Subspace-Based Suppression of Cortical Stimulation Artifacts

Po. T. Wang[1], Colin M. McCrimmon[2], Payam Heydari[3], An. H. Do[4], and Zoran Nenadic[1,3]

***Abstract*— Bi-directional brain-computer interfaces for the restoration of movement and sensation must simultaneously record neural signals and deliver cortical stimulation. This poses a challenge since stimulation artifacts can be orders of magnitude stronger than neural signals. In this article, we propose a novel subspace-based method for the removal of cortical electrical stimulation artifacts. We demonstrate the practical application of our approach on experimentally recorded electroencephalogram data, where artifacts were suppressed by as much as 30–40 dB. Our method is computationally simple, yet it achieves superior results to the state-of-the art methods.**

I. Introduction

The ability of brain-computer interfaces (BCIs) to provide somatosensory feedback by electrical stimulation of the sensory cortex has been demonstrated in animal [1] and human [2] studies. Ideally, these bi-directional BCIs should simultaneously record neural signals and deliver cortical stimulation [3]. This poses a significant challenge due to stimulation voltages being orders of magnitude stronger than neural potentials. Specifically, these stimulation artifacts may severely interfere with neural signals and even drive amplifiers into a saturation regime. Current bi-directional BCIs circumvent this problem by temporally multiplexing recording and stimulation [1], [2], which may be suboptimal for applications that require continuous somatosensory feedback and decoding. This strategy also constrains the BCI decoder, which may compromise the overall BCI function.

An alternative approach to this problem is to optimize the parameters critical for artifact propagation and avoid the saturation of the BCI's analog front-end. Based on volume conduction theory, electrical stimulation is expected to produce dipole-like potential distributions. Thus, the relative distance and orientation of stimulating and recording electrodes, the tissue conductivity, and the choice of a reference electrode significantly affect artifact propagation [3]. Factors such as the stimulation frequency, amplitude and waveform can also be optimized. Assuming amplifiers are not saturated, an additional level of artifact suppression can be achieved by utilizing array signal processing techniques [4]. Typically, these methods decompose measurements into a signal and noise/interference subspace, followed by the suppression of unwanted components. In the context of neural signal processing, independent component analysis (ICA) is the most popular representative of these techniques [5].

Work supported by the National Science Foundation (Award #1446908)

[1]Department of Biomedical Engineering, University of California Irvine (UCI), Irvine, CA 92697, USA {ptwang,znenadic}@uci.edu

[2]Medical Science Training Program, UCI cmccrimm@uci.edu

[3]Department of Electrical Engineering and Computer Science, UCI payam@uci.edu

[4]Department of Neurology, UCI and@uci.edu

In this article we propose a straightforward method for the removal of stimulation artifacts from recorded neural data. Similar to other subspace-based techniques, our method yields a spatial filter that can be trained offline and implemented efficiently in real time. Our approach takes advantage of stimulation artifacts being much stronger than neural signals. It also exploits strong spatial correlations that are generally present in neural signals measured by an array of sensors [6]. The performance of our method was tested on experimentally collected electroencephalogram (EEG) data, recorded in the presence of a strong electrical interference. Our method produces results that are competitive, if not superior, to the state-of-the-art ICA-based artifact removal techniques, yet at a significantly lower computational cost and with a much simpler implementation.

II. Methods

A. Null Projection for Artifact Suppression

In the presence of stimulation dipoles, the measurements from an array of n sensors can be modeled as:

$$\mathbf{X} = \mathbf{A}_D\mathbf{S}_D + \mathbf{A}_S\mathbf{S}_S + \mathbf{N} \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{n\times t}$, t is the number of time samples ($t \gg n$), $\mathbf{S}_D \in \mathbb{R}^{d\times t}$ are the time-dependent moment magnitudes of d equivalent dipoles ($d < n$), $\mathbf{S}_S \in \mathbb{R}^{s\times t}$ are the activities of s neural sources, and $\mathbf{N} \in \mathbb{R}^{n\times t}$ is background noise [7]. The columns of $\mathbf{A}_D \in \mathbb{R}^{n\times d}$ and $\mathbf{A}_S \in \mathbb{R}^{n\times s}$ are the lead field vectors (LFVs) of the stimulation dipoles and neural sources, respectively. In the BCI context, $\mathbf{S}_S$ represents task-related neural activity, whereas $\mathbf{N}$ subsumes background neural activity, biological artifacts, and electronic/sensor noise [8].

If stimulation dipoles are much stronger than neural sources and noise in (1), most of the energy in $\mathbf{X}$ will be confined to a d-dimensional subspace spanned by the LFVs in $\mathbf{A}_D$. The stimulation artifacts can then be removed by a null projection (NP) procedure [7], whereby $\mathbf{X}$ is projected to the orthogonal complement of this dipole subspace. While reasonable, this strategy may produce suboptimal results due to the stimulation-free response $\mathbf{A}_S\mathbf{S}_S+\mathbf{N}$ exhibiting strong spatial correlations. These correlations, primarily caused by the physical proximity of individual sensors and correlated background, may hinder the accurate identification of the dipole subspace.

To address the above challenge, we employ a dual-condition experimental design [9]. Specifically, in addition to measurements in the presence of a stimulator, referred to as active state data, we collect data while the stimulator is

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

turned off (control state). Expressed mathematically:

$$\mathbf{X}_A = \mathbf{A}_D\mathbf{S}_D + \mathbf{A}_S\mathbf{S}_S + \mathbf{N}_A \quad (2)$$

$$\mathbf{X}_C = \mathbf{A}_S\mathbf{S}_S + \mathbf{N}_C \quad (3)$$

where the statistical properties of noise in the active and control state are not necessarily the same.

To improve the signal-to-noise ratio (SNR) and accuracy of a dipole subspace estimate [10], we calculate the prewhitening matrix from the control state data: $\mathbf{\Sigma}_C^{-\frac{1}{2}} = \mathbf{V}_C^{\mathrm{T}}\mathbf{\Lambda}_C^{-\frac{1}{2}}\mathbf{V}_C$. Here, $\mathbf{\Sigma}_C \in \mathbb{R}^{n\times n}$ is the covariance of $\mathbf{X}_C$, and $\mathbf{V}_C \in \mathbb{R}^{n\times n}$ and $\mathbf{\Lambda}_C \in \mathbb{R}^{n\times n}$ are its eigenvector and eigenvalue matrix, respectively. Note that $\mathbf{X}_C$ is generally a full-rank matrix, and so $\mathbf{\Sigma}_C^{-\frac{1}{2}}$ is well defined. Subsequently, we de-mean and prewhiten the active state data:

$$\mathbf{X}'_A = \mathbf{\Sigma}_C^{-\frac{1}{2}}\left(\mathbf{X}_A - \mu_A\mathbf{1}^{\mathrm{T}}\right) \quad (4)$$

where $\mu_A = \frac{1}{t_A}\sum_{i=1}^{t_A}\mathbf{X}_A(i) \in \mathbb{R}^n$, t_A is the number of samples in the active state, and the elements of $\mathbf{1} \in \mathbb{R}^{t_A}$ are all 1. Since this operation nearly whitens the stimulation-free response in (2), the identification of the dipole subspace is readily achieved using the singular value decomposition [7]:

$$\mathbf{X}'_A = \mathbf{U}_A\mathbf{\Sigma}_A\mathbf{V}_A^{\mathrm{T}} = \begin{bmatrix} \mathbf{U}_d & \mathbf{U}_d^c \end{bmatrix}\begin{bmatrix} \mathbf{\Sigma}_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_d^c & \mathbf{0} \end{bmatrix}\mathbf{V}_A^{\mathrm{T}}$$

The dipole subspace is spanned by the columns of $\mathbf{U}_d \in \mathbb{R}^{n\times d}$, which are the left singular vectors of $\mathbf{X}'_A$ corresponding to its largest d singular values (the diagonal of $\mathbf{\Sigma}_d$). Its orthogonal complement is spanned by the remaining $n-d$ left singular eigenvectors (columns of $\mathbf{U}_d^c$). When $\mathbf{X}'_A$ is projected onto this subspace, the effect of the stimulation dipole will be minimized. Formally, let this null projection be represented by a matrix $\mathbf{H} = \mathbf{U}_d^c$, then the projected data $\mathbf{H}^{\mathrm{T}}\mathbf{X}'_A$ will be largely free of stimulation artifacts. Reconstructing this projection in the original space followed by "coloring" and restoring the mean value yields:

$$\mathbf{X}_A^{\text{clean}} = \mathbf{\Sigma}_C^{\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{X}'_A + \mu_A\mathbf{1}^{\mathrm{T}} \quad (5)$$

The coloring matrix, $\mathbf{\Sigma}_C^{\frac{1}{2}}$, is simply the inverse of the prewhitening matrix. We will refer to this method as prewhitening plus null projection (PW+NP) algorithm.

B. Experimental Data Collection

The study was approved by the Institutional Review Board at the University of California Irvine. An able-body subject was fitted with a 20-electrode EEG cap (Compumedics USA, Charlotte, NC). Conductive gel was applied and the scalp was abraded until the 30-Hz impedances were lowered to $<$10 kΩ. The signals were referenced to a frontal electrode (see Fig. 1), and were amplified ($\times$5,000) and band-pass filtered (1–35 Hz) using 19 single-channel bioamplifiers (EEG100C, Biopac Systems, Goleta, CA). The signals were digitized (4,000 Hz, 16 bits) and recorded by a data acquisition system (MP150, Biopac Systems). The system's auxiliary analog inputs were used to synchronize the recorded EEG and experimental cues, controlled by custom-made Matlab (Mathworks, Natick, MA) scripts.

Prior to mounting the EEG cap, three individual electrodes were affixed to the scalp with an adhesive cream (EC2, Natus Neurology, Middleton, WI), in the vicinity of the central-midline electrode Cz (Fig. 1). They were connected (two at a time) to a hand-held, battery-powered impedance monitor (EIM105, General Devices, Ridgefield, NJ). The monitor produces a 30-Hz sinusoidal artifact that is much stronger than EEG, and therefore acts like a surrogate for a cortical stimulator. The electrodes were connected to mimic both "horizontal" and "vertical" dipole configurations. The stimulator signal was split by a cable, and its copy was acquired by the MP150 system at the same rate as EEG.

Auditory cues instructed the subject to alternate between keeping his eyes open and closed for 5 min total. The duration of each eyes-open and eyes-closed epoch was 15 sec (20 epochs total). This experiment was performed under the no stimulation, horizontal dipole, and vertical dipole conditions. All the data were saved for subsequent analyses.

C. Analysis

1) PW + NP: The no-stimulation EEG data were used as a control state and the prewhitening matrix was calculated as explained in Section II-A. The EEG data from the horizontal-dipole condition (active state) were then processed according to (4–5). In the presence of a single stimulator, the theoretical dimension of the dipole subspace is $d = 1$. Practically, d may be higher due imperfections in the model (2). To determine d, we note that prewhitening renders the non-dipole components of $\mathbf{X}'_A$ nearly white, and so the smallest $n-d$ eigenvalues of its covariance matrix are $\approx$1. Utilizing the connection between singular values and eigenvalues, it follows that the smallest $n-d$ singular values of $\mathbf{X}'_A$ satisfy $\sigma \approx \sqrt{t_A - 1}$. To account for noise in the singular value distribution, we determined d by counting the number of singular values that satisfy $\sigma > 2\sqrt{t_A - 1}$. More elaborate techniques for determining d can be found in [11]. The null projection matrix $\mathbf{H}$ and $\mathbf{\Sigma}_C^{\frac{1}{2}}$ were then calculated as explained in Section II-A, and the artifact-free data were obtained by (5). The above procedure was then repeated for the vertical-dipole data.

2) Performance Metrics: The quality of artifact suppression algorithms (PW+NP and ICA) was quantified by measuring the SNR and signal-to-interference ratio (SIR). The SNR was defined as a deflection coefficient [12]:

$$\text{SNR}(f) = \sqrt{\frac{(\mu_{\text{c}}(f) - \mu_{\text{o}}(f))^2}{0.5(\sigma_{\text{c}}^2(f) + \sigma_{\text{o}}^2(f))}}, \quad f \in [8,\, 12]\,\text{Hz}$$

where $\mu_{\text{c}}(f)$ and $\mu_{\text{o}}(f)$ are the power spectral densities (PSDs) of EEG data averaged across the eyes-closed and eyes-open epochs, respectively, and $\sigma_{\text{c}}(f)$ and $\sigma_{\text{o}}(f)$ are their corresponding standard deviations. The 8–12 Hz band was chosen since closing the eyes is known to increase the power of the occipital α wave over these frequencies [13]. Thus, we formally designate EEG under eyes-closed and eyes-open conditions as signal and noise, respectively. The overall SNR was calculated by averaging SNR(f) over the 8–12 Hz band.

The SIR was defined as:

$$\text{SIR} = \frac{\max\limits_{8\le f\le 12} \mu_c(f)}{\max\limits_{29\le f\le 31} \mu_c(f)} \tag{6}$$

with interference being a narrow-band signal around 30 Hz due to the stimulator's sinusoidal waveform. The SIR and SNR were calculated on a per channel basis. For visualization purposes they were expressed in decibels ($10\log_{10}$).

3) ICA: To compare the performance of the PW+NP algorithm to the state-of-the-art, the FastICA algorithm [14] was employed. Matlab scripts were downloaded from the FastICA web site, and the algorithm was executed with default parameters. The independent components (ICs) corresponding to the horizontal and vertical dipole data were calculated and stored for subsequent analysis, together with the mixing and unmixing matrices. The ICs were visualized in order identify the presence of stimulation artifacts, and this process was aided by evaluating their SIRs. Independent components deemed to contain the stimulation artifacts were then discarded by the backprojection method [5].

III. Results and Discussion

The impedance meter produced strong artifacts, primarily around 30 Hz. Fig. 1 shows the contour-plot distribution of EEG data peaks calculated for both the horizontal and vertical dipole conditions. Briefly, the positive peaks of the stimulator signal were detected and multi-channel EEG data were extracted at these time points. The contours were then generated by averaging these peak values over ∼9,000 stimulation cycles. The artifacts exhibited a dipole-like distribution, and were an order of magnitude stronger than the α wave. Stronger artifacts could have been generated, however the impedances of the stimulating electrodes were kept intentionally high in order to prevent the saturation of bioamplifiers. The mismatch between the location of the stimulating electrodes and the maximum/minimum contours is caused by the low channel density and spatial interpolation artifacts.

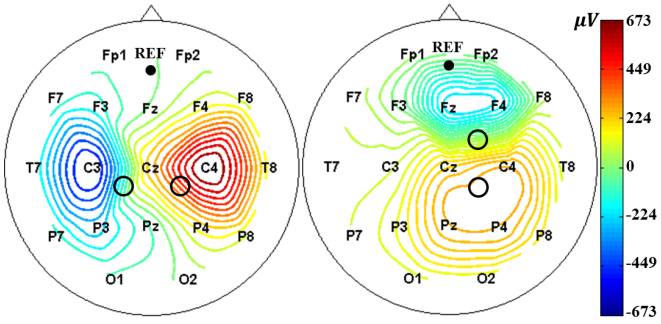

Fig. 1. The spatial distribution of stimulation artifacts represented by contours. Left: horizontal dipole. Right: vertical dipole. The approximate locations of the stimulating electrodes are marked by the open circles, and the location of the reference electrode is shown as the black dot.

Additional analysis showed that the signals on different channels did not peak at the same time, and that these phase shifts, calculated with respect to the stimulator signal, varied between 0.2 and 8.9 msec across the channels and experimental conditions. These phase differences were likely caused by the impedances at the electrode-scalp interface.

The estimated dipole subspace dimension was $\hat{d} = 3$ for both the horizontal and vertical dipole data. These estimates were corroborated by decomposing the active state data via $\mathbf{C} = \mathbf{U}_A^{\mathrm{T}}\mathbf{X}'_A \in \mathbb{R}^{n\times t_A}$, and visualizing the PSDs of these components (the rows of $\mathbf{C}$). The top three components exhibited a significant peak at 30 Hz, suggesting that they resided in the dipole subspace, $\mathbf{U}_d$. The remaining components had a much smaller peak at this frequency. The departure of $\hat{d}$ from the theoretical value, $d = 1$, is consistent with the observed phase shifts. Since a single LFV in the model (2) cannot account for phase differences, the rank of $\mathbf{A}_D$ is inflated and phase-shifted artifacts are treated as additional dipoles. Upon cleaning the data according to (5), the SNR and SIR values were calculated. The results are shown on the top of Figs. 2 and 3 for the horizontal and vertical dipole data, respectively. The original SIRs and SNRs (prior to processing) are also shown.

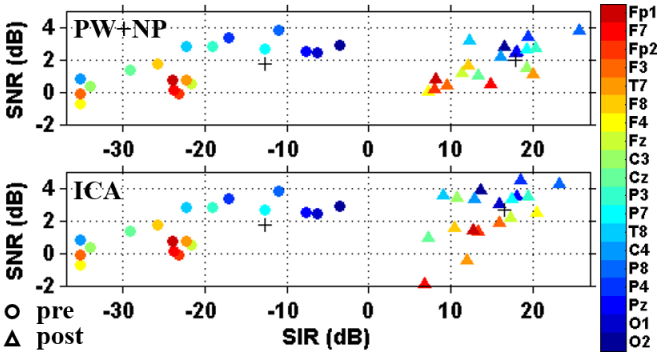

Fig. 2. SNRs and SIRs across channels (color coded) before (○) and after (△) processing the horizontal-dipole data. Top: PW+NP. Bottom: ICA. The crosses are the means as defined in Table I.

Upon execution of ICA on the horizontal and vertical dipole data, 16 and 17 ICs converged respectively. The ICs were analyzed and those that showed a dominant peak at 30 Hz were identified (IC 8 for horizontal and ICs 5 and 16 for vertical dipole). When EEG signals were reconstructed without these ICs, the calculated SIRs were not competitive to those obtained by PW+NP, especially for the vertical dipole data (compare the 2nd and 3rd column of Table I). The ICs were then sorted according to their SIRs, and a combinatorial search was performed by discarding low-SIR components until the reconstructed EEG achieved optimal SIR values. This resulted in the exclusion of an additional IC for the horizontal dipole data, and 3 additional ICs for the vertical dipole data. These values are shown on the bottom of Figs. 2 and 3, and their summary statistics are denoted by ICA_{opt} in Table I.

Figs. 2 and 3 show that both PW+NP and ICA substantially improved the SIR without compromising SNR. In fact the mean SNR improved slightly, but these gains were not statistically significant. On the other hand, the gains in

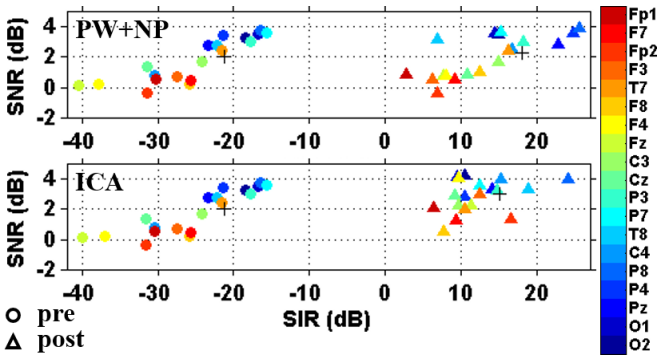

Fig. 3. Equivalent of Fig. 2 for the vertical dipole data.

SIR were statistically significant for both PW+NP and ICA and for both dipole conditions (Kolmogorov-Smirnov test, $p<10^{-8}$). The figures also show that PW+NP outperformed ICA in rejecting the artifacts, especially under the vertical-dipole condition. While these gains were not statistically significant due to an excessively large variance, these differences were not negligible. For the horizontal dipole, the mean SIR after PW+NP was 1.31 dB (35%) higher than that of ICA. For the vertical dipole, the PW+NP SIR was 2.96 dB (98%) higher than its ICA counterpart. With respect to the original data, PW+NP achieved overall SIR gains of 30.57 dB (horizontal dipole) and 39.29 dB (vertical dipole).

TABLE I

THE MEAN SNR AND SIR FOR THE HORIZONTAL (TOP) AND VERTICAL (BOTTOM) DIPOLE DATA. THE STATISTICS WERE CALCULATED OVER CHANNELS AND TRANSFORMED INTO DECIBELS BY $10\log_{10}$. THE NUMBERS IN PARENTHESES SHOW THE EXCLUDED ICS.

Horizontal		Original	PW+NP	ICA (8)	ICA_{opt}
SIR dB	Mean	-12.63	17.94	15.90	16.63
SNR dB	Mean	1.74	1.96	2.69	2.70
Vertical		Original	PW+NP	ICA (5,16)	ICA_{opt}
SIR dB	Mean	-21.28	18.01	10.42	15.05
SNR dB	Mean	2.03	2.27	2.80	3.00

The advantages of our method are several-fold. First, the stimulation dipole identification is straightforward with theoretically justified thresholds (see Section II-C). To be competitive, ICA required a combinatorial search over its low-SIR components. For generalizable results, this dipole identification must be implemented within a cross-validation framework, which is computationally demanding in ICA, but very simple in PW+NP using standard linear algebra tools. In addition, our method produces superior artifact suppression results. This is not surprising since phase-shifted artifacts cannot be captured by a single IC, yet they are not likely to be fully independent. The PW+NP method handles this problem by inflating the dipole subspace. On the other hand, ICA must fit these residual artifacts into other ICs, thus remixing them with other sources, including neural ones. PW+NP requires additional data (control state), but this is not a problem since the stimulator can easily be turned off. When we ran ICA on the concatenation of control and active state data, its performance did not improve. Thus, the advantage of PW+NP cannot be attributed to it having more data at its disposal.

With respect to our method, ICA achieved slightly superior SNRs, although these gains were not statistically significant. However, the SNR was evaluated to merely demonstrate that artifact suppression did not compromise the signal quality. To maximize the SNR, a more appropriate supervised learning techniques could be used [15].

IV. CONCLUSION

Our study demonstrates that a straightforward array signal processing method can achieve 30–40 dB of artifact suppression. In combination with artifact mitigation strategies at the front-end, this method could yield usable neural signals and, in turn, enable simultaneous recording and stimulation in bi-directional BCIs. Our future plans include testing of the PW+NP algorithm on additional subjects as well as with other types of neural data.

REFERENCES

[1] J. E. O'Doherty, M. A. Lebedev, P. J. Ifft, K. Z. Zhuang, S. Shokur, H. Bleuler, and M. A. L. Nicolelis, "Active tactile exploration using a brain-machine-brain interface," *Nature*, 479(7372):228-231, 2011.

[2] S. Flesher, J. Downey, J. Collinger, S. Foldes, J. Weiss, E. Tyler-Kabara, S. Bensmaia, A. Schwartz, M. Boninger, and R. Gaunt, "Intracortical Microstimulation as a Feedback Source for Brain-Computer Interface Users," In *Brain-Computer Interface Research*, pp. 43-54. Springer, Cham, 2017.

[3] A. G. Rouse, S. R. Stanslaski, P. Cong, R. M. Jensen, P. Afshar, D. Ullestad, R. Gupta, G. F. Molnar, D. W. Moran, and T. J. Denison, "A chronic generalized bi-directional brain-machine interface," *J. Neural Eng.*, 8(3):036018, 2011.

[4] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Proc. Mag.*, 13(4):67-94, 1996.

[5] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects, *Clin. Neurophysiol.*, 111:1745-1758, 2000.

[6] E. J. Hwang and R. A. Andersen, The utility of multichannel local field potentials for brain-machine interfaces, *J. Neural Eng.*, 10(4):046005, 2013.

[7] S.-C. Wu, A. L. Swindlehurst, P. T. Wang, and Z. Nenadic, "Projection vs. prewhitening for EEG interference suppression," *IEEE T. Bio-med. Eng.*, 59(5):1329-1338, 2012.

[8] S.-C. Wu, A. L. Swindlehurst, P. T. Wang, and Z. Nenadic, "Efficient dipole parameter estimation in EEG systems with near-ML performance," *IEEE T. Bio-med. Eng.*, 59(5):1339-1348, 2012.

[9] J. J. Ermer, J. C. Mosher, M. Huang, and R. M. Leahy, "Paired MEG data set source localization using recursively applied and projected (RAP) MUSIC," *IEEE T. Bio-med. Eng.*, 47(9):1248-1260, 2000.

[10] J. P. C. L. Da Costa, K. Liu, H. C. So, S. Schwarz, M. Haardt, and F. Römer, "Multidimensional prewhitening for enhanced signal reconstruction and parameter estimation in colored noise with Kronecker correlation structure," *Signal Process.*, 93(11):3209-3226, 2013.

[11] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE T. Acoust. Speech*, 33(2):387392, 1985.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing. Detection Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1998.

[13] D. Lehmann, Multichannel topography of human alpha EEG fields," *Electroen. Clin. Neuro.*, 31(5):439-449, 1971.

[14] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE T. Neural Networ.*, 10(3):626-634, 1999.

[15] Z. Nenadic, "Information discriminant analysis: Feature extraction with an information-theoretic objective," *IEEE T. Pattern Anal.*, 29(8): 1394-1407, 2007.