iteration (HOOI) algorithm, and an lSVM. The tensor decomposition was used to map for each trial, the original (#channels × #samples ×) tensor onto a (8×10) core tensor, which was used as features for the lSVM.

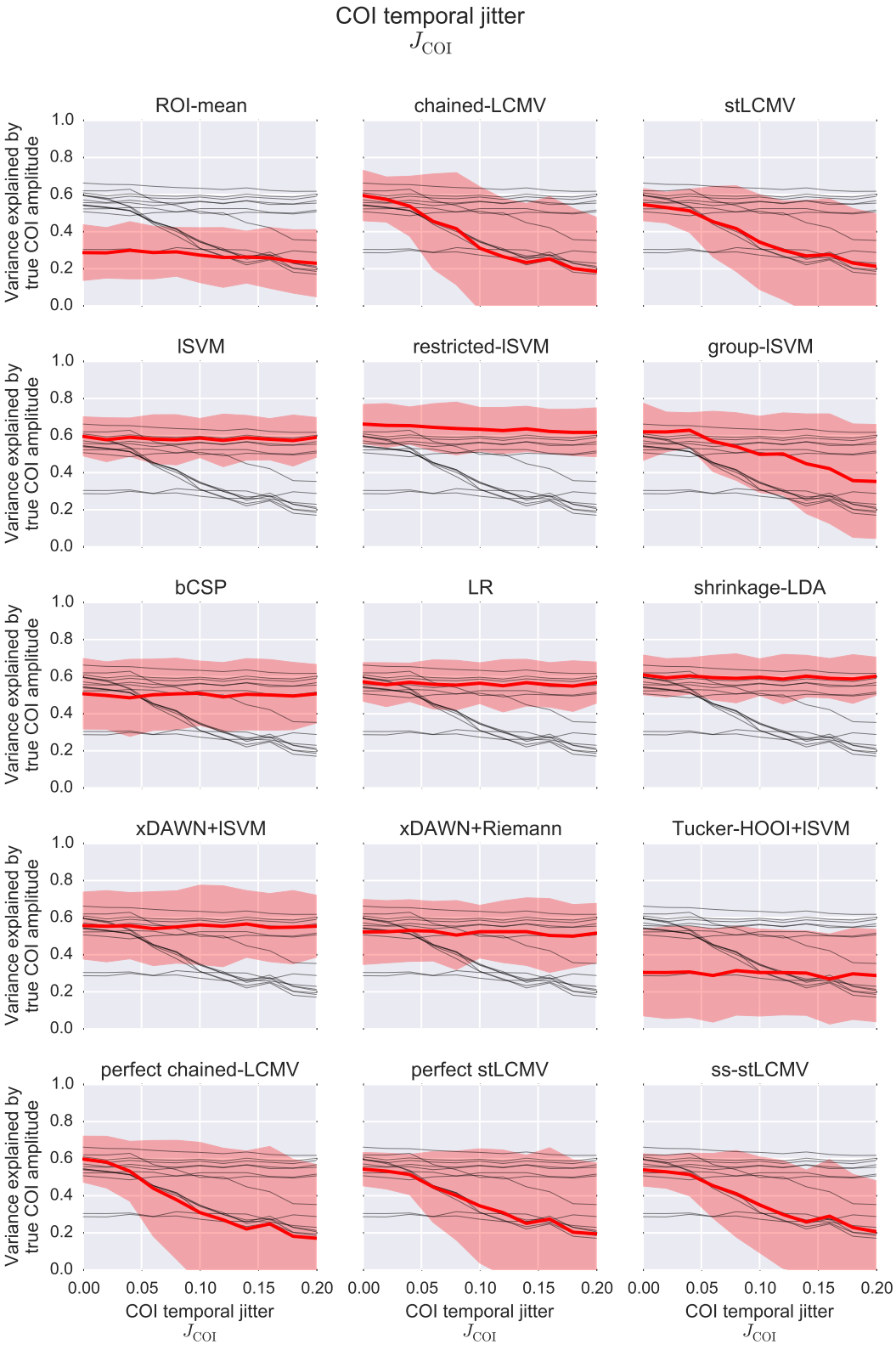
The following graphs show the performance of above techniques to estimate COI amplitude under various simulated conditions. The performance is quantified by two metrics: the variance of the output explained (R^2 stat) by the true amplitude of theCOI (higher is better) and the variance explained by least squares regression with the structured noise sources (lower is better). Each curve is the average of 100 runs. The 95% confidence interval is plotted as a shaded area. In each graph, the curve corresponding to a particular method is highlighted in red, with the curves corresponding to the other methods shown as thin black lines for comparison.

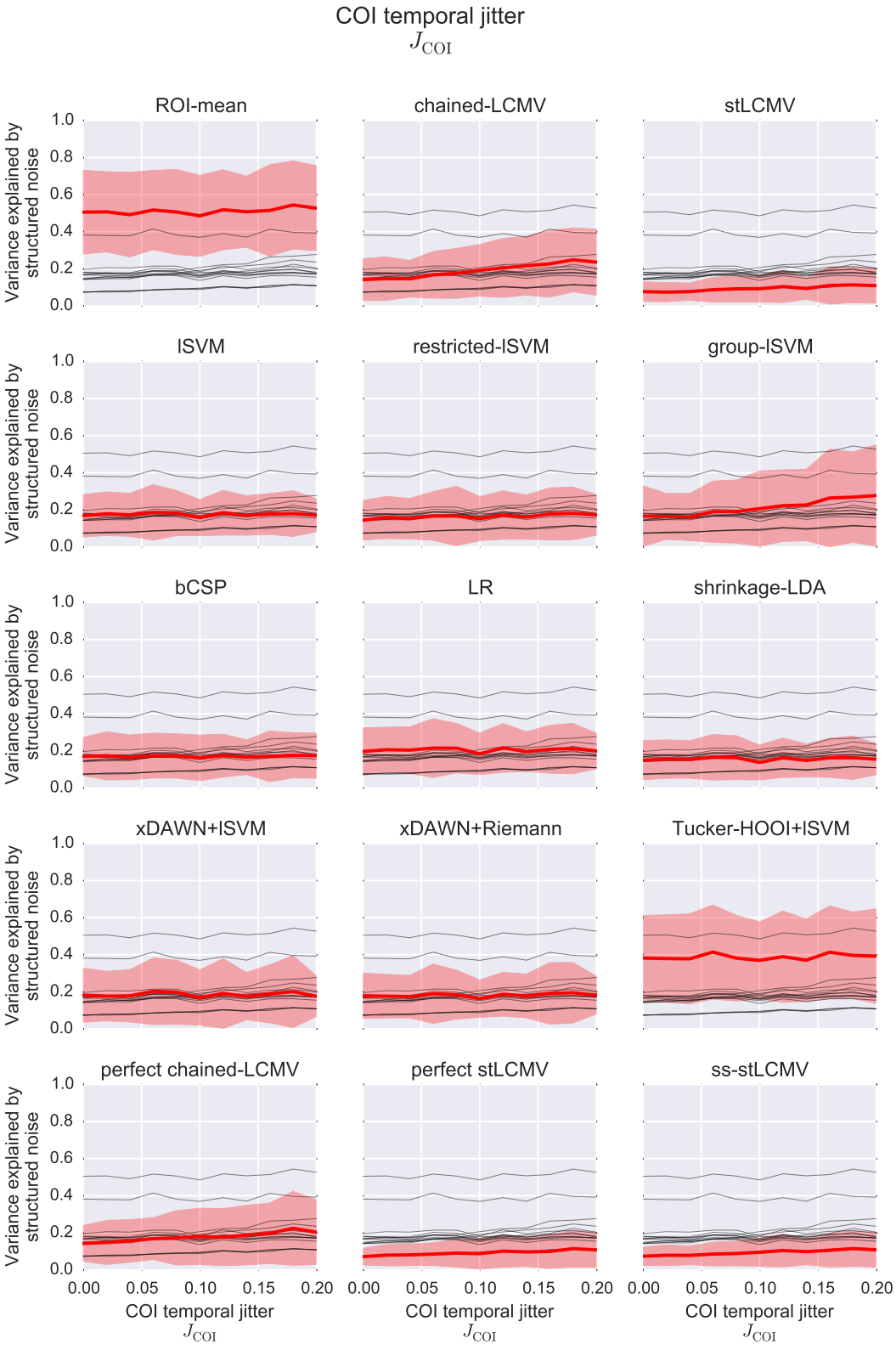

Unstructured noise
S_{unstruct}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by true COI amplitude
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Unstructured noise
S_{unstruct}

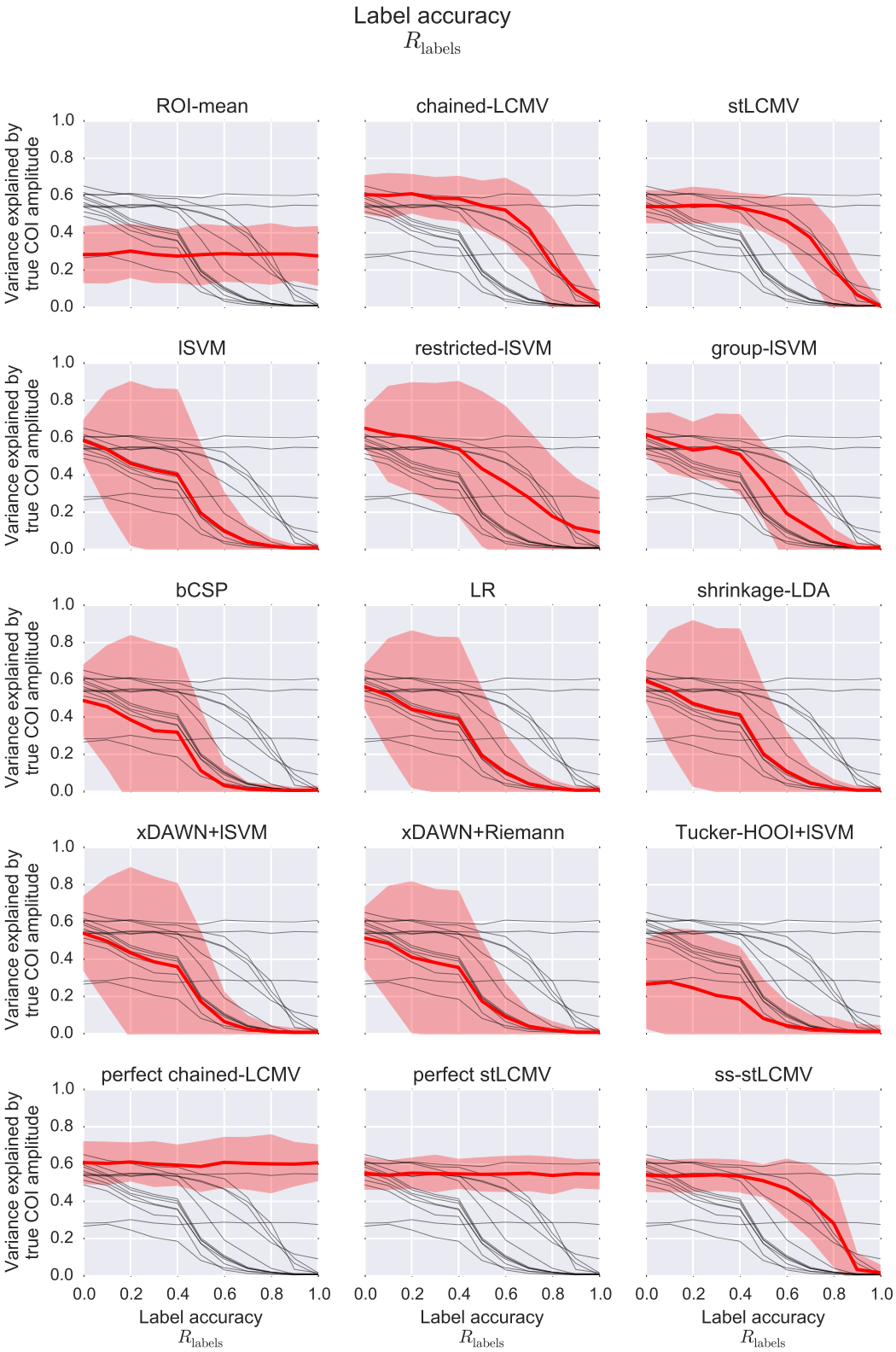

Unstructured noise
S_{unstruct}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by structured noise
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Unstructured noise
S_{unstruct}

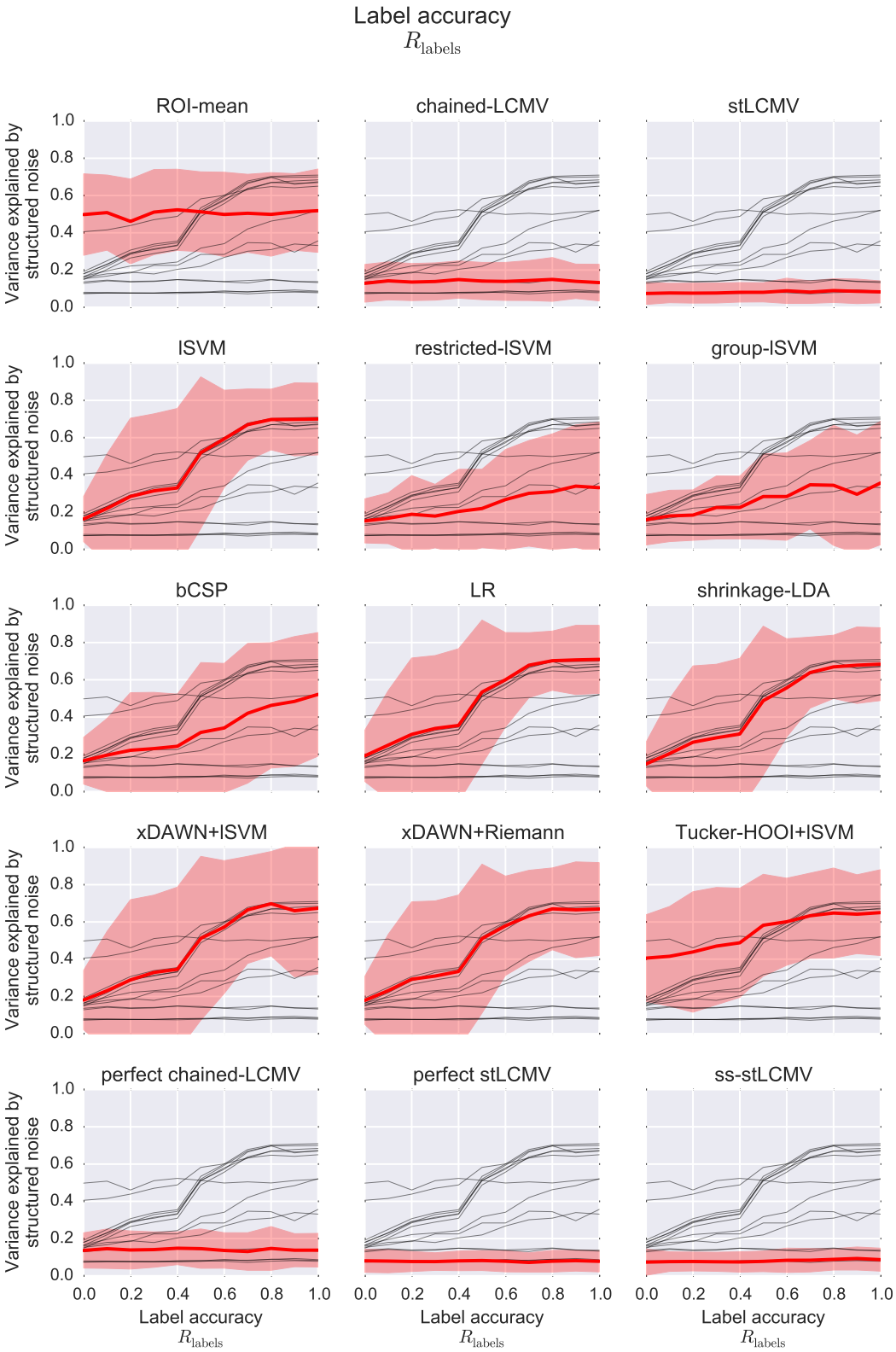

Structured noise
S_{struct}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by true COI amplitude
0.0
0.2
0.4
0.6
0.8
1.0
0 5 10 15 20 25 30 35 40
Structured noise
S_{struct}


Structured noise
S_{struct}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by structured noise
0.0
0.2
0.4
0.6
0.8
1.0
0 5 10 15 20 25 30 35 40
Structured noise
S_{struct}


COI temporal jitter
J_{COI}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by true COI amplitude
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
COI temporal jitter
J_{COI}


COI temporal jitter
J_{COI}
ROI-mean
chained-LCMV
stLCMV
ISVM
restricted-ISVM
group-ISVM
bCSP
LR
shrinkage-LDA
xDAWN+ISVM
xDAWN+Riemann
Tucker-HOOI+ISVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by structured noise
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
COI temporal jitter
J_{COI}

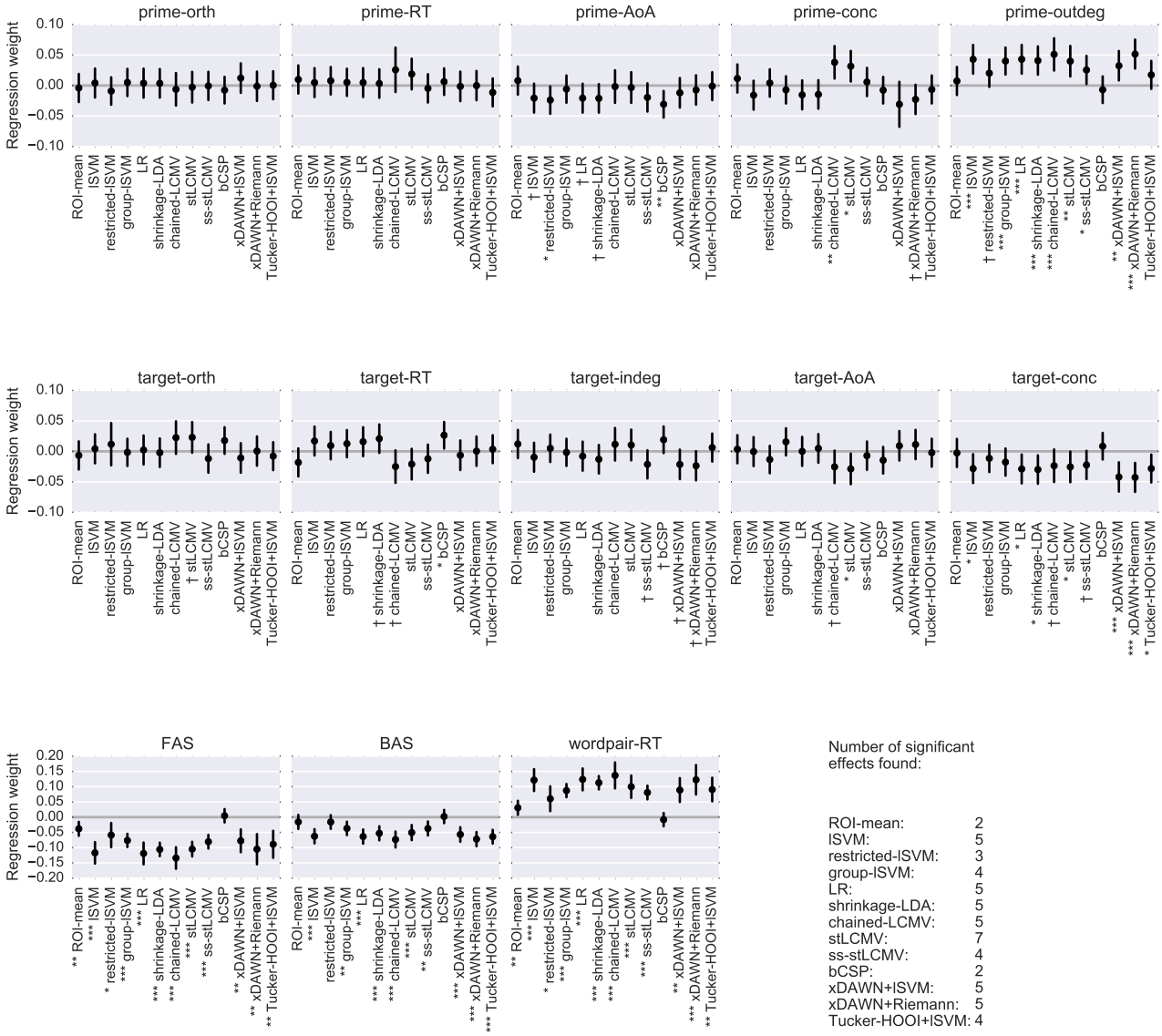

Label accuracy
R_{labels}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by true COI amplitude
0.0
0.2
0.4
0.6
0.8
1.0
Label accuracy
R_{labels}


Label accuracy
R_{labels}
ROI-mean
chained-LCMV
stLCMV
lSVM
restricted-lSVM
group-lSVM
bCSP
LR
shrinkage-LDA
xDAWN+lSVM
xDAWN+Riemann
Tucker-HOOI+lSVM
perfect chained-LCMV
perfect stLCMV
ss-stLCMV
Variance explained by structured noise
0.0
0.2
0.4
0.6
0.8
1.0
Label accuracy
R_{labels}

Evaluation on real EEG data

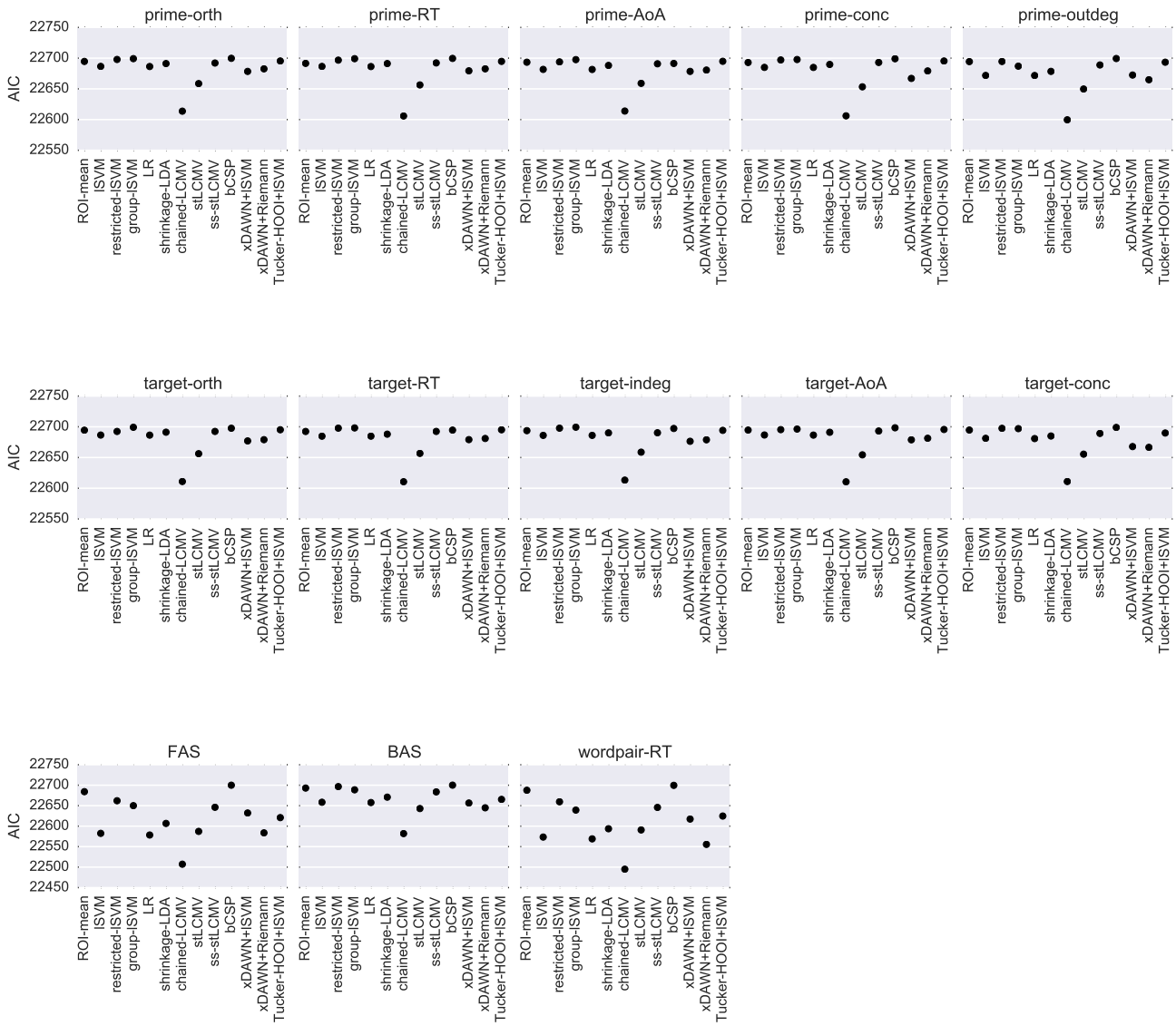
Univariate regression between each analysis method and each stimulus property. Dots indicate the regression weights obtained through the LME model and vertical lines indicate the 95% confidence interval. *P*-values are given for the null-hypothesis that the regression coefficient is zero. All variables were *z*-transformed before being entered into the

model, hence regression weights can be interpreted as Pearson's correlation coefficients.



† p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001

Akaike information criterion (AIC) measurements for the regression models used in the previous figure.



The following tables contain the raw values obtained during the regression study on real EEG data. Given are the regression coefficients (coeff), confidence interval of these coefficients (low-hight), t-values, p-values, and the AIC.

N400 amplitude estimation through ROI-mean

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.004040	-0.027168	0.019088	-0.342396	0.73214	22694.635799
prime-RT	0.009939	-0.013180	0.033058	0.842584	0.39971	22691.466089
prime-AoA	0.008094	-0.015029	0.031217	0.686067	0.49287	22693.355484
prime-conc	0.011547	-0.011568	0.034663	0.979078	0.32784	22692.909825
prime-outdeg	0.007461	-0.015663	0.030584	0.632353	0.52734	22694.353255
target-orth	-0.006705	-0.029830	0.016420	-0.568268	0.57001	22694.430162
target-RT	-0.018138	-0.041234	0.004957	-1.539292	0.12413	22692.387113
target-indeg	0.012063	-0.011052	0.035177	1.022847	0.30669	22693.707495
target-AoA	0.003501	-0.019627	0.026629	0.296705	0.76677	22694.664996
target-conc	-0.002645	-0.025774	0.020484	-0.224133	0.82271	22694.701653
FAS	-0.038215	-0.061192	-0.015237	-3.259705	0.00116	22684.197385
BAS	-0.015554	-0.038659	0.007550	-1.319473	0.18739	22693.013910
wordpair-RT	0.030776	0.007745	0.053807	2.619115	0.00898	22687.635866

N400 amplitude estimation through lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.004132	-0.019706	0.027970	0.339735	0.73415	22686.536836
prime-RT	0.004865	-0.018972	0.028702	0.400012	0.68925	22686.492254
prime-AoA	-0.020635	-0.044432	0.003162	-1.699565	0.0896	22681.683334
prime-conc	-0.015865	-0.039680	0.007949	-1.305746	0.19201	22684.949092
prime-outdeg	0.042976	0.019324	0.066629	3.561209	0.000390993	22671.824437
target-orth	0.004167	-0.019671	0.028005	0.342628	0.73197	22686.534863
target-RT	0.016951	-0.006859	0.040762	1.395335	0.1633	22684.707656
target-indeg	-0.009654	-0.033484	0.014176	-0.794003	0.42743	22686.022056
target-AoA	-0.000406	-0.024246	0.023434	-0.033377	0.97338	22686.651134
target-conc	-0.028301	-0.052060	-0.004542	-2.334664	0.01981	22681.220098
FAS	-0.117075	-0.152427	-0.081723	-6.490809	8.93444×10^{-5}	22582.440434
BAS	-0.062610	-0.086051	-0.039169	-5.234984	2.11096×10^{-7}	22658.404367
wordpair-RT	0.121639	0.085557	0.157722	6.607303	8.13534×10^{-5}	22573.382098

N400 amplitude estimation through restricted-lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.009039	-0.031624	0.013546	-0.784387	0.43305	22697.848354
prime-RT	0.007834	-0.014753	0.030421	0.679804	0.49683	22696.755648
prime-AoA	-0.023880	-0.046413	-0.001347	-2.077116	0.03811	22693.882183
prime-conc	0.004155	-0.018437	0.026747	0.360448	0.71861	22697.057827
prime-outdeg	0.020282	-0.002268	0.042832	1.762850	0.07831	22694.551178
target-orth	0.011686	-0.023001	0.046373	0.660311	0.52473	22692.426463
target-RT	0.009407	-0.013177	0.031992	0.816417	0.41451	22697.797122
target-indeg	0.004980	-0.017611	0.027571	0.432092	0.66579	22697.742386
target-AoA	-0.013379	-0.035954	0.009195	-1.161605	0.24574	22695.360678
target-conc	-0.011426	-0.034006	0.011154	-0.991778	0.32161	22697.480361
FAS	-0.058786	-0.098656	-0.018915	-2.889806	0.0175	22662.072540
BAS	-0.015827	-0.038394	0.006740	-1.374601	0.16964	22696.576084
wordpair-RT	0.060365	0.019553	0.101177	2.898960	0.01727	22659.583732

N400 amplitude estimation through group-lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.005026	-0.017347	0.027399	0.440312	0.65983	22699.181945
prime-RT	0.005237	-0.017136	0.027610	0.458803	0.6465	22699.116376
prime-AoA	-0.005855	-0.028227	0.016517	-0.512953	0.60813	22697.583772
prime-conc	-0.007159	-0.029529	0.015211	-0.627224	0.53069	22697.748803
prime-outdeg	0.040133	0.017931	0.062335	3.542895	0.00041872	22686.921270
target-orth	-0.001627	-0.024002	0.020749	-0.142480	0.88674	22699.199613
target-RT	0.012554	-0.009805	0.034913	1.100496	0.27145	22698.165623
target-indeg	-0.001631	-0.024007	0.020744	-0.142885	0.88642	22699.337064
target-AoA	0.015583	-0.006766	0.037933	1.366597	0.17214	22696.285612
target-conc	-0.017602	-0.039944	0.004741	-1.544086	0.12296	22696.995146
FAS	-0.076316	-0.098167	-0.054466	-6.845464	8.1859×10^{-12}	22650.034361
BAS	-0.036827	-0.059057	-0.014597	-3.247006	0.00121	22688.729961
wordpair-RT	0.087092	0.065260	0.108923	7.818924	5.9952×10^{-15}	22639.256399

N400 amplitude estimation through LR

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.003946	-0.019906	0.027798	0.324227	0.74585	22686.354425
prime-RT	0.004862	-0.018989	0.028713	0.399547	0.6896	22686.299919
prime-AoA	-0.020730	-0.044540	0.003080	-1.706437	0.08832	22681.570630
prime-conc	-0.015333	-0.039162	0.008497	-1.261078	0.20765	22684.870805
prime-outdeg	0.043071	0.019405	0.066737	3.567061	0.000382502	22671.719503
target-orth	0.002228	-0.021625	0.026081	0.183069	0.85479	22686.426027
target-RT	0.015814	-0.008014	0.039642	1.300778	0.19371	22684.769306
target-indeg	-0.007924	-0.031772	0.015923	-0.651301	0.51504	22686.035460
target-AoA	-0.000009	-0.023862	0.023845	-0.000713	0.99943	22686.459541
target-conc	-0.028891	-0.052660	-0.005122	-2.382303	0.01744	22680.804234
FAS	-0.119048	-0.154880	-0.083215	-6.511703	8.85733×10^{-5}	22578.378878
BAS	-0.063613	-0.087056	-0.040171	-5.318630	1.35836×10^{-7}	22657.666512
wordpair-RT	0.123920	0.087866	0.159974	6.736539	7.08355×10^{-5}	22568.987026

N400 amplitude estimation through shrinkage-LDA

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.003632	-0.019844	0.027109	0.303241	0.76179	22691.136628
prime-RT	0.003284	-0.020193	0.026761	0.274141	0.78405	22691.153547
prime-AoA	-0.021084	-0.044517	0.002348	-1.763547	0.07819	22688.124636
prime-conc	-0.014632	-0.038088	0.008824	-1.222655	0.22182	22689.733660
prime-outdeg	0.040924	0.017618	0.064230	3.441589	0.000608349	22678.435013
target-orth	-0.002185	-0.025662	0.021293	-0.182390	0.85532	22691.195431
target-RT	0.020786	-0.002648	0.044219	1.738484	0.08251	22687.984372
target-indeg	-0.012939	-0.036400	0.010522	-1.080946	0.28005	22690.061106
target-AoA	0.004831	-0.018644	0.028307	0.403352	0.6868	22691.066020
target-conc	-0.029850	-0.053236	-0.006463	-2.501623	0.01256	22684.994958
FAS	-0.105816	-0.128118	-0.083514	-9.299325	0	22606.672059
BAS	-0.052891	-0.076081	-0.029701	-4.470262	8.94185×10^{-6}	22670.815820
wordpair-RT	0.112956	0.090823	0.135089	10.002662	0	22593.667117

N400 amplitude estimation through chained-LCMV

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.006158	-0.032971	0.020654	-0.450170	0.65271	22613.682123
prime-RT	0.025763	-0.010770	0.062296	1.382155	0.18875	22605.842816
prime-AoA	-0.001758	-0.028574	0.025058	-0.128513	0.89778	22613.877213
prime-conc	0.038022	0.011336	0.064709	2.792519	0.00535	22606.083472
prime-outdeg	0.051086	0.024505	0.077667	3.766800	0.000177495	22599.625823
target-orth	0.022466	-0.004305	0.049237	1.644807	0.1004	22610.953388
target-RT	-0.025026	-0.051786	0.001734	-1.832972	0.06718	22610.576280
target-indeg	0.011587	-0.015217	0.038392	0.847285	0.39709	22613.211453
target-AoA	-0.025434	-0.052192	0.001325	-1.862948	0.06284	22610.460809
target-conc	-0.023575	-0.050342	0.003191	-1.726286	0.08468	22610.954501
FAS	-0.133987	-0.169488	-0.098486	-7.397234	7.56465×10^{-6}	22507.132270
BAS	-0.073358	-0.099687	-0.047029	-5.460765	6.33312×10^{-8}	22581.880911
wordpair-RT	0.136897	0.094242	0.179553	6.290252	5.86531×10^{-5}	22494.922880

N400 amplitude estimation through stLCMV

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.002728	-0.028002	0.022546	-0.211541	0.83252	22658.543987
prime-RT	0.018824	-0.006417	0.044065	1.461664	0.14423	22656.384533
prime-AoA	-0.003273	-0.028547	0.022000	-0.253844	0.79968	22658.971728
prime-conc	0.031654	0.006475	0.056834	2.463963	0.01395	22653.304790
prime-outdeg	0.039872	0.014749	0.064995	3.110584	0.00193	22649.735339
target-orth	0.022895	-0.002330	0.048120	1.778916	0.07563	22656.194661
target-RT	-0.020971	-0.046204	0.004262	-1.628890	0.10373	22656.704066
target-indeg	0.010416	-0.014849	0.035680	0.808046	0.4193	22658.700280
target-AoA	-0.028793	-0.053989	-0.003598	-2.239816	0.02538	22654.351862
target-conc	-0.025503	-0.050716	-0.000290	-1.982528	0.04776	22655.430540
FAS	-0.104778	-0.128986	-0.080570	-8.483133	0	22587.230850
BAS	-0.050568	-0.075598	-0.025537	-3.959599	8.17716×10^{-5}	22642.986334
wordpair-RT	0.099884	0.062942	0.136827	5.299295	0.000256578	22590.774994

N400 amplitude estimation through ss-stLCMV

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.000663	-0.023926	0.022599	-0.055880	0.95545	22692.034744
prime-RT	-0.004635	-0.027896	0.018626	-0.390552	0.69623	22692.210717
prime-AoA	-0.019508	-0.042732	0.003715	-1.646425	0.10007	22690.802730
prime-conc	0.005807	-0.017452	0.029067	0.489370	0.62471	22692.971923
prime-outdeg	0.025509	0.002313	0.048704	2.155442	0.03143	22688.876381
target-orth	-0.011864	-0.035112	0.011384	-1.000221	0.31751	22692.442320
target-RT	-0.012153	-0.035400	0.011095	-1.024591	0.30587	22692.459760
target-indeg	-0.021352	-0.044568	0.001863	-1.802645	0.07182	22690.265920
target-AoA	-0.007080	-0.030338	0.016177	-0.596686	0.55089	22693.152902
target-conc	-0.022295	-0.045506	0.000916	-1.882601	0.06012	22689.194723
FAS	-0.080526	-0.103108	-0.057943	-6.988844	5.84865×10^{-12}	22646.099686
BAS	-0.037115	-0.060235	-0.013995	-3.146358	0.00171	22683.670117
wordpair-RT	0.080815	0.058237	0.103393	7.015503	4.88853×10^{-12}	22645.747919

N400 amplitude estimation through bCSP

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.007585	-0.029499	0.014328	-0.678431	0.49752	22699.699278
prime-RT	0.006358	-0.015556	0.028272	0.568646	0.56961	22699.555197
prime-AoA	-0.030699	-0.052603	-0.008795	-2.746894	0.00603	22691.344031
prime-conc	-0.007626	-0.029539	0.014288	-0.682031	0.49524	22699.010468
prime-outdeg	-0.006883	-0.028797	0.015031	-0.615606	0.53817	22699.358077
target-orth	0.017754	-0.004157	0.039665	1.588113	0.1123	22697.637828
target-RT	0.026399	0.004492	0.048305	2.361833	0.01821	22694.583221
target-indeg	0.018994	-0.002916	0.040905	1.699089	0.08934	22697.273149
target-AoA	-0.014552	-0.036464	0.007360	-1.301603	0.19309	22698.378897
target-conc	0.008427	-0.013487	0.030340	0.753678	0.45106	22698.923887
FAS	0.004726	-0.017188	0.026641	0.422720	0.67251	22699.980843
BAS	0.002099	-0.019815	0.024013	0.187734	0.85109	22700.124289
wordpair-RT	-0.008095	-0.030008	0.013819	-0.723988	0.46909	22699.618566

N400 amplitude estimation through xDAWN+lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.012187	-0.012104	0.036479	0.983325	0.32575	22678.403599
prime-RT	-0.001437	-0.025743	0.022870	-0.115850	0.9078	22679.356521
prime-AoA	-0.011940	-0.036233	0.012352	-0.963364	0.33566	22678.412169
prime-conc	-0.030780	-0.067748	0.006189	-1.631865	0.13156	22667.003945
prime-outdeg	0.032481	0.008279	0.056683	2.630406	0.00869	22672.480694
target-orth	-0.010846	-0.035141	0.013449	-0.874997	0.38184	22676.994454
target-RT	-0.006418	-0.030720	0.017884	-0.517607	0.60488	22679.102071
target-indeg	-0.021390	-0.045651	0.002871	-1.727991	0.08438	22676.389555
target-AoA	0.009068	-0.015230	0.033366	0.731442	0.46472	22678.834819
target-conc	-0.042019	-0.066150	-0.017888	-3.412821	0.000675295	22667.802775
FAS	-0.077537	-0.115237	-0.039836	-4.030950	0.00225	22632.140305
BAS	-0.056740	-0.080726	-0.032754	-4.636414	4.13978×10^{-6}	22656.491877
wordpair-RT	0.088882	0.049475	0.128289	4.420712	0.00127	22617.324799

N400 amplitude estimation through xDAWN+Riemann

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	-0.001219	-0.025330	0.022892	-0.099084	0.9211	22682.589326
prime-RT	-0.000110	-0.024222	0.024001	-0.008954	0.99286	22682.599063
prime-AoA	-0.007310	-0.031416	0.016796	-0.594325	0.55246	22680.589373
prime-conc	-0.022737	-0.046797	0.001323	-1.852221	0.06436	22679.175766
prime-outdeg	0.051539	0.027694	0.075384	4.236272	2.53731×10^{-5}	22664.851713
target-orth	0.000465	-0.023646	0.024577	0.037810	0.96985	22679.022175
target-RT	0.000300	-0.023811	0.024412	0.024396	0.98054	22680.985276
target-indeg	-0.023649	-0.047705	0.000407	-1.926840	0.05435	22678.849632
target-AoA	0.011183	-0.012916	0.035282	0.909535	0.36334	22681.322913
target-conc	-0.042766	-0.066695	-0.018837	-3.502924	0.000485746	22666.534226
FAS	-0.105084	-0.154679	-0.055490	-4.152887	0.00222	22583.658832
BAS	-0.072020	-0.095609	-0.048431	-5.984080	3.2833×10^{-9}	22644.821789
wordpair-RT	0.122593	0.074117	0.171069	4.956608	0.000704892	22555.531296

N400 amplitude estimation through Tucker-HOOI+lSVM

	coeff	conf-low	conf-high	t	p	AIC
feature						
prime-orth	0.000494	-0.022531	0.023520	0.042091	0.96644	22695.366497
prime-RT	-0.011288	-0.034300	0.011724	-0.961424	0.33663	22694.715256
prime-AoA	-0.001047	-0.024072	0.021978	-0.089110	0.92902	22694.853662
prime-conc	-0.006462	-0.029483	0.016559	-0.550165	0.58236	22695.336435
prime-outdeg	0.017462	-0.005531	0.040456	1.488510	0.13701	22693.367150
target-orth	-0.007909	-0.030927	0.015110	-0.673400	0.50089	22695.185720
target-RT	0.003484	-0.019540	0.026508	0.296572	0.76687	22695.136349
target-indeg	0.006201	-0.016820	0.029222	0.527966	0.59767	22694.204781
target-AoA	-0.002191	-0.025215	0.020834	-0.186471	0.85212	22695.604288
target-conc	-0.028147	-0.051090	-0.005205	-2.404645	0.01641	22689.877562
FAS	-0.088982	-0.133312	-0.044653	-3.934208	0.00331	22620.853750
BAS	-0.064052	-0.086645	-0.041459	-5.556654	3.74762×10^{-8}	22665.335565
wordpair-RT	0.090631	0.051688	0.129573	4.561396	0.00129	22624.687438

Software used in this study

This study used the following software packages:

- Stimulus presentation was performed using MATLAB in combination with the Psychophysics toolbox.[63]
- Dipole modeling and forward model computation was provided by the OpenMEEG package.[64]
- Data analysis was performed using Python in combination with the NumPy and SciPy packages.[65]
- Plots were created using the Matplotlib package.[66]
- The implementation of the lSVM, logistic regression (LR), shrinkage-linear discriminant analysis (LDA) and OAS algorithms was provided by the Scikit-learn package.[67]
- The implementation of the xDAWN algorithm and Riemannian distance calculations were provided by the pyRiemann package.[68]
- The implementation of the Tucker-HOOI tensor decomposition algorithm was provided by the Scikit-tensor package.[69]
- Statistical analysis was performed using R[70] in combination with the LME4[71] and lmerTest[72] packages.

An implementation of the chained-LCMV and stLCMV beamformers used in this study, as well as methods to estimate a suitable template, can be found at: https://github.com/wmvanvliet/ERP-beamformer

[63] Hunter, 2007

[64] Gramfort, Alexandre and Papadopoulo, Théodore and Olivi, Emmanuel and Clerc, Maureen, 2011; Kybic, Clerc, Faugeras, Keriven, and Papadopoulo, 2006

[65] Oliphant, 2007

[66] Hunter, 2007

[67] Pedregosa et al., 2011

[68] Barachant, 2015

[69] Nickel, 2015

[70] R Development Core Team, 2008

[71] Bates, Maechler, Bolker, and Walker, 2015

[72] Kuznetsova, Bruun Brockhoff, and Haubo Bojesen Christensen, 2014

References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*(4), 491–6.

Baayen, R. H., Davidson, D. J., & M., B. D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bajo, M.-T. (1988). Semantic facilitation with pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 579–589.

Barachant, A. (2015). Pyriemann v0.2.2. doi:10.5281/zenodo.18982

Barachant, A. & Congedo, M. (2014). *A plug & play P300 BCI using information geometry*. GIPSA-lab, CNRS, Grenoble Universities. Grenoble, France. arXiv: 1409.0107v1. Retrieved from http://arxiv.org/abs/1409.0107

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. ArXiv e-print; in press, *Journal of Statistical Software*. Retrieved from http://arxiv.org/abs/1406.5823

Blankertz, Benjamin and Lemm, Steven and Treder, Matthias and Haufe, Stefan and Müller, Klaus Robert. (2011). Single-trial analysis and classification of ERP components - a tutorial. *NeuroImage*, *56*(2), 814–825.

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia*, *50*(11), 2659–68.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta psychologica*, *150*, 80–4.

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*. NetGames '06, *2*(2), 121–167.

Chen, Y., Wiesel, A., Eldar, Y. C., & Hero, A. O. (2010). Shrinkage algorithms for MMSE covariance estimation. *IEEE Transactions on Signal Processing*, *58*(10), 5016–5029.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour*, *12*(4), 335–359.

Comon, P. & Jutten, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Elsevier Science.

Croft, R. J. & Barry, R. J. (2000). Removal of ocular artifact from the EEG: a review. *Clinical Neurophysiology*, *30*(1), 5–19.

Cutler, A. (1981). Making up materials is a confounded nuisance, or: will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, *10*, 65–70.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain research*, *1084*(1), 89–103.

De Deyne, S. & Storms, G. (2008). Word associations: network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231.

Donchin, E. (1969). Discriminant analysis in average evoked response studies: the study of single trial data. *Electroencephalography and clinical neurophysiology*, *27*, 311–314.

Donchin, E. [E.] & Heffley, E. (1978). *Multivariate analysis of event-related potential data: A tutorial review* (D. Otto, Ed.). Government Printing Office.

Engemann, D. A. & Gramfort, A. (2015). Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. *NeuroImage*, *108*, 328–342.

Faraway, J. J. (2006). *Extending the linear model with R*. Boca Raton, FL: Chapman & Hall/CRC.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and language*, *115*(3), 149–61.

Grainger, J. & Holcomb, P. J. (2009). Watching the word go by: on the time-course of component processes in visual word recognition. *Language and Linguistics Compass*, *3*(1), 128–156.

Gramfort, Alexandre and Papadopoulo, Théodore and Olivi, Emmanuel and Clerc, Maureen. (2011). Forward field computation with OpenMEEG. *Computational Intelligence and Neuroscience*, *2011*, 923703.

Grosse-Wentrup, M., Liefhold, C., Gramann, K., & Buss, M. (2009). Beamforming in noninvasive brain-computer interfaces. *IEEE Transactions on Bio-medical Engineering*, *56*(4), 1209–19.

Handy, T. C. (2004). *Event-related potentials: a methods handbook*. Cambridge: MIT Press.

Hoormann, J., Falkenstein, M., Schwarzenau, P., & Hohnsbein, J. (1998). Methods for the quantification and statistical testing of ERP differences across conditions. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 103–109.

Hunter, J. (2007). Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, *9*(3), 90–95.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813.

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *Quarterly journal of experimental psychology (2006)*, *61*(7), 1036–66.

Jonmohamadi, Y., Poudel, G., Innes, C., & Weiss, D. (2014). Comparison of beamformers for EEG source signal reconstruction. *Biomedical Signal Processing and Control*, *14*, 175–188.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behavior research methods*, *42*(3), 643–650.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: a lexical decision study on 14,000 dutch mono- and disyllabic words and nonwords. *Frontiers in psychology*, *1*(November), 174.

Kutas, M. [M.]. (1993). In the company of other words: electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*, 533–572.

Kutas, M. [M.] & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and clinical neurophysiology*, *108*(5), 456–471.

Kutas, M. [Marta] & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M. [Marta] & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2014). *Lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. R package version 2.0-6. Retrieved from http://CRAN.R-project.org/package=lmerTest

Kybic, J., Clerc, M., Faugeras, O., Keriven, R., & Papadopoulo, T. (2006). Generalized head models for MEG/EEG: boundary element method beyond nested volumes. *Physics in medicine and biology*, *51*(5), 1333–1346.

Laszlo, S. & Federmeier, K. D. (2010). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, *48*, 176–186.

Lemm, Steven and Blankertz, Benjamin and Dickhaus, Thorsten and Müller, Klaus-Robert. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, *56*(2), 387–399.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Massachussetts: The MIT Press.

McNamara, T. P. (2005). What is semantic priming and why should anyone care about it? In *Semantic priming: Perspectives from memory and word recognition* (Chap. 1, pp. 3–7). New York.

Münte, T. F., Urbach, T. P., Düzel, E., & Kutas, M. (2000). Event-related brain potentials in the study of human cognition and neuropsychology. In F. Boller, J. Grafman, & G. Rizzolatti (Eds.), *Handbook of neuropsychology* (2nd, Chap. 7, Vol. 1, pp. 1–97). Elsevier Science.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (Chap. 9, pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nickel, M. (2015). Scikit-tensor. https://github.com/mnick/scikit-tensor. GitHub.

Niedermeyer, E. & da Silva, F. H. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., & Cadusch, P. J. (1997). EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology*, *103*, 499–515.

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science Engineering*, *9*(3), 10–20.

Parra, L. C., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., & Sajda, P. (2003). Single-trial detection in EEG and MEG: Keeping it linear. *Neurocomputing*, *52-54*, 177–183.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pernet, C. R., Sajda, P., & Rousselet, G. a. (2011). Single-trial analyses: Why bother? *Frontiers in Psychology*, *2*(November), 1–2.

Petten, C. V. & Van Petten, C. (2014). Examining the N400 semantic context effect item-by-item: relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology*, *94*(3), 407–419.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org

Rivet, B., Souloumiac, A., Attina, V., & Gibert, G. (2009). xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. *IEEE Transactions on Bio-medical Engineering*, *56*(8), 2035–2043.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*(6), 110–114.

Schäfer, J. & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, *4*(1), Article32.

Van Veen, B. D. & Buckley, K. M. (1988). Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, *5*(April), 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 Transactions on Bio-medical Engineering, 44*(9), 867–880.

van Vliet, M., Manyakov, N. V., Storms, G., Fias, W., Wiersema, J. R., & Van Hulle, M. M. (2014). Response-related potentials during semantic priming: the effect of a speeded button response task on ERPs. *PLoS ONE, 9*(2), e87650.

Vossen, H., van Breukelen, G., Hermens, H., van Os, J., & Lousberg, R. (2011). More potential in statistical analyses of event-related potentials: a mixed regression approach. *International journal of methods in psychiatric research, 20*(3), e56–68.

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: a practical guide using statistical software.* Boca Raton, FL: Taylor & Francis Group, LLC.

Yu, K., Shen, K., Shao, S., Ng, W. C., Kwok, K., & Li, X. (2012). A spatio-temporal filtering approach to denoising of single-trial ERP in rapid image triage. *Journal of Neuroscience Methods, 204*(2), 288–95.