STIMULUS-AWARE SPATIAL FILTERING FOR SINGLE-TRIAL NEURAL RESPONSE AND TEMPORAL RESPONSE FUNCTION ESTIMATION IN HIGH-DENSITY EEG WITH APPLICATIONS IN AUDITORY RESEARCH

Authors

Neetha Das **(Corresponding Author)**

Dept. Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

Dept. Neurosciences, ExpORL, KU Leuven, Herestraat 49 bus 721, B-3000 Leuven, Belgium

email: ndas@esat.kuleuven.be

Jonas Vanthornhout

Dept. Neurosciences, ExpORL, KU Leuven, Herestraat 49 bus 721, B-3000 Leuven, Belgium

Tom Francart

Dept. Neurosciences, ExpORL, KU Leuven, Herestraat 49 bus 721, B-3000 Leuven, Belgium

Alexander Bertrand

Dept. Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

Document details

Number of pages: 24

Number of figures: 6

Number of tables: 0

Declarations of interest: none.

February 5, 2019

Stimulus-aware spatial filtering for single-trial neural response and temporal response function estimation in high-density EEG with applications in auditory research

Neetha Das[†⋆], Jonas Vanthornhout[⋆], Tom Francart[⋆], Alexander Bertrand[†]

Abstract

Objective. Neural responses recorded using electroencephalography (EEG) and magnetoencephalography (MEG) can be used to study how our brain functions, as well as for various promising brain computer interface (BCI) applications. However, a common problem is the low signal to noise ratio (SNR) which makes it challenging to estimate task-related neural responses or the temporal response function (TRF) describing the linear relationship between the stimulus and the neural response, particularly over short data windows. To address these, we present an algorithm that takes advantage of the multi-channel nature of the recordings, and knowledge of the presented stimulus, to achieve a joint noise reduction and dimensionality reduction using spatial filtering. *Methods.* Forward modeling is used to project the stimulus onto the electrode space. The second-order statistics of this estimated desired signal and the raw neural data are used to estimate spatial filters that maximize the SNR of the neural response, based on a generalized eigenvalue decomposition. *Main Results.* 1. For synthesized EEG data, over a range of SNRs, our filtering resulted in significantly better TRF estimates from 20 s trials, compared to unfiltered EEG data. 2. On a dataset from 28 subjects who listened to a single-talker stimulus, our method resulted in correlations between predicted neural responses and the original EEG data that were significantly higher compared to standard approaches. 3. On a dataset of 16 subjects attending to 1 speaker in a two-speaker scenario, our method resulted in attention decoding accuracies which were higher compared to existing forward modelling methods. *Significance.* Our algorithm presents a data-driven way to denoise and reduce dimensionality of neural data, thus aiding further analysis, by utilizing the knowledge of the stimulus. The method is computationally efficient, and does not require repeated trials, thereby relieving experiment design from the necessity of presenting repeated stimuli to the subjects.

Index Terms

EEG processing, temporal response function, speech entrainment, spatial filtering, denoising, forward modeling, attention decoding.

I. Introduction

Understanding how auditory stimuli influence neural activity is one of the goals of auditory neuroscience research. Towards this goal, several methods have been proposed to model the relationship between the auditory stimuli and

The work is funded by KU Leuven Special Research Fund C14/16/057 and OT/14/119, FWO project nrs. 1.5.123.16N and G0A4918N, FWO PhD grant awarded to Jonas Vanthornhout (1S10416N), the ERC (637424 and 802895) under the European Union's Horizon 2020 research and innovation programme. The scientific responsibility is assumed by its authors.

[†] KU Leuven, Dept. Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics. Kasteelpark Arenberg 10, B-3001 Leuven, Belgium.

[⋆] KU Leuven, Dept. Neurosciences, ExpORL. Herestraat 49 bus 721, B-3000 Leuven, Belgium.

the elicited neural responses. In magnetoencephalography (MEG) or electroencephalography (EEG) studies, forward models based on linear temporal response functions (TRFs) are often used to model the path between the stimulus and each of the electrodes/sensors [1, 2]. In this work we focus on the EEG modality but the methods discussed are not limited to EEG. In case of speech stimuli and multitalker scenarios, TRFs have often been used to linearly map speech stimulus features such as stimulus envelopes, speech spectrograms, phonemes etc. of both attended and unattended speakers, to the neural activity of the listener [2–5]. TRFs not only have high temporal precision, but are also sensitive to attentional modulation [3, 6]. Forward modeling thus comes with the advantages of being able to investigate the TRFs and gain a better understanding of how our brain handles auditory stimuli [2, 4, 7], and also the possibility to identify the brain regions involved with stimulus processing [2, 3, 8, 9].

On the other hand, linear backward modeling, where the stimulus features are reconstructed from the neural activity, is also a commonly used method [2, 10–14]. Unlike the forward model, this approach makes use of inter-channel covariances to design the decoder. The correlations between the reconstructed and the original stimulus features are thus higher, than those of predicted and original neural responses in the forward modeling approach. However, the decoder coefficients themselves can not directly be interpreted, unlike forward models, where TRFs for different channels can be visualized using topoplots.

In algorithms that deal with neural activity, often dimensionality reduction is key [6, 15, 16]. Dimensionality reduction works on the assumption that the data of interest lies in a lower dimensional space than its original representation. As mentioned earlier, forward models map the auditory stimulus to each of the electrodes, thus preserving spatial information of stimulus-related cortical activity. However, unlike algorithms that employ backward modeling, they do not use cross-channel information to regress out non-stimulus related activity [17]. In such cases, dimensionality reduction can help to transform neural recordings from a multi-electrode system, into a signal subspace with fewer components (than electrodes) and better SNR of the stimulus following responses. The algorithm may then use these components themselves to achieve its goal, or project the components back to the electrode space, effectively performing a denoising operation, before further processing. Thus, dimensionality reduction goes hand-in-hand with denoising.

Principal component analysis (PCA) is often used for dimensionality reduction, in which case the principal components corresponding to lower variance are discarded [18, 19]. However, this approach relies on the assumption that low variance corresponds to non-relevant activity, which can be a rather restrictive assumption to make, particularly for EEG data, where the SNRs are poor, and large artifacts with high variance are common. Another approach, independent component analysis (ICA) [20], works on the assumption that the components (or sources) are statistically independent. It is often used to remove components corresponding to artifacts with specific patterns (like eye-blinks) [21]. ICA, however, does not perform well in the extraction of signal components that are far below the noise floor, as is the case for neural responses to speech. Another method, joint decorrelation [22] follows the formulation of linear denoising source separation (DSS) [23] to improve the SNR of the activity of interest in the neural data through a double PCA step, one for pre-whitening the data and one for dimensionality reduction. In the most commonly used version of DSS, a criterion of stimulus-evoked reproducibility is used, maximizing the evoked to induced ratio. This, however, required repeated trials, which renders it impractical for many EEG applications.

Canonical correlation analysis (CCA) also reduces dimensionality by finding separate linear transformations for the stimulus as well as neural responses, such that in the projected subspace, the neural response and the stimulus are maximally correlated [19, 24, 25]. This method, however, could filter out certain (frequency) components of the stimulus and/or the neural responses to maximize the correlation coefficient, thereby leading to distorted TRFs or neural response estimates.

Except for CCA, all methods mentioned above do not exploit the knowledge of the stimulus. In this work, we propose a data-driven stimulus-aware method for dimensionality reduction of neural responses, which finds a set of spatial filters using the generalized eigenvalue decomposition, with the goal of maximizing the SNR of stimulus following responses. For neural responses to continuous speech stimuli, we show that the proposed method results in effective dimensionality reduction/denoising without the need for data from repeated stimulus trials as in the DSS method or phase-locked averaging techniques. We validate the performance of the proposed method in the following three contexts.

1) Short-term TRF estimation, where short trials are used to estimate TRFs that map the auditory stimulus envelope to the neural responses. Estimation of TRFs from short trials has multiple advantages: a) TRFs to different stimuli can be visualized to track the effect of attention on the TRF shapes, and eventually to even decode attention in realtime without any prior training of decoders [6, 26], b) TRFs also may be visualized using topoplots to gain an understanding of the brain regions that are actively involved in the auditory attention process.
2) Stimulus envelope tracking, where in a single speaker scenario, neural responses to the stimulus envelope are predicted using forward modeling, i.e., finding TRFs that map the stimulus envelope to the neural responses. The correlations between the original neural responses and the predicted stimulus following responses are analyzed [5, 27]. The analysis of these correlations can not only contribute to advancing our knowledge of how the brain responds to auditory stimuli under different conditions, but also has potential to act as objective measures of speech intelligibility [4, 28–31].
3) Auditory attention decoding, where an estimate is made as to which of multiple speakers, a person is attending, using forward modeling [17]. Attention decoding finds application in brain computer interfaces (BCIs) such as neuro-steered hearing prostheses where accurate information about a person's auditory attention can be used to steer the noise suppression beamformer towards the direction of the attended speaker [32–35]. In particular, for real-time tracking of auditory attention, forward models are popular [6, 16, 26], in which case, dimensionality reduction is an important ingredient.

The outline of the paper is as follows. In Section II, we describe the algorithm for denoising and dimensionality reduction. In Section III, we describe the validation within the three aforementioned contexts. The strengths of our method as well as similarities and differences with existing denoising and dimensionality reduction methods are discussed in Section IV. Finally, we summarize and draw conclusions in Section V.

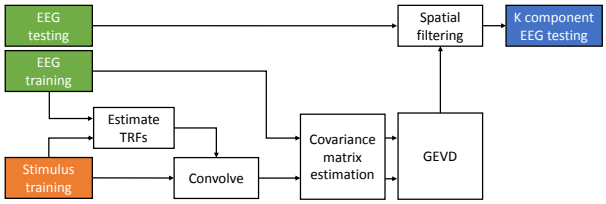

Fig. 1: Block diagram of the proposed method.

II. Algorithm

The main goal of this work is to find a spatial filter, or a group of spatial filters, that combine the EEG channels in such a way that the power of the auditory stimulus following responses is increased in comparison to the rest of the neural activity captured by the electrodes. We aim to achieve this goal using the following steps to train the filters: 1) Estimate the spatial covariance of the desired neural response across the different EEG channels, using the known stimulus as side information, 2) find a signal subspace were the ratio of the power of the desired neural response to the power of the background EEG signal (i.e. SNR) is maximized, 3) project the EEG data into the new signal subspace, thereby performing a joint dimensionality reduction and denoising, 4) back-project the data into the channel space if necessary, and then perform the necessary analysis. The steps of the algorithm are illustrated in figure 1 and explained in detail below.

A. Neural response covariance estimation

The EEG signal, at a time index t, is defined as a C-dimensional vector $\mathbf{m}(t) = [m_1(t), m_2(t), ...m_C(t)]^T \in \mathbb{R}^C$ where C denotes the number of channels, and $m_i(t)$ represents the EEG sample from the i^{th} channel at time index t. In a scenario where the subject listens to a speech signal $s(t)$, it is known that the neural responses of the person entrain to the envelope of the speech stream. Therefore, the EEG signal can be assumed to be the sum of the activity driven by the speech stimulus $\mathbf{x}(t)$ and the rest of the neural activity $\mathbf{n}(t)$ at time t.

$$\mathbf{m}(t) = \mathbf{x}(t) + \mathbf{n}(t). \tag{1}$$

The stimulus following neural response $\mathbf{x}(t)$ can be modeled by a linear temporal response function (TRF) [1, 2, 4] that maps the stimulus envelope (including N_l-1 time lagged versions of it) $\mathbf{s}(t) = [s(t), s(t-1), ..., s(t-N_l+1)]^T \in \mathbb{R}^{N_l}$ to the neural response.

$$\mathbf{x}(t) = \mathbf{W}^T\mathbf{s}(t), \tag{2}$$

where $\mathbf{W} \in \mathbb{R}^{N_l \times C}$ is a matrix containing the per-channel TRFs in its columns. $\mathbf{W}$ can be estimated by minimizing the mean square error (MMSE) between $\mathbf{x}(t)$ and $\mathbf{m}(t)$:

$$\tilde{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}} \; E\{||\mathbf{W}^T\mathbf{s}(t) - \mathbf{m}(t)||^2\}, \tag{3}$$

where $E\{.\}$ denotes the expected value operator. The solution of (3) is given by

$$\tilde{\mathbf{W}} = \mathbf{R}_{\mathbf{ss}}^{-1}\mathbf{r}_{\mathbf{sm}}, \tag{4}$$

where $\mathbf{R}_{\mathbf{ss}} = E\{\mathbf{s}(t)\mathbf{s}(t)^T\} \in \mathbb{R}^{N_l \times N_l}$ is the covariance matrix of the stimulus envelope, and $\mathbf{r}_{\mathbf{sm}} = E\{\mathbf{s}(t)\mathbf{m}(t)^T\} \in \mathbb{R}^{N_l \times C}$ is the cross-correlation matrix between the stimulus envelope and the EEG data.

Consider we have N samples of data, with which to estimate the TRF $\tilde{\mathbf{W}}$. The stimulus envelope $\mathbf{S} = [\mathbf{s}(1), \mathbf{s}(2), ..., \mathbf{s}(N)] \in \mathbb{R}^{N_l \times N}$ and the EEG signal $\mathbf{M} = [\mathbf{m}(1), \mathbf{m}(2), ..., \mathbf{m}(N)] \in \mathbb{R}^{C \times N}$ taken over N samples, can be used to estimate the covariance matrix of the stimulus envelope as $\mathbf{R}_{\mathbf{ss}} \approx (\mathbf{S}\mathbf{S}^T)/N$ and the cross-correlation matrix between the stimulus envelope and the EEG data as $\mathbf{R}_{\mathbf{sm}} \approx (\mathbf{S}\mathbf{M}^T)/N$ which can then be used in (4) to estimate the TRF matrix $\tilde{\mathbf{W}}$. With the estimated TRF, the desired neural response $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), ..., \mathbf{x}(N)] \in \mathbb{R}^{C \times N}$ can be computed using (2), the spatial covariance matrix of which can then be estimated as $\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{X}\mathbf{X}^T\} \approx (\mathbf{X}\mathbf{X}^T)/N$.

Here, we are essentially estimating long-term TRFs over a large amount of data (i.e., N is very large). This will ensure that we have a good estimate of the second order statistics of the neural response, allowing us to compute spatial filters (the next step) to improve the SNR of the EEG data. This higher SNR EEG data should enable us to better estimate short-term TRFs over much shorter windows with fewer samples.

B. Spatial filter estimation

Having estimated the spatial covariance matrix $\mathbf{R}_{\mathbf{xx}}$ of the desired neural response $\mathbf{x}(t)$, the next step is to find a set of spatial filters that maximizes the SNR at their outputs. Consider the matrix $\mathbf{P}_K \in \mathbb{R}^{C \times K}$ which contains K spatial filters in its columns. By multiplying the EEG data matrix with $\mathbf{P}_K^T$, the C EEG channels are combined into K output channels where $K < C$, in such a way that the SNR at the output is maximized. For the sake of simplicity, we first assume $K = 1$, thus reducing $\mathbf{P}_K$ to a single spatial filter denoted by the vector $\mathbf{p} \in \mathbb{R}^C$. The maximum-SNR criterion thus becomes:

$$\underset{\mathbf{p}}{\operatorname{argmax}} \ \frac{E\{(\mathbf{p}^T\mathbf{x}(t))^2\}}{E\{(\mathbf{p}^T\mathbf{n}(t))^2\}} = \underset{\mathbf{p}}{\operatorname{argmax}} \ \frac{\mathbf{p}^T\mathbf{R}_{\mathbf{xx}}\mathbf{p}}{\mathbf{p}^T\mathbf{R}_{\mathbf{nn}}\mathbf{p}}. \tag{5}$$

The stationary points of (5) can be shown to satisfy[1]

$$\mathbf{R}_{\mathbf{xx}}\mathbf{p} = \lambda\mathbf{R}_{\mathbf{nn}}\mathbf{p}. \tag{6}$$

which defines a generalized eigenvalue problem for the matrix pencil $(\mathbf{R}_{\mathbf{xx}}, \mathbf{R}_{\mathbf{nn}})$ [36–38], which can also be solved by considering the equivalent eigenvalue problem

$$\mathbf{R}\mathbf{p} = \lambda\mathbf{p}. \tag{7}$$

involving the matrix $\mathbf{R} = \mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{R}_{\mathbf{xx}}$. From (6), it follows that

[1]A brief derivation is given in Appendix A for the interested reader.

$$\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p} = \lambda\mathbf{p}^T\mathbf{R_{nn}}\mathbf{p} \Leftrightarrow \lambda = \frac{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}}{\mathbf{p}^T\mathbf{R_{nn}}\mathbf{p}} \tag{8}$$

which implies that λ is equal to the output SNR of the spatial filter. Therefore, in order to maximize the SNR, we should set $\mathbf{p}$ equal to the (generalized) eigenvector corresponding to the largest (generalized) eigenvalue λ.

So far, we have considered a spatial filter $\mathbf{p}$ which combines the C EEG channels in an optimal way to obtain a channel with SNR maximized. This can be further extended to solving the problem of finding a filter bank $\mathbf{P}_K \in \mathbb{R}^{C\times K}$ consisting of K spatial filters that maximizes the total output SNR, by finding K generalized eigenvectors corresponding to the K highest eigenvalues from the GEVD of $(\mathbf{R_{xx}}, \mathbf{R_{nn}})$.

In this work, we apply the GEVD to the matrix pair $(\mathbf{R_{xx}}, \mathbf{R_{mm}})$ where $\mathbf{R_{mm}} = E\{\mathbf{m}(t)\mathbf{m}(t)^T\} \in \mathbb{R}^{C\times C}$ contains both noise and desired neural responses. This is equivalent to solving the optimization problem of maximizing the signal to signal plus noise ratio (SSNR). $\mathbf{R_{mm}}$ can easily be computed from the raw EEG data as $\mathbf{R_{mm}} \approx (\mathbf{M}\mathbf{M}^T)/N$, while $\mathbf{R_{xx}}$ is estimated by first estimating long-term TRFs as discussed in section II-A. Assuming the stimulus following responses $\mathbf{x}(t)$ and the background EEG (noise) $\mathbf{n}(t)$ are uncorrelated, the maximal SSNR criterion reduces to the maximal SNR criterion as shown in Appendix B. We refer to the resulting spatial filterbank as the stimulus-informed GEVD (SI-GEVD) filter.

C. Project EEG signal onto the new subspace

The top K eigenvectors from the GEVD corresponding to the K largest λ values can be used to reduce the dimensionality of the raw EEG data with C channels, to K filtered EEG components. If $\mathbf{P}_K = [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_K]$ is the matrix containing the top K eigenvectors in its columns, then the EEG signal in the new subspace is

$$\mathbf{m}_{proj}(t) = \mathbf{P}_K^T\mathbf{m}(t). \tag{9}$$

There are many ways to choose K. The cumulative λ values can be plotted and the top K values for which the cumulative sum shows a significant increase can be chosen. The optimum K can also be chosen based on the application. For example, in the case of auditory attention decoding using forward modeling or correlation analysis, one can use cross validation to select the number of components K that eventually results in the highest attention decoding accuracy or the best correlations respectively.

D. Back-projection to the electrode space

Back-projection to the electrode space can be done based on the matrix $\mathbf{Q}_C = \mathbf{P}_C^{-T}$, where $\mathbf{P}_C$ contains all the generalized eigenvectors of the pencil $(\mathbf{R_{xx}}, \mathbf{R_{mm}})$ in its columns. Note that $\mathbf{P}_K$ contains a subset of the columns of $\mathbf{P}_C$. By selecting the subset with the same column indices from $\mathbf{Q}_C$, we obtain the matrix $\mathbf{Q}_K$, which can be used to project the compressed data $\mathbf{m}_{proj}(t)$ back to the electrode space with minimal error in the least squares sense (see Appendix C):

$$\bar{\mathbf{m}}(t) = \mathbf{Q}_K\mathbf{m}_{proj}(t). \tag{10}$$

This results in a denoised version $\bar{\mathbf{m}}(t)$ of the original EEG channels $\mathbf{m}(t)$. The denoised signals can then be used for short-term TRF estimation, attention decoding, etc.

III. Validated Experiments

A. Data collection and pre-processing

1) Dataset I: This dataset consists of EEG data collected during the experiments described in [8]. The data consists of 64 channel EEG recordings at 8192 Hz collected from 16 normal hearing subjects (8 male, 8 female) between the age of 17 and 30 years. The subjects were asked to attend to one of two stories that were simultaneously presented to them. We refer to [8] for details on the experiment conditions. For the analyses in this paper (including those in section III-D), we used only the data corresponding to conditions where the presented stimuli were filtered by head-related transfer functions (HRTFs) (≈38 minutes per subject), in order to provide a realistic acoustic impression of the location of the speakers. In [8], it has been shown that this data resulted in significantly better attention decoding performance in comparison to dichotically presented unfiltered stimuli. The multi-channel Wiener filtering (MWF) method in [39] was used on the EEG data for artifact removal. The EEG data was bandpass filtered between 1 and 9 Hz. In [2, 8, 40, 41], it was shown that cortical envelope tracking in this frequency range results in the best attention decoding performance. The audio envelope was determined by filtering the speech waveform with a gammatone filterbank (with 15 filters) followed by powerlaw compression (exponent = 0.6) on the absolute value of each filter's output signal [12]. The resulting signals from all subbands were then summed, after which the resulting signal was downsampled to 32 Hz and filtered using the same 1-9 Hz bandpass filter as for the EEG signal. The EEG data was referenced to the Cz electrode (therefore, $C = 63$). All data were normalized to have zero mean and unit standard deviation.

2) Dataset II: This dataset consists of EEG data from [30]. In this study the EEG (64 channels at 8192 Hz) was recorded from 27 normal hearing subjects (8 male, 19 female, average age: 23 years). The participants listened to a 14-minute story in silence (Milan, narrated in Flemish by Stijn Vranken) and to Matrix stimuli. A Matrix stimulus consists of a concatenation of 40 Matrix sentences [42] with a 1-s silence between each sentence. As each sentence is approximately 2 s long, this yields a stimulus of 120 s with 80 s of speech. Each Matrix stimulus was repeated 3 or 4 times, yielding 6 to 8 minutes of EEG recordings. All stimuli were presented diotically. In the original study [30] the Matrix sentences were presented in silence and in noise. For this study, we only used the data without background noise.

Preprocessing of the EEG data was done similar to [30]. The EEG was highpass filtered (second order Butterworth with cut-off at 0.5 Hz) and downsampled to 256 Hz before applying the MWF for artifact rejection [39]. It was then re-referenced to the Cz electrode (therefore $C = 63$). Next, the EEG was bandpass filtered within the delta band (0.5 - 4 Hz). This band was found to be the most useful to predict speech intelligibility [30]. Finally, the EEG was downsampled to 128 Hz. The stimulus envelope was estimated using a gammatone filterbank with 28 filters. A powerlaw compression was then applied (exponent = 0.6) to the absolute values in each of the subbands. The resulting 28 subband envelopes were bandpass filtered (as was done for the EEG data) and averaged to obtain one single envelope, and downsampled to 128 Hz. The EEG data and stimulus data were normalized to have zero mean and unit standard deviation.

B. Short-term TRF estimation

1) EEG synthesis: In order to reliably estimate any improvement that SI-GEVD filtering might bring to the problem of TRF estimation from short trials, we use synthesized EEG built from an underlying template of a TRF. The EEG data synthesis was based on a real EEG recording from [8]. Details of the data collection and preprocessing can be found in section III-A1. The data from each subject was split into 38 trials of 60 s duration. The 'mTRF' toolbox [43] was used to find the TRFs mapping the attended speech envelope (and it's lagged versions up to 400 ms) to the EEG data, per subject per trial. The TRFs were then averaged across trials and subjects. The TRF corresponding to the channel 'Tp8' was then taken to generate a base TRF template. The TRFs for all other channels were taken to be scaled versions of this base template, with the scaling corresponding to the relative root mean square value of the TRFs in those channels, with respect to that of 'Tp8'. Thus a C-channel TRF matrix $\mathbf{W}_{\text{base}} \in \mathbb{R}^{N_l \times C}$ containing the per-channel TRFs in its columns was synthesized. Lags of 0 to 400 ms were used corresponding to N_l = 32 Hz $\times$ 400 ms + 1 = 14 samples.

The stimulus following response was then synthesized by convolving the per-channel TRFs with the attended speech envelope and the results are put in a matrix $\mathbf{X} \in \mathbb{R}^{N \times C}$ where N is the total number of EEG samples (38 minutes). Noise $\mathbf{N}$ to be added to the stimulus following response was synthesized by flipping in time, the concatenated 38 minute EEG from a subject. The flipping operation ensures that the neural response in the EEG data is uncorrelated to the synthesized neural response. EEG responses $\mathbf{M}$ to the attended speech were simulated by adding noise $\mathbf{N}$ to $\mathbf{X}$ at different SNRs. SNR is defined here as 20 times the logarithm (base 10) of the ratio of the mean (over channels) of root mean square (rms) values of the neural response ($\mathbf{X}$) to the mean of the rms values of the noise ($\mathbf{N}$).

2) TRF estimation: Once EEG data was synthesized using the base TRF template, we focused on estimating short-term TRFs from the synthesized EEG as well as it's SI-GEVD filtered version, and assessing the quality of these TRF estimates in comparison to the base TRF template. We estimated TRFs $\mathbf{W}_{\text{raw}} \in \mathbb{R}^{N_l \times C}$ from 20 s trials of the unfiltered EEG data $\mathbf{M}$ using a standard ridge regression technique using the mTRF toolbox [43]. For the forward modeling, the regularization parameter for ridge regularization was automatically chosen (on a per trial basis) as the maximum absolute value among all the elements in the corresponding trial's speech covariance matrix $\mathbf{R}_{SS}$ (default choice in the 'mTRF' toolbox). In addition, a leave-one-trial-out cross validation was used for estimating the SI-GEVD filter, and estimating the TRFs $\mathbf{W}_{\text{SI-GEVD}} \in \mathbb{R}^{N_l \times C}$ on the SI-GEVD filtered data. This is explained as follows. For the i^{th} trial, the EEG data $\mathbf{M}_{test}[i] = [\mathbf{m}((i-1)L+1), \mathbf{m}((i-1)L+2), ..., \mathbf{m}(iL)] \in \mathbb{R}^{C \times L}$ formed the i^{th} test trial, while the EEG data from all trials except the test trial, formed the training set $\mathbf{M}_{train}[i] \in \mathbb{R}^{C \times (N-L)}$, where L = 32 Hz $\times$ 20 s = 640 samples and N = 32 Hz $\times$ 20 s $\times$ 115 trials = 73600 samples. Similarly, the stimulus envelope of the test trial was taken as $\mathbf{S}_{test}[i] = [\mathbf{s}((i-1)L+1), \mathbf{s}((i-1)L+2), ..., \mathbf{s}(iL)] \in \mathbb{R}^{N_l \times L}$ where $\mathbf{s}(t)$ is a vector containing N_l = 14 samples corresponding to N_l delays, as seen in (2). Stimulus envelopes from all trials except the test trial of index i, were concatenated to form the training set $\mathbf{S}_{train}[i] \in \mathbb{R}^{N_l \times (N-L)}$.

We estimate the SI-GEVD filters $\mathbf{P}_K[i]$ on the training data $\mathbf{S}_{train}[i]$ and $\mathbf{M}_{train}[i]$ based on the method described in section II. Since, in all cases, there was a single dominant generalized eigenvalue, only one generalized eigenvector (K = 1) was used as the SI-GEVD filter. The test trial was then SI-GEVD filtered and back-projected to obtain the denoised EEG data $\bar{\mathbf{M}}_{test}[i]$. The TRFs $\mathbf{W}_{\text{SI-GEVD}}[i]$ were then estimated from $\bar{\mathbf{M}}_{test}[i]$ and $\mathbf{S}_{test}[i]$ within the

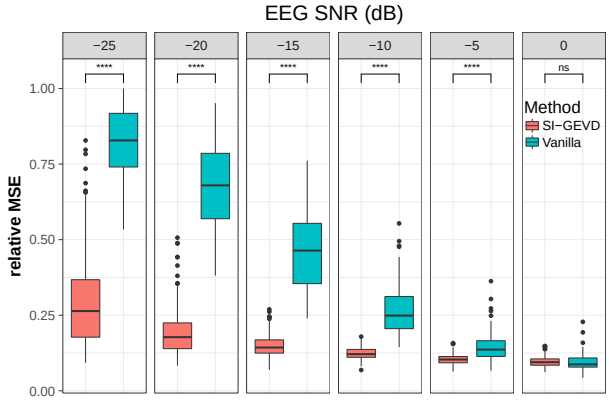

Fig. 2: Relative MSE between base TRFs and estimated TRFs using the two approaches: Vanilla using raw EEG Vs SI-GEVD filtered EEG. Each box contains one datapoint per trial. Comparison between the methods were done using Wilcoxon's signed-rank test: '****' for $p < 0.0001$.

20 s trial, in the same way as how $\mathbf{W}_{\text{raw}}[i]$ were estimated.

3) Results: For each trial, the estimated TRFs of all channels were concatenated into a single vector estimated TRF: $\hat{\mathbf{w}}_{\text{raw}}[i] \in \mathbb{R}^{CN_l \times 1}$ from the raw EEG data, and $\hat{\mathbf{w}}_{\text{SI-GEVD}}[i] \in \mathbb{R}^{CN_l \times 1}$ from the denoised EEG data. The base TRFs of all channels were also concatenated to obtain a single vector base TRF $\hat{\mathbf{w}}_{\text{base}}[i] \in \mathbb{R}^{CN_l \times 1}$. In order to eliminate any differences due to scaling, a scaling factor was estimated by applying a least squares fitting such that the estimated TRF vector (referred to, in general, as $\hat{\mathbf{w}}$) was scaled to fit $\hat{\mathbf{w}}_{\text{base}}[i]$ in the minimum mean squared error sense. This scaling factor was found as follows:

$$\alpha[i] = (\hat{\mathbf{w}}^T[i]\hat{\mathbf{w}}[i])^{-1}\hat{\mathbf{w}}^T[i]\hat{\mathbf{w}}_{\text{base}}[i]. \tag{11}$$

The mean squared error (MSE) between the scaled estimated TRF vector and the base TRF vector was computed for all the trials. The MSE values were normalized by the square of the norm of the base TRF vectors to find relative MSEs in the range of 0 to 1 resulting in the following definition:

$$MSE_{rel}[i] = \frac{||\alpha[i]\hat{\mathbf{w}}[i] - \hat{\mathbf{w}}_{\text{base}}[i]||^2}{||\hat{\mathbf{w}}^2_{\text{base}}[i]||^2}. \tag{12}$$

For a range of SNRs, figure 2 shows the relative MSE for TRF estimation from raw EEG ('vanilla') and from SI-GEVD filtered EEG (after back-projection). The relative MSEs for the two approaches were compared using Wilcoxon's signed-rank test (with Bonferroni correction). While no significant difference was found between the performance of the two approaches at SNR = 0 dB (i.e. signal power = noise power), the relative MSEs corresponding to the TRFs estimated from SI-GEVD filtered EEG were found to be significantly lower ($p < 0.0001$) for more realistic SNRs ranging from -5 dB to -25 dB, showing that SI-GEVD filter is effective in denoising the data which translates into better TRF estimates over short trials.

C. Speech envelope tracking

1) Data Analysis: To analyze the effectiveness of the SI-GEVD filter in measures of envelope tracking, we used the data from [30]. Details of the data collection and preprocessing can be found in section III-A2. The goal of

this analysis was to use forward modeling to predict speech following neural responses. Correlations of the original EEG and the predicted EEG were then used to quantify the effectiveness of denoising using the SI-GEVD filter. The data from the story presentation was used only for training the SI-GEVD filters and the TRFs, and not for testing. On the other hand, the Matrix dataset was used both in the training as well as the testing set while performing leave-one-trial-out cross validation. This dataset was split into 6 trials, each trial consisting of 20 s of EEG data. Thus, for each test trial (Matrix), the SI-GEVD filter and the TRF was trained on a) the remaining 5 Matrix trials and their repetitions (3 repetitions $\times$ 5 trials $\times$ 20 s), and b) the data from the story presentation (14 minutes).

The i^{th} test trial $\mathbf{M}_{test}[i] \in \mathbb{R}^{C\times L}$ contained L = 128 Hz $\times$ 20 s = 2560 time samples per channel, while the training set $\mathbf{M}_{train}[i] \in \mathbb{R}^{C\times N}$ consisted of N = (3 Matrix repetitions $\times$ 5 trials $\times$ L) + (14 minutes of story $\times$ 60 s $\times$ 128 Hz) time samples per channel. Stimulus envelopes $\mathbf{S}_{test}[i] \in \mathbb{R}^{N_l\times L}$ and $\mathbf{S}_{train}[i] \in \mathbb{R}^{N_l\times N}$ consisted of lags up to 75 ms, and hence N_l = 128 Hz $\times$ 75 ms + 1 = 11 samples.

For each trial, the training set data was used to find the SI-GEVD spatial filter $\mathbf{P}_C[i]$ and the back-projection matrix $\mathbf{Q}_C[i]$. After choosing the number of components to be used (the method of choosing K is explained further on), the EEG data in the test set was denoised by applying the SI-GEVD spatial filter and back-projected to the C-channel space resulting in

$$\bar{\mathbf{M}}_{test}[i] = \mathbf{Q}_K[i]\mathbf{P}_K^T[i]\mathbf{M}_{test}[i]. \tag{13}$$

This denoised EEG data was then used for TRF estimation. From the denoised training set and the raw training set, the TRFs $\mathbf{W}_{\text{SI-GEVD}}[i]$ and $\mathbf{W}_{\text{raw}}[i] \in \mathbb{R}^{N_l\times C}$ that mapped the attended stimulus data $\mathbf{S}_{train}[i]$ to the denoised EEG data $\bar{\mathbf{M}}_{train}[i]$ and the raw EEG data $\mathbf{M}_{train}[i]$ respectively were computed as in section II-B using the mTRF toolbox. The stimulus following responses of the test trial were predicted by convolving the obtained TRFs with the stimulus envelope $\mathbf{S}_{test}[i]$ in the testing set:

$$\mathbf{X}_{\text{raw}}[i] = \mathbf{W}_{\text{raw}}^T[i]\mathbf{S}_{test}[i], \tag{14}$$

$$\mathbf{X}_{\text{SI-GEVD}}[i] = \mathbf{W}_{\text{SI-GEVD}}^T[i]\mathbf{S}_{test}[i]. \tag{15}$$

The Spearman correlation coefficients between the EEG data ($\mathbf{M}_{test}[i]$ (raw) and $\bar{\mathbf{M}}_{test}[i]$ (denoised)) and the corresponding predicted stimulus following responses ($\mathbf{X}_{\text{raw}}[i]$ and $\mathbf{X}_{\text{SI-GEVD}}[i]$) were computed.

The correlation coefficients obtained were averaged across repetitions (from repeated stimulus trials) yielding 378 correlations (6 trials, 63 sensors). Next, we averaged the correlations in the sensor dimension using a trimmed mean (percentage = 50%) to get one correlation per trial per subject. The number of components K was chosen based on leave-one-trial-out cross-validation. For each of the 6 trials, we choose the number of components that maximized the average correlation between the predicted EEG and the actual EEG on the other 5 trials.

In order to compare the results of SI-GEVD filtering with other methods which also aim for dimensionality reduction, we also implemented the DSS approach [23]. DSS first performs a pre-whitening on the EEG data, then averages the repeated trials in order to boost the SNR of the stimulus response, and then applies a PCA step on the resulting signal to find a set of spatial filters for dimensionality reduction. It is noted that, in our case, only the

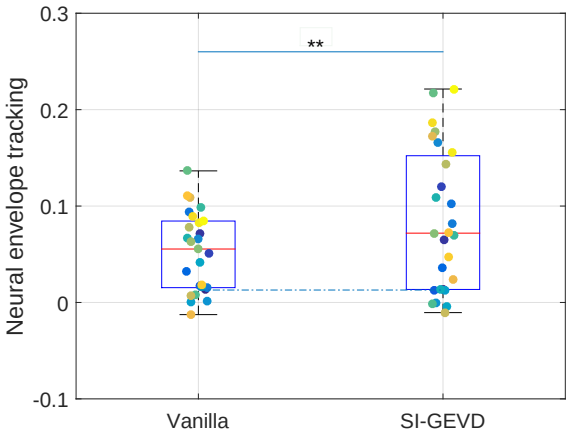

Fig. 3: Mean correlations on 20 s trials: Comparison of mean correlations from forward modeling using raw data ('vanilla') versus SI-GEVD filtered data (SI-GEVD). The dotted line indicates significance level (95%) for the correlations. Each box contains one datapoint per subject. In the plot, comparison between the two methods was done using Wilcoxon's signed-rank test: '**' for $p < 0.0100$.

Matrix data contains repetitions, so we can only use this part of the data for training the DSS filters. These filters then replace the SI-GEVD filters in the procedure mentioned above. For the normal forward model (using $\mathbf{W}_{\text{raw}}$) based on unfiltered EEG data (referred to as 'vanilla'), the correlations from the best 8 channels (with the highest correlations) across subjects are averaged (channels Pz, POz, P2, CPz, Oz, O2, PO3 and FC1).

2) Results: Correlations from the 2 approaches, vanilla and SI-GEVD, were compared using Wilcoxon's signed-rank test with $\alpha = 0.05$ (figure 3). The correlations of the SI-GEVD filtering approach were found to be significantly higher than those of the vanilla approach ($p < 0.0100, W = 81$). For the DSS approach, in contrast to the SI-GEVD approach, we found that the power ratio plot of DSS components, indicating SNR corresponding to each component space, degraded slowly over the components, and therefore, it was clear that choosing a few top components from the DSS approach for spatial filtering would not result in good denoising. As expected, the DSS approach resulted in correlations that were significantly lower than those from the SI-GEVD approach ($p < 0.0010$, for any choice of K), and not significantly different from those from the vanilla approach. This is probably due to the low number of repetitions available.

In order to have a better understanding of the influence of the various GEVD components in the SI-GEVD approach, we investigated the weights of the back-projection matrix (each column corresponds to a component), for multiple components. Figure 4 shows the weights of each channel for the first component of the back-projection matrix ($\mathbf{Q}_C$) averaged over all training sets (each with one trial removed) and subjects. Note that the different spatial filters consist of generalized eigenvectors, which are only defined up to an arbitrary scaling and sign. Therefore, for each component, we normalized across channels, so that the sum of the absolute values of the weights equal one, and averaged the normalized spatial filter of each trial after correcting for polarity. To find those channel weights which are significantly different from zero across subjects, using one sample cluster mass statistics [44], a reference distribution is built by repeatedly ($n = 5000$) and randomly swapping the sign of the weights in $\mathbf{Q}_K$

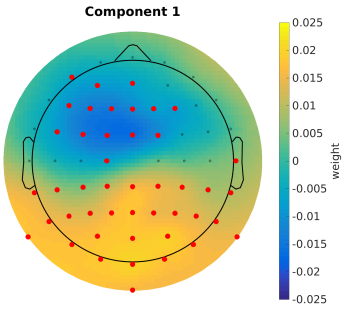

Fig. 4: Visualization of the SI-GEVD spatial filter's first component: Red dots indicate channel weights that are significantly different from zero.

across subjects and calculating a t-statistic per channel. The actual t-statistic per channel is then calculated without swapping the sign. We obtained the p-value of each channel by calculating the proportion of random samplings that have a higher t-statistic than the actual t-statistic. In this test we also take the position of the channels into account as neighbouring channels can have similar weights, using this information to reduce the family-wise error rate (details in [44]). This test will show a significance when the weights of the channel show a low variability over subjects. In figure 4, channel weights significantly different from zero ($p < 0.0100$, one sample cluster mass statistics) across subjects are shown in red.

When using only the first component, we observed two regions with an opposite polarity that contribute most to the significant difference. One is the fronto-central region, and the other is the parietal-occipital region. This is in line with other EEG research that estimates spatial filters based on speech processing [45] and EEG research that shows the topography of speech processing in the brain [46–48]. For the second component we did not find any significant channels. This is consistent with the results of the number of components needed to have optimal correlations and with the SSNR of each component (1 GEVD component for 16 out of 27 subjects). All of this is converging evidence that the GEVD-based spatial filter is very effective in denoising and dimensionality reduction.

D. Auditory attention decoding

1) Data Preprocessing: For attention decoding, we used the dataset described in section III-A1. Also as described in III-A1, the preprocessing of the data was done similar to [8], and hence has differences with the preprocessing done in section III-A2 (based on [30]). The preprocessed data set at a sampling rate of 32 Hz was divided into 112 trials of 20 s per subject.

2) Data Analysis: Adopting a leave-one-trial-out cross-validation approach, for each test trial, we used the remaining 111 trials as the training set for spatial filter estimation and TRF estimation. The difference of this application's algorithm with that of envelope tracking is that here EEG is predicted from both the stimuli (attended and unattended), and an attention decoding decision is made based on how well these predictions correlate to the original EEG data, as will be explained further on. We perform forward modeling in the SI-GEVD component space, and no back-projection is performed.

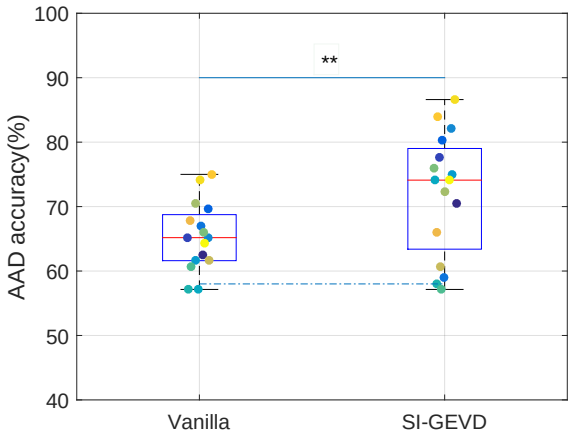

Fig. 5: Attention decoding performance on 20 s trials: SI-GEVD based forward modeling compared to forward modeling with mean of correlations over all channels ('vanilla'). The dotted line indicates chance level (95 percentile). Each box contains one mean decoding accuracy per subject, also indicated by colored points. In the plot, comparisons between methods are done using Wilcoxon's signed-rank tests: '**' for $p < 0.0100$.

In the i^{th} trial, EEG data $\mathbf{M}_{test}[i] \in \mathbb{R}^{C\times L}$ and the attended stimulus envelope $\mathbf{S}_{test}[i] \in \mathbb{R}^{N_l\times L}$ consisted of L = 32 Hz $\times$ 20 s = 640 time samples, while the training set consisted of EEG data $\mathbf{M}_{train}[i] \in \mathbb{R}^{C\times N}$ and attended stimulus envelope $\mathbf{S}_{train}[i] \in \mathbb{R}^{N_l\times N}$ with N = 111 $\times$ L time samples. Stimulus envelope lags up to 400 ms were used. Therefore N_l = 32 Hz $\times$ 400 ms + 1 = 14 samples.

For each trial i, the training set data was used to find the SI-GEVD spatial filter $\mathbf{P}_C[i]$. TRFs with respect to the attended stimulus were estimated from the training set using the raw EEG data on the one hand, or the SI-GEVD filtered data (no back projection) on the other hand. This resulted in two different TRF matrices, denoted as $\mathbf{W}_{\text{raw}} \in \mathbb{R}^{N_l\times C}$ and $\mathbf{W}_{\text{SI-GEVD}} \in \mathbb{R}^{N_l\times K}$, respectively. The speech stimuli in the test trial, which was left out during the training of the TRFs and the SI-GEVD filters, was convolved with these TRFs to predict the stimulus following responses in the test trial. The predicted responses (from both the stimuli) were then compared with the original EEG data $m(t)$ (for $\mathbf{W}_{\text{raw}}$) and the SI-GEVD filtered EEG data $\mathbf{m}_{proj}(t)$ (for $\mathbf{W}_{\text{SI-GEVD}}$) by computing the Spearman correlation over each channel.

In each case, i.e. for the raw EEG data as well as SI-GEVD filtered EEG data, the stimulus that resulted in a reconstruction that yielded a higher mean correlation with the original EEG components was then considered to be the attended stimulus. In this application, we used a different number of spatial filters K for each subject. In order to do this, an optimal number of components was found for each trial by cross-validating the difference between attended and unattended correlations when GEVD components were added one at a time, until there was no more improvement. Then, for each subject, K was chosen to be the highest among the optimal number of components among all trials. In the forward modeling approach using raw EEG data, referred to as 'vanilla' in figure 5, the correlations from all the channels were averaged before making the attention decision. Since the dataset (from [8]) included recordings where the stimuli were repeated (3 repetitions $\times$ 2 minutes $\times$ 2 story parts), we also tested attention decoding on EEG data which was filtered based on the DSS approach.

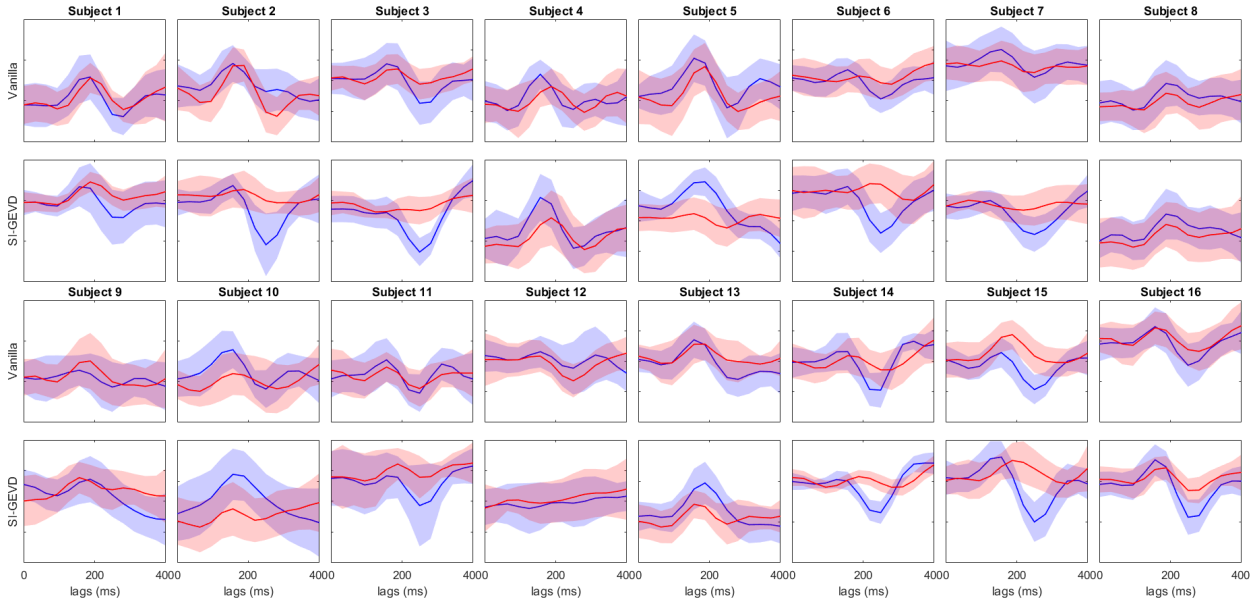

Fig. 6: Mean attended (blue) and unattended TRFs from 120 s trials of each subject, estimated from raw data ('vanilla') as well as after SI-GEVD filtering. The horizontal axis represents lags in ms.

3) Results: Attention decoding performance in the vanilla approach was compared with the SI-GEVD filtering approach. In the SI-GEVD filtering approach, as per the criterion described before for the choice of the number of GEVD components K, 3 GEVD components were used for 3 subjects, 2 GEVD components for 5 subjects, and 1 GEVD component for the remaining 8 subjects. As can be seen in figure 5, the median decoding accuracy with the SI-GEVD approach was found to be 74.1%, which was 8.9% higher than that of the vanilla approach (65.2%). Using Wilcoxon's signed rank test, this difference was found to be significant ($p = 0.0050, W = 119.5$). We also used the available repetition trials in the data, to extract DSS components. However, similar to the results from section III-C, the power ratio plot showed a gradual decrease of power ratio over components. Hence, there were not just a few dominant components that could explain most of the variance in the biased matrix, and as a consequence, choosing a small K would not be enough to ensure a high SNR in the filtered data. In short, this approach did not meet the purpose of dimensionality reduction. Also, using only the first DSS component for AAD resulted in accuracies of 9 out of 16 subjects being below chance level, and the results were significantly lower than with the vanilla approach (Wilcoxon's signed-rank test: $p = 0.0020, W = 118$).

To gain a better understanding of the denoising ability of our approach, in the context of TRF estimation on short trials, we estimated TRFs (as in section III-B) for the attended and unattended stimuli on 120 s trials, employing the vanilla as well as the SI-GEVD approach (in C channel space). Figure 6 shows the mean TRFs (averaged over the channels 'Tp7' and 'Tp8', and over trials), and their standard deviation (area around the mean) for each of the 16 subjects. We can see clear differences between the two approaches. The SI-GEVD approach results in patterns that have a better separation between attended and unattended TRFs compared to the vanilla approach. This could be the reason for the better attention decoding accuracies that we see when the forward models are trained on

SI-GEVD filtered data.

IV. Discussion

We presented an algorithm for a joint denoising and dimensionality reduction of EEG data, with the goal of maximizing SNR, in the context of auditory stimulus following responses. In order to obtain the spatial filters that perform the dimensionality reduction, we employed a stimulus-informed generalized eigenvalue decomposition (SI-GEVD) of the covariance matrix of the desired signal and the covariance matrix of the raw EEG signal. The desired signal is the stimulus following response (attended stimulus in the case of attention decoding) which is estimated from the raw EEG signal and the stimulus by forward model estimation (TRFs), and then convolving the stimulus with the TRFs. We analyzed the performance of the proposed algorithm in 3 experiments in the context of auditory neuroscience - short-term TRF estimation, speech envelope tracking and auditory attention decoding.

In the context of TRF estimation from short trials, we used synthesized EEG in order to have access to the ground truth TRFs to assess performance. We found that, over a range of SNRs, SI-GEVD filtering effectively denoised the EEG data, resulting in significantly lower relative MSEs with respect to the base TRFs, in comparison to TRF estimation from unfiltered EEG data (vanilla). For speech envelope tracking in a single speaker scenario, the correlations between the EEG data and the stimulus following responses (predicted) were also found to be significantly better when using the SI-GEVD filter, compared to DSS or hand-picking channels. For multi-talker scenarios, SI-GEVD filtering resulted in significantly higher attention decoding accuracies than by averaging correlations over a set of raw EEG channels, and then making attention decoding decisions. The attention decoding performance was also found to be better than when DSS-based spatial filtering was employed. It is important to note here that the DSS-based approach relies on data from stimulus repetitions, and the lower performance for this method can be attributed to the lack of repetitions present in our data to have a good estimate of the covariance matrix of the desired signal. Another difference is that DSS makes an estimate of the stimulus response by averaging repetitions, without knowledge of the stimulus, whereas our method uses forward modeling for a better estimate of the desired signal using the attended stimulus envelope. DSS is often used to denoise MEG responses with a small number of repetitions [2, 16, 26], but the SNR in EEG responses is considerably lower [49].

The xDAWN algorithm [50] addressed the problem of dimensionality reduction in the context of brain computer interfaces - specifically a P300 speller. The goal of discriminating epochs containing a P300 potential evoked by a target stimulus (visual) and epochs corresponding to non-target stimuli was tackled by estimating the P300 subspace from raw EEG data, and projecting the raw EEG onto this subspace, effectively enhancing the P300 evoked potentials [50, 51]. The xDAWN algorithm also uses a forward encoding model to estimate a template for the event-related potentials (ERPs) using knowledge of the stimulus onsets, and consequently, the second order statistics of the neural responses evoked by the target stimulus pulses. In [51], it is also stated that the computationally expensive QR decompositions and singular value decomposition (SVD) used for estimation of the spatial filter can be replaced by a GEVD on the matrix pair of the correlation matrices of the stimulus evoked responses and the raw EEG data. Thus, we can draw parallels between our proposed method and the xDAWN algorithm, with the difference that, here, we aim for SNR improvement of neural responses evoked by continuous stimuli, for which the statistics are

estimated using knowledge of the continuous stimulus waveform (instead of activation times of repeated pulsed stimuli in [51]), and we make use of a GEVD, which is computationally more efficient.

Spatial filtering to denoise neural data thereby separating stimulus-related activity from non-stimulus related activity was also addressed by [52]. They utilized DSS [23] where PCA and normalization is used to pre-whiten the data. The whitened data is then submitted to a bias function, followed by another PCA to determine the orientations that maximize the bias function, i.e., essentially the power of the biased data. The bias function averages the epochs under the same stimulus condition, thereby reinforcing stimulus-evoked activity. PCA on the biased data results in a rotation matrix which can be applied to the whitened data to get signal components which can be kept or discarded depending on their bias score. The remaining components can be projected back to the sensor space resulting in denoised responses. The steps of pre-whitening and PCA to find the rotation matrix can in principle be replaced by a GEVD on the correlation matrix of the biased data (with enhanced stimulus following responses) and the correlation matrix of the raw data, as done in the procedure explained in Section II. The key difference between our proposed method and [52] using DSS is the way we estimate the covariance matrix of the stimulus response. Our method estimates a TRF between the (known) stimulus and the recorded EEG data, after which the stimulus is convolved with the TRF to obtain an estimate of the stimulus response in each channel. [52] uses the method of epoch-averaging on pre-whitened data. Here, the stimulus does not have to be known, but a major practical limitation comes with the condition that it can only be performed if there are enough repetitions of the same stimulus in the data to effectively enhance the stimulus following responses by averaging. In addition, the GEVD approach renders pre-whitening and the PCA steps unnecessary, resulting in improved computational efficiency. The joint decorrelation method proposed by [22] generalized the DSS based denoising approach with the freedom to choose a bias filter based on the task at hand, and also cited equivalence in results with the two-step procedure for diagonalizing two correlation matrices described by [53, 54].

Our method also has similarities with the method of common spatial pattern analysis (CSP) [55–57] that aims to solve a binary classification task. To achieve this, the algorithm finds a set of spatial filters that maximize the variance of the projected data for one class, and minimize the variance for the other, such that the output power of the filters can be used as features in a classification task. Similar to DSS and our approach, the solution is found by a GEVD, yet on a different set of covariance matrices, namely the two covariance matrices corresponding to both classes. Similar to the DSS approach in [52], in CSP, epochs of data are sometimes averaged, per class, in order to have a good estimate of the covariance matrix of the desired signal.

Another common way of dimensionality reduction is to handpick channels from brain regions that are expected to have the strongest neural responses to the stimulus of interest, or simply use all channels, and calculate a mean over these channels for the parameter (e.g., correlation) analyzed in the problem [5, 17, 58]. Another method is to exhaustively train and test the system (eg. attention decoders, or forward models) dropping one channel at a time [13, 14, 59] until the desired goal is achieved with fewer channels. Other methods include PCA which has an underlying assumption that the desired responses are also those that exhibit the largest variance in the data (which is definitely not the case for auditory responses), independent component analysis (ICA) which produces statistically independent components, etc. These approaches, however, are not guided by the signal of interest - neural responses

that entrain to the stimulus envelope. In [52], the better performance of the DSS approach compared to these methods was already demonstrated. With our work, we present a data-driven dimensionality reduction/denoising approach which makes use of stimulus information to find a spatial filter that can project the neural responses to a signal space where the power of the stimulus following responses is maximized. The method is computationally more efficient than DSS, and it also relieves the experiment design from the necessity of recording neural responses for multiple repetitions of the stimulus for the purpose of averaging condition specific epochs.

V. Conclusion

Dimensionality reduction and denoising are important steps in the procedure to analyze neural responses, and particularly EEG data which often has low SNRs. In the context of analyzing stimulus following neural responses, methods like DSS and PCA are often used in order to denoise and/or reduce the dimensionality of the data. When it comes to forward modeling approaches, often researchers handpick channels based on their knowledge of locations of the strongest stimulus following cortical activity. In this paper, we present and evaluate a stimulus-informed GEVD-based filtering approach which makes use of predicted stimulus following responses to find a spatial filter that maximizes the SNR at its output. Not only is the proposed method computationally efficient in dimensionality reduction/denoising, but also has the advantage of not relying on repeated trials for the spatial filter estimation. We have shown the benefits of using our approach by analyzing 3 different applications in the field of auditory neuroscience, and compared the performance of the proposed method with other approaches like DSS or averaging over selected channels.

Appendix

A. Spatial filter estimation

Since $\mathbf{p}$ is only defined up to an arbitrary scaling, we can fix the output noise to have unit variance, thus fixing the denominator in (5). The optimization problem can then be redefined as:

$$\begin{aligned} \mathbf{p} &= \underset{\mathbf{p}}{\operatorname{argmax}} \ \mathbf{p}^T \mathbf{R}_{\mathbf{xx}} \mathbf{p}, \\ &\text{s.t.}, \mathbf{p}^T \mathbf{R}_{\mathbf{nn}} \mathbf{p} = 1. \end{aligned} \tag{A.1}$$

The Lagrangian of this constrained optimization problem is given by:

$$\mathcal{L}(\mathbf{p}, \lambda) = \mathbf{p}^T \mathbf{R}_{\mathbf{xx}} \mathbf{p} + \lambda(\mathbf{p}^T \mathbf{R}_{\mathbf{nn}} \mathbf{p} - 1). \tag{A.2}$$

The solution can then be found by setting the partial derivative of (A.2) with respect to $\mathbf{p}$ to zero.

$$\frac{\partial \mathcal{L}(\mathbf{p}, \lambda)}{\partial \mathbf{p}} = 2(\mathbf{R}_{\mathbf{xx}} - \lambda \mathbf{R}_{\mathbf{nn}})\mathbf{p} = 0, \tag{A.3}$$

$$\mathbf{R}_{\mathbf{xx}} \mathbf{p} = \lambda \mathbf{R}_{\mathbf{nn}} \mathbf{p}. \tag{A.4}$$

B. Optimization problem equivalence

The SI-GEVD filter is found by solving of the optimization problem

$$\underset{\mathbf{p}}{\operatorname{argmax}} \ \frac{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}}{\mathbf{p}^T\mathbf{R_{nn}}\mathbf{p}} \tag{B.1}$$

resulting in the GEVD of the matrix pencil $(\mathbf{R_{xx}}, \mathbf{R_{nn}})$. It is shown below that solving the optimization problem for maximizing the SSNR (derivation was taken from [60])

$$\underset{\mathbf{p}}{\operatorname{argmax}} \ \frac{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}}{\mathbf{p}^T\mathbf{R_{mm}}\mathbf{p}} \tag{B.2}$$

results in the same solution as that for maximizing the SNR (B.1).

$$\begin{aligned}
\underset{\mathbf{p}}{\operatorname{argmax}} \ \frac{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}}{\mathbf{p}^T\mathbf{R_{mm}}\mathbf{p}} &= \underset{\mathbf{p}}{\operatorname{argmax}} \ \frac{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}}{\mathbf{p}^T(\mathbf{R_{xx}}+\mathbf{R_{nn}})\mathbf{p}} \\
&= \underset{\mathbf{p}}{\operatorname{argmin}} \ \frac{\mathbf{p}^T(\mathbf{R_{xx}}+\mathbf{R_{nn}})\mathbf{p}}{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}} \\
&= \underset{\mathbf{p}}{\operatorname{argmin}} (1 + \frac{\mathbf{p}^T\mathbf{R_{nn}}\mathbf{p}}{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}}) \\
&= \underset{\mathbf{p}}{\operatorname{argmin}} \ \frac{\mathbf{p}^T\mathbf{R_{nn}}\mathbf{p}}{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}} \\
&= \underset{\mathbf{p}}{\operatorname{argmax}} \ \frac{\mathbf{p}^T\mathbf{R_{xx}}\mathbf{p}}{\mathbf{p}^T\mathbf{R_{nn}}\mathbf{p}}
\end{aligned} \tag{B.3}$$

Thus, it can be concluded that the GEVD of the matrix pencil $(\mathbf{R_{xx}}, \mathbf{R_{mm}})$ would result in the same SI-GEVD filter as for $(\mathbf{R_{xx}}, \mathbf{R_{nn}})$.

C. Back-projection to the electrode space

If $\mathbf{P_C}$ contains the set of generalized eigenvectors of the matrix pencil $(\mathbf{R_{xx}}, \mathbf{R_{mm}})$ in its columns, and $\mathbf{Q}_C = \mathbf{P}_C^{-T}$, it follows that

$$\mathbf{R_{mm}} = \mathbf{Q}_C\mathbf{\Lambda_m}\mathbf{Q}_C^T \tag{C.1}$$

$$\mathbf{R_{xx}} = \mathbf{Q}_C\mathbf{\Lambda_x}\mathbf{Q}_C^T \tag{C.2}$$

where $\mathbf{\Lambda_m}$ and $\mathbf{\Lambda_x}$ are diagonal matrices such that $\mathbf{\Lambda} = \mathbf{\Lambda_m}^{-1}\mathbf{\Lambda_x}$ [36]. Here $\mathbf{\Lambda}$ is the diagonal matrix containing all the generalized eigenvalues. The goal is to find a filter $\mathbf{V} \in \mathbb{R}^{K\times C}$ which projects the compressed K-component data $\mathbf{m}_{proj}(t)$ back to the electrode space with minimal error in the least squares sense:

$$\mathbf{V} = \underset{\mathbf{V}}{\operatorname{argmin}} \ E\{||\mathbf{V}^T\mathbf{m}_{proj}(t) - \mathbf{m}(t)||^2\}. \tag{C.3}$$

This is a standard minimum mean squared error problem of which the solution is given as

$$\mathbf{V} = (E\{\mathbf{m}_{proj}(t)\mathbf{m}_{proj}(t)^T\})^{-1}E\{\mathbf{m}_{proj}(t)\mathbf{m}^T\}. \tag{C.4}$$

This can be rewritten using the notation introduced in section II as

$$\begin{aligned}\mathbf{V} &= (\mathbf{P}_K^T \mathbf{R}_{\mathbf{mm}} \mathbf{P}_K)^{-1} \mathbf{P}_K^T \mathbf{R}_{\mathbf{mm}} \\ &= (\mathbf{E}_K^T \mathbf{P}_C^T \mathbf{R}_{\mathbf{mm}} \mathbf{P}_C \mathbf{E}_K)^{-1} \mathbf{E}_K^T \mathbf{P}_C^T \mathbf{R}_{\mathbf{mm}}\end{aligned} \tag{C.5}$$

where $\mathbf{E}_K \in \mathbb{R}^{C \times K}$ is a matrix that chooses the first K columns of $\mathbf{P}_C$ so that $\mathbf{P}_K = \mathbf{P}_C \mathbf{E}_K$. Since $\mathbf{Q}_C = \mathbf{P}_C^{-T}$, from (C.1) it follows:

$$\begin{aligned}\mathbf{V} &= (\mathbf{E}_K^T \mathbf{Q}_C^{-1} (\mathbf{Q}_C \mathbf{\Lambda}_{\mathbf{m}} \mathbf{Q}_C^T) \mathbf{Q}_C^{-T} \mathbf{E}_K)^{-1} \mathbf{E}_K^T \mathbf{Q}_C^{-1} (\mathbf{Q}_C \mathbf{\Lambda}_{\mathbf{m}} \mathbf{Q}_C^T) \\ &= (\mathbf{E}_K^T \mathbf{\Lambda}_{\mathbf{m}} \mathbf{E}_K)^{-1} \mathbf{E}_K^T \mathbf{\Lambda}_{\mathbf{m}} \mathbf{Q}_C^T \\ &= \mathbf{\Lambda}_{\mathbf{m,k}}^{-1} [\mathbf{\Lambda}_{\mathbf{m,k}} | 0] \mathbf{Q}_C^T = [\mathbf{I}_K | 0] \mathbf{Q}_C^T \\ &= \mathbf{E}_K^T \mathbf{Q}_C^T = \mathbf{Q}_K^T\end{aligned} \tag{C.6}$$

where $\mathbf{\Lambda}_{\mathbf{m,k}}$ is a diagonal matrix containing the first K diagonal values of $\mathbf{\Lambda}_{\mathbf{m}}$, and $[\mathbf{\Lambda}_{\mathbf{m,k}}|0]$ represents the concatenation of $\mathbf{\Lambda}_{\mathbf{m,k}}$ with $C - K$ columns of all zero values (equivalent to choosing the first K rows of $\mathbf{\Lambda}_{\mathbf{m}}$). The derivation shows that the filter that should be used to back-project $\mathbf{m}_{proj}(t)$ to the electrode space is $\mathbf{V}^T = \mathbf{Q}_K = \mathbf{P}_C^{-T} \mathbf{E}_K$.

ACKNOWLEDGMENTS

The authors would like to thank all the subjects for their participation in the study.

REFERENCES

[1] E. C. Lalor and J. J. Foxe, "Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution," *European journal of neuroscience*, vol. 31, no. 1, pp. 189–193, 2010.

[2] N. Ding and J. Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," *Journal of neurophysiology*, vol. 107, no. 1, pp. 78–89, 2012.

[3] A. J. Power, J. J. Foxe, E.-J. Forde, R. B. Reilly, and E. C. Lalor, "At what time is the cocktail party? a late locus of selective attention to natural speech," *European Journal of Neuroscience*, vol. 35, no. 9, pp. 1497–1503, 2012.

[4] N. Ding and J. Z. Simon, "Emergence of neural encoding of auditory objects while listening to competing speakers," *Proc. National Academy of Sciences*, vol. 109, no. 29, pp. 11 854–11 859, 2012.

[5] G. M. Di Liberto, J. A. O'Sullivan, and E. C. Lalor, "Low-frequency cortical entrainment to speech reflects phoneme-level processing," *Current Biology*, vol. 25, no. 19, pp. 2457–2465, 2015.

[6] S. Akram, J. Z. Simon, and B. Babadi, "Dynamic estimation of the auditory temporal response function from MEG in competing-speaker environments," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1896–1905, 2017.

[7] G. M. Di Liberto and E. C. Lalor, "Isolating neural indices of continuous speech processing at the phonetic level," in *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*. Springer, 2016, pp. 337–345.

[8] N. Das, W. Biesmans, A. Bertrand, and T. Francart, "The effect of head-related filtering and ear-specific decoding bias on auditory attention detection," *Journal of neural engineering*, vol. 13, no. 5, p. 056014, 2016.

[9] O. Etard, M. Kegler, C. Braiman, A. E. Forte, and T. Reichenbach, "Real-time decoding of selective attention from the human auditory brainstem response to continuous speech," *bioRxiv*, p. 259853, 2018.

[10] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, no. 7397, pp. 233–236, 2012.

[11] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cerebral Cortex*, p. bht355, 2014.

[12] W. Biesmans, N. Das, T. Francart, and A. Bertrand, "Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 5, pp. 402–412, 2017.

[13] B. Mirkovic, S. Debener, M. Jaeger, and M. De Vos, "Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications," *Journal of neural engineering*, vol. 12, no. 4, p. 046007, 2015.

[14] S. A. Fuglsang, T. Dau, and J. Hjortkjær, "Noise-robust cortical tracking of attended speech in real-world acoustic scenes," *Neuroimage*, vol. 156, pp. 435–444, 2017.

[15] N. Ding, M. Chatterjee, and J. Z. Simon, "Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure," *Neuroimage*, vol. 88, pp. 41–46, 2014.

[16] S. Akram, A. Presacco, J. Z. Simon, S. A. Shamma, and B. Babadi, "Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling," *NeuroImage*, vol. 124, pp. 906–917, 2016.

[17] D. D. Wong, S. A. A. Fuglsang, J. Hjortkjær, E. Ceolini, M. Slaney, and A. de Cheveigné, "A comparison of temporal response function estimation methods for auditory attention decoding," *bioRxiv*, p. 281345, 2018.

[18] J. P. Cunningham and M. Y. Byron, "Dimensionality reduction for large-scale neural recordings," *Nature neuroscience*, vol. 17, no. 11, p. 1500, 2014.

[19] A. de Cheveigné, D. D. Wong, G. M. D. Liberto, J. Hjortkjaer, M. Slaney, and E. Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206 – 216, 2018.

[20] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[21] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal-state-of-the-art and guidelines," *Journal of neural engineering*, vol. 12, no. 3, p. 031001, 2015.

[22] A. de Cheveigné and L. C. Parra, "Joint decorrelation, a versatile tool for multichannel data analysis," *Neuroimage*, vol. 98, pp. 487–505, 2014.

[23] J. Särelä and H. Valpola, "Denoising source separation," *Journal of machine learning research*, vol. 6, no. Mar, pp. 233–272, 2005.

[24] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[25] A. de Cheveigné, G. M. Di Liberto, D. Arzounian, D. D. Wong, J. Hjortkjær, S. Fuglsang, and L. C. Parra, "Multiway canonical correlation analysis of brain data," *NeuroImage*, 2018.

[26] S. Miran, S. Akram, A. Sheikhattar, J. Z. Simon, T. Zhang, and B. Babadi, "Real-time tracking of selective auditory attention from M/EEG: A bayesian filtering approach," *Frontiers in neuroscience*, vol. 12, 2018.

[27] S. J. Aiken and T. W. Picton, "Human cortical responses to the speech envelope," *Ear and hearing*, vol. 29, no. 2, pp. 139–157, 2008.

[28] M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, "Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech," *Current Biology*, vol. 28, no. 5, pp. 803–809, 2018.

[29] G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, "Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech," *eNeuro*, pp. ENEURO–0084, 2018.

[30] J. Vanthornhout, L. Decruy, J. Wouters, J. Z. Simon, and T. Francart, "Speech intelligibility predicted from neural entrainment of the speech envelope," *Journal of the Association for Research in Otolaryngology*, vol. 19, no. 2, pp. 181–191, Apr 2018.

[31] D. Lesenfants, J. Vanthornhout, E. Verschueren, L. Decruy, and T. Francart, "Predicting individual speech intelligibility from the neural tracking of acoustic-and phonetic-level speech representations," *bioRxiv*, p. 471367, 2018.

[32] S. Van Eyndhoven, T. Francart, and A. Bertrand, "EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 1045–1056, 2017.

[33] N. Das, S. Van Eyndhoven, T. Francart, and A. Bertrand, "EEG-based attention-driven speech enhancement for noisy speech mixtures using N-fold multi-channel Wiener filters," in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 1660–1664.

[34] A. Aroudi, D. Marquardt, and S. Doclo, "EEG-based auditory attention decoding using steerable binaural superdirective beamformer," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada*, 2018.

[35] J. O'Sullivan, Z. Chen, J. Herrero, G. M. McKhann, S. A. Sheth, A. D. Mehta, and N. Mesgarani, "Neural decoding of attentional selection in multi-speaker environments without access to clean sources," *Journal of neural engineering*, vol. 14, no. 5, p. 056001, 2017.

[36] G. H. Golub and C. F. Van Loan, "Matrix computations, 3rd," 1996.

[37] L. Parra and P. Sajda, "Blind source separation via generalized eigenvalue decomposition," *Journal of Machine Learning Research*, vol. 4, no. Dec, pp. 1261–1269, 2003.

[38] W. Biesmans, A. Bertrand, J. Wouters, and M. Moonen, "Optimal spatial filtering for auditory steady-state response detection using high-density EEG," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 857–861.

[39] B. Somers, T. Francart, and A. Bertrand, "A generic EEG artifact removal algorithm based on the multi-channel wiener filter," *Journal of neural engineering*, vol. 15, no. 3, p. 036007, 2018.

[40] E. M. Z. Golumbic, N. Ding, S. Bickel, P. Lakatos, C. A. Schevon, G. M. McKhann, R. R. Goodman, R. Emerson, A. D. Mehta, J. Z. Simon *et al.*, "Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party"," *Neuron*, vol. 77, no. 5, pp. 980–991, 2013.

[41] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex," *PLoS-Biology*, vol. 10, no. 1, p. 175, 2012.

[42] H. Luts, S. Jansen, W. Dreschler, and J. Wouters, "Development and normative data for the Flemish/Dutch Matrix test," 2014.

[43] M. J. Crosse, G. M. Di Liberto, A. Bednar, and E. C. Lalor, "The multivariate temporal response function (mtrf) toolbox: a matlab toolbox for relating neural signals to continuous stimuli," *Frontiers in human neuroscience*, vol. 10, p. 604, 2016.

[44] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG-and MEG-data," *Journal of neuroscience methods*, vol. 164, no. 1, pp. 177–190, 2007.

[45] J. P. Dmochowski, J. J. Ki, P. DeGuzman, P. Sajda, and L. C. Parra, "Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity," *NeuroImage*, vol. 180, pp. 134–146, 2018.

[46] J. Hjortkjaer, J. Märcher-Rørsted, S. A. Fuglsang, and T. Dau, "Cortical oscillations and entrainment in speech processing during working memory load," *European Journal of Neuroscience*, Feb. 2018.

[47] C. Braiman, E. A. Fridman, M. M. Conte, H. U. Voss, C. S. Reichenbach, T. Reichenbach, and N. D. Schiff, "Cortical response to the natural speech envelope correlates with neuroimaging evidence of cognition in severe brain injury," *Current Biology*, vol. 28, no. 23, pp. 3833–3839, 2018.

[48] J. A. O'Sullivan, S. A. Shamma, and E. C. Lalor, "Evidence for neural computations of temporal coherence in an auditory scene and their enhancement during active listening," *Journal of Neuroscience*, vol. 35, no. 18, pp. 7256–7263, 2015.

[49] Y.-Y. Kong, A. Somarowthu, and N. Ding, "Effects of spectral degradation on attentional modulation of cortical auditory responses to continuous speech," *Journal of the Association for Research in Otolaryngology*, vol. 16, no. 6, pp. 783–796, 2015.

[50] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xdawn algorithm to enhance evoked potentials: application to brain–computer interface," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, 2009.

[51] B. Rivet, H. Cecotti, A. Souloumiac, E. Maby, and J. Mattout, "Theoretical analysis of xdawn algorithm: application to an efficient sensor selection in a p300 bci," in *19th European Signal Processing Conference (EUSIPCO 2011)*, 2011, pp. 1382–1386.

[52] A. de Cheveigne and J. Z. Simon, "Denoising based on spatial filtering," *Journal of neuroscience methods*, vol. 171, no. 2, pp. 331–339, 2008.

[53] K. Fukunaga and W. L. Koontz, "Application of the Karhunen-Loeve expansion to feature selection and ordering," *IEEE Transactions on computers*, vol. 100, no. 4, pp. 311–318, 1970.

[54] K. Fukunaga, *Introduction to statistical pattern recognition*. Elsevier, 2013.

[55] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.

[56] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[57] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE transactions on biomedical engineering*, vol. 53, no. 11, pp. 2274–2281, 2006.

[58] J. Zou, J. Feng, T. Xu, P. Jin, C. Luo, F. Chen, J. Zhang, and N. Ding, "Auditory and language contributions to neural encoding of speech features in noisy environments," *bioRxiv*, p. 377838, 2018.

[59] A. M. Narayanan and A. Bertrand, "The effect of miniaturization and galvanic separation of EEG sensor devices in an auditory attention detection task," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 77–80.

[60] J. Wouters, F. Kloosterman, and A. Bertrand, "Towards online spike sorting for high-density neural probes using discriminative template matching with suppression of interfering spikes," *Journal of neural engineering*, 2018.