A HIERARCHICAL BAYESIAN M/EEG IMAGING METHOD CORRECTING FOR INCOMPLETE SPATIO-TEMPORAL PRIORS

Carsten Stahlhut[a], Hagai T. Attias[b], Kensuke Sekihara[c], David Wipf[d]
Lars K. Hansen[a], Srikantan S. Nagarajan[e]

[a] Technical University of Denmark, DTU Compute, Denmark
[b]Convex Imaging, San Francisco, CA, USA
[c]Tokyo Metropolian University, Dept. of System Design & Engineering, Japan
[d]Visual Computing Group, Microsoft Research Asia, China
[e]UCSF, Biomagnetic Imaging Laboratory, San Francisco, CA, USA

ABSTRACT

In this paper we present a hierarchical Bayesian model, to tackle the highly ill-posed problem that follows with MEG and EEG source imaging. Our model promotes spatio-temporal patterns through the use of both spatial and temporal basis functions. While in contrast to most previous spatio-temporal inverse M/EEG models, the proposed model benefits of consisting of two source terms, namely, a spatio-temporal pattern term limiting the source configuration to a spatio-temporal subspace and a source correcting term to pick up source activity not covered by the spatio-temporal prior belief. Both artificial data and real EEG data is used to demonstrate the efficacy of the model.

Index Terms— MEG, EEG, inverse problem, variational Bayes, spatio-temporal prior

1. INTRODUCTION

While functional Magnetic Resonance Imaging (fMRI) has made profound contribution to the neuroscience due to its high spatial resolution brain images even though it has relative low temporal resolution compared to functional imaging approaches such as magnetoencephalography (MEG) and electroencephaloghy (EEG) offering temporal resolution less than 0.1s. However, the nature of M/EEG only allows an inverse problem to be tackled, which in the current state-of-the-art consequently has meant spatial specificity lacking behind. To obtain reasonable spatial resolution, the number of dipoles (source locations) is typically in the order of $N_d = 5,000 - 80,000$ dipoles, which outnumber the number of channels, $N_c = 32 - 300$, significantly. Consequently, the inverse problem, estimating the current sources, $\mathbf{S}$, given the observed M/EEG signal, $\mathbf{Y}$, is severely ill-posed and all efforts at source reconstruction are heavily dependent on prior assumptions. The relation between the measured M/EEG signal and the brain's current sources can be expressed as a linear instantaneous form in the sources. The forward relation can be written as

$$\mathbf{Y} = \mathbf{F}\mathbf{S} + \boldsymbol{\mathcal{E}}, \tag{1}$$

where $\boldsymbol{\mathcal{E}}$ is assumed additive noise and the observed M/EEG signal is denoted $\mathbf{Y} \in \Re^{J\times N}$, the current sources $\mathbf{S} \in \Re^{V\times N}$, and with J, V, and N being the number of sensors, voxels/dipoles, and time points, respectively. The lead field/forward model $\mathbf{F} \in \Re^{J\times V}$ express the relationsship between the sensor space and the source space. Several approaches exist for computing $\mathbf{F}$ based on Maxwell's equations and physical properties of the brain such as geometry and tissue conductivity values.

Besides having to deal with the ill-posed nature of M/EEG imaging, the inverse algorithms have to address the challenge of searching for true source signals while minimizing the many sources of noise that interfere with the true signals. Electrical, thermal and biological noise as well as background room interference can be present. To overcome this both spatial and temporal assumptions have been pursued by several authors [1]–[6].

In this contribution we also pursue spatio-temporal patterns, in order to facilitate neurophysiological plausiable source configurations of networks of brain regions being activate together. Here, we embody spatio-temporal patterns using probability distributions over the sources. In contrast to most related work using spatio-temporal basis functions we seek source configurations that are likely to have support in the spatio-temporal space spanned by the basis functions, however, importantly not limited to this space. Whether or not the sources are within the subspace given by spatial and temporal basis functions is purely determined by the data itself. In contrast to [6] we propose a hierarchical Bayesian model, using ideas from probabilistic graphical modeling, which reflects uncertainty by allowing us to compute a tractable posterior distribution on the unknown sources.

This work was supported by Lundbeck foundation and Danish Council for Independent Research (CS) and R01 DC004855 / R21 NS76171 (SSN).

2. METHODS

To solve the ill-posed M/EEG source imaging problem we propose a hierarchical Bayesian spatio-temporal model, with the current sources $\mathbf{S}$ decomposed into two source parts, $\mathbf{S} = \mathbf{\Psi W \Phi} + \mathbf{V}$. The first source part is a spatio-temporal contribution expressed in the spatio-temporal maps, $\mathbf{W}$, whereas the latter is the spatio-temporal source correction part (or source 'noise' contribution) often regarded as independent identical distributed (iid), in methods such as the Champagne algorithm [7]. Thus, the Aquavit model spells

$$\mathbf{Y} = \mathbf{F}\left(\mathbf{\Psi W \Phi} + \mathbf{V}\right) + \boldsymbol{\mathcal{E}} = \mathbf{G}\sum_k \mathbf{w}_k \boldsymbol{\phi}_k^T + \mathbf{FV} + \boldsymbol{\mathcal{E}} \quad (2)$$

with $\mathbf{w}_k$ being the k'th column of the spatio-temporal maps $\mathbf{W}$ and $\mathbf{G} = \mathbf{F\Psi}$ representing a forward model mapping spatial regions (groups of dipoles/voxels clustered by the spatial basis functions $\mathbf{\Psi}$) of interest to the sensors. Given J sensors, V voxels/dipoles, R spatial basis functions (regions), K temporal basis functions, and N time points, the spatio-temporal maps $\mathbf{W}$, the forward model matrices $\mathbf{G}$ and $\mathbf{F}$, are $R \times K$, $J \times R$ and $J \times V$, respectively. In the following we will make use of the notation $\mathbf{\Phi} = [\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_N] = [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_K]^T$ interchangeably, with $\boldsymbol{\varphi}_n$ and $\boldsymbol{\phi}_k$ representing all temporal basis functions at time point n and the k'th temporal basis function, respectively. The model distribution is given by

$$\begin{aligned} p\left(\mathbf{Y}\,|\mathbf{W}, \mathbf{V}\right) &= \prod_{n=1} \mathcal{N}\left(\mathbf{y}_n \,|\mathbf{GW}\boldsymbol{\varphi}_n + \mathbf{F}\mathbf{v}_n, \mathbf{\Sigma}_\varepsilon\right) \\ p\left(\mathbf{W}\right) &= \prod_{k=1} \mathcal{N}\left(\mathbf{w}_{k\cdot} \,\middle|\mathbf{0}, \mathbf{A}^{-1}\right) \\ p\left(\mathbf{V}\right) &= \prod_{n=1} \mathcal{N}\left(\mathbf{v}_n \,\middle|\mathbf{0}, \mathbf{\Gamma}^{-1}\right), \end{aligned} \quad (3)$$

with $\mathbf{\Sigma}_\varepsilon$ denoting the noise covariance matrix, $\mathbf{A} = \mathrm{diag}\left(\boldsymbol{\alpha}\right)$ as a shared precision matrix for the spatio-temporal weights among the columns of $\mathbf{W}$, and $\mathbf{\Gamma} = \mathrm{diag}\left(\boldsymbol{\gamma}\right)$ the precision matrix for the source term $\mathbf{V}$. Tuning a source's variance to zero effectively prunes out the source parameter [8]. Compared to the model in [9] our Aquavit model here includes spatial-basis functions which allows to cluster/weight set of voxels to form regions/groups. Ideally, we would like to integrate out all unknowns (including hyperparameters) and then compute the posterior over sources $p\left(\mathbf{S}|\mathbf{Y}\right) = p\left(\mathbf{V}, \mathbf{W}|\mathbf{Y}\right)$, which contains all possible information about $\mathbf{S}$ conditioned on the observed data $\mathbf{Y}$. However, the exact posterior in our model is computationally intractable. Here we choose to infer the hyperparameters using the maximum a-posteriori (MAP) point estimate. Given this, the joint posterior over the source 'noise' and STE maps $p\left(\mathbf{V}, \mathbf{W}|\,\mathbf{Y}\right)$ is Gaussian, and can be computed analytically. Nevertheless, here we apply a variational Bayesian (VB) [10] approximation $p\left(\mathbf{V}, \mathbf{W}|\,\mathbf{Y}\right) \approx q\left(\mathbf{V}, \mathbf{W}\right)$ to speed up the computation. In this contribution we seek computational efficient approaches to calculate the posterior distribution, thus we impose a factorization in the columns of $\mathbf{W}$, $q\left(\mathbf{W}\right) = \prod_k q\left(\mathbf{w}_k\right)$. Note it can be shown that indeed such a factorization is natural and is actually only required when the temporal basis functions are non-orthonormal. Given orthonormal temporal basis functions the full variational posterior distribution of $\mathbf{W}$ factorizes in the columns of $\mathbf{W}$. Importantly, we could skip the column factorization and use $q\left(\mathbf{W}\right)$, which would still leave the posterior tractable but more expensive to compute especially for large sets of spatial and temporal basis functions. Other than the factorial assumption $q\left(\mathbf{W}, \mathbf{V}\right) = q\left(\mathbf{V}\right)\prod_k q\left(\mathbf{w}_k\right)$, we place no constraints on this distribution. We then iteratively maximize the objective function $\mathcal{F}$ with respect to $q\left(\mathbf{W}\right)$, $q\left(\mathbf{V}\right)$, and any hyperparameters (i.e., $\mathbf{A}$, $\mathbf{\Gamma}$) we choose to update. The objective function $\mathcal{F}$ is the lower bound of the log marginal likehood

$$\mathcal{F} = \left\langle \log p\left(\mathbf{Y}, \mathbf{W}, \mathbf{V}\right) - \log q\left(\mathbf{W}, \mathbf{V}\right)\right\rangle_{q(\mathbf{W})q(\mathbf{V})} \leq \mathcal{L}. \quad (4)$$

Upon convergence, it can be shown that $q\left(\mathbf{V}\right)$ becomes a principled Gaussian approximation to $p\left(\mathbf{V}|\mathbf{Y}\right)$ with analytically computable moments.

Optimization of variational distributions: In turn each of the marginal posterior distributions can be maximized as follows

$$\begin{aligned} q\left(\mathbf{V}\right) &\propto \mathrm{e}^{\left\langle \log p(\mathbf{Y},\mathbf{W},\mathbf{V})\right\rangle_{q(\mathbf{W})}} = \prod_n \mathcal{N}\left(\mathbf{v}_n \,|\bar{\mathbf{v}}_n, \mathbf{\Sigma}_v\right) \\ \bar{\mathbf{v}}_n &= \mathbf{\Sigma}_v \mathbf{F}^T \mathbf{\Sigma}_\varepsilon^{-1}\left(\mathbf{y}_n - \mathbf{G}\bar{\mathbf{W}}\boldsymbol{\varphi}_n\right), \mathbf{\Sigma}_v^{-1} = \mathbf{F}^T\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{F} + \mathbf{\Gamma}, \end{aligned} \quad (5)$$

which can be calculated efficiently with the use of the matrix inversion lemma. Similarly, we maximize the variational posterior distribution of $q\left(\mathbf{W}\right)$ by

$$\begin{aligned} q\left(\mathbf{W}\right) &\propto \mathrm{e}^{\left\langle \log p(\mathbf{Y},\mathbf{W},\mathbf{V})\right\rangle_{q(\mathbf{V})}} = \prod_k \mathcal{N}\left(\mathbf{w}_k \,|\bar{\mathbf{w}}_k, \mathbf{\Omega}_k\right) \\ \bar{\mathbf{w}}_k &= \mathbf{\Omega}_k \mathbf{G}^T \mathbf{\Sigma}_\varepsilon^{-1}\mathbf{X}, \mathbf{\Omega}_k^{-1} = \boldsymbol{\phi}_k^T\boldsymbol{\phi}_k \mathbf{G}^T\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{G} + \mathbf{A} \\ \mathbf{X} &= \left(\mathbf{Y} - \mathbf{F}\bar{\mathbf{V}}\right)\boldsymbol{\phi}_k - \mathbf{G}\sum_{l\neq k}\bar{\mathbf{w}}_l \boldsymbol{\phi}_l^T\boldsymbol{\phi}_k \end{aligned} \quad (6)$$

with $\mathbf{X}$ expressing the residual between the actual M/EEG observations $\mathbf{Y}$ and M/EEG data accounted for by the model terms $\mathbf{V}$ and $\mathbf{W}_{\setminus k}$ projected to the space spanned by the temporal basis function $\boldsymbol{\phi}_k$. To obtain computational efficient calculation of the posterior distribution of $q\left(\mathbf{W}\right)$ we can again invoke the matrix inversion lemma profitably in cases with the number of groups/regions, R, exceeding the number of sensors, J. We note that with $\mathbf{\Phi}$ being a set of orthonormal temporal basis functions, i.e. $\mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{I}$, the posterior distribution over the spatio-temporal maps is easier computed as this reduces to a single shared covariance matrix $\mathbf{\Omega}$ among all $\mathbf{w}_k$ and the dependent term on $\mathbf{W}_{\setminus k}$ (appearing as $\mathbf{w}_l$ in $\mathbf{X}$) vanishes. The Aquavit model has a number of interesting properties such as efficient calculation of the variational posterior distribution of the spatio-temporal maps, even for

large set of regions, which potential could be larger than the number of voxels if e.g. different degrees of spatial smoothness is desired similar to spatial basis functions used in [3]. Besides a different factorizing approach of the joint posterior distribution over the source space and the present of spatial basis functions applied in [9], the fact that Aquavit can handle potential large spatial basis is a key benefit. Yet, another interesting observation of the Aquavit is that it reduces to the Champagne model [7], when $\mathbf{W} = \mathbf{A} = 0$, meaning that in this case all reconstructed sources are expressed by the spatio-temporal source correction term, $\mathbf{V}$.

Optimization of hyperparameters: In accordance with VB framework we should alternate between an expectation (VB-E) step and a maximization (VB-M) step. While the VB-E step includes calcuation of the sufficient statistics used by the varatiational posterior distributions, the VB-M step covers the optimization of the hyperparameters, $\mathbf{A}$ and $\mathbf{\Gamma}$. Standard VB expected maximization prusuits hyperparameter maximization through the derivative of the complete data log likelihood w.r.t. the hyperparameters. However, taking convergency of the VB-EM into consideration this can be extremely slow in practice when V is large. In order to gain high convergence speed techniques such as [11] have demonstrated its applicability. Yet, these updates often provide convergence speed-up through fix-point iterations, it does not provide any guarantee of an increasing variational bound. Thus, we here focus on an alternative approach which make use of convex bounding techniques [7] to ensure that the variational bound increases in every step. Starting from the variation bound $\mathcal{F}$, which can be shown to be

$$\mathcal{F} = \log p\left(\mathbf{Y}\left|\mathbf{W}=\bar{\mathbf{W}},\mathbf{V}=\bar{\mathbf{V}}\right.\right) + \log p\left(\mathbf{W}=\bar{\mathbf{W}}\right) + \log p\left(\mathbf{V}=\bar{\mathbf{V}}\right) - \frac{1}{2}\sum_k \log\left|\frac{\mathbf{\Omega}_k^{-1}}{2\pi}\right| - \frac{N}{2}\log\left|\frac{\mathbf{\Sigma}_v^{-1}}{2\pi}\right| \quad (7)$$

we next define two functions

$$h\left(\boldsymbol{\alpha}\right) = K\log|\mathbf{A}| - \sum_k \log\left|\mathbf{\Omega}_k^{-1}\right| \quad (8)$$

$$f\left(\boldsymbol{\gamma}\right) = \log|\mathbf{\Gamma}| - \log\left|\mathbf{\Sigma}^{-1}\right| \quad (9)$$

Since h and f are convex in $\mathbf{A}^{-1}$ and $\mathbf{\Gamma}^{-1}$, respectively, we can use the bounds

$$h \geq \boldsymbol{\zeta}^T\boldsymbol{\alpha}^{-1} + \tilde{h}\left(\boldsymbol{\zeta}\right), \quad f \geq \boldsymbol{\rho}^T\boldsymbol{\gamma}^{-1} + \tilde{f}\left(\boldsymbol{\rho}\right) \quad (10)$$

with $\boldsymbol{\zeta}$ and $\boldsymbol{\rho}$ as auxiliary variables, and $\tilde{h}$, $\tilde{f}$ the duals whose functional forms are not needed here. By substituting (10) into $\mathcal{F}$ and setting the gradients $\frac{\partial\mathcal{F}}{\partial\boldsymbol{\alpha}^{-1}} = \frac{\partial\mathcal{F}}{\partial\boldsymbol{\gamma}^{-1}} = 0$, we obtain

$$\alpha_i^{-2} = -\frac{1}{\zeta_i}\sum_k \bar{w}_{ik}^2, \quad \gamma_i^{-2} = -\frac{1}{\rho_i N}\sum_n \bar{v}_{in}^2. \quad (11)$$

Finally, we have $\boldsymbol{\zeta} = \frac{\partial h}{\partial\boldsymbol{\alpha}^{-1}}$ and $\boldsymbol{\rho} = \frac{\partial f}{\partial\boldsymbol{\gamma}^{-1}}$,

$$\zeta_i = -\sum_k \mathbf{g}_i^T\left(\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T + \frac{1}{\boldsymbol{\phi}_k^T\boldsymbol{\phi}_k}\mathbf{\Sigma}_\varepsilon\right)^{-1}\mathbf{g}_i \quad (12)$$

$$\rho_i = -\mathbf{f}_i^T\left(\mathbf{F}\mathbf{\Gamma}^{-1}\mathbf{F}^T + \mathbf{\Sigma}_\varepsilon\right)^{-1}\mathbf{f}_i, \quad (13)$$

which completes the specification of the Aquavit algorithm.

3. EXPERIMENTS

To demonstrate the efficacy of the of Aquavit we test the model on both simulated and real EEG data. We benchmark our model against state-of-the-art algorithms such as Bolstad et al.'s spatio-temporal event model (here denoted STE) [6] and Wipf et al.'s Champagne algorithm [7]. As forward model we use a BEM obtained from OpenMEEG toolbox [12], with a cortical resolution of 8,124 vertices. Of spatial basis functions we apply anatomical regions corresponding to the Brodmann's areas.

Simulations M/EEG recordings are known to consist of both transient and oscillatory behaviour, thus, we follow the [13] approach of simulating both these patterns. Similar as [13] we examine temporal decompostion obtained through stationary wavelet transform with Symlets wavelets as our temporal basis functions. To mimic some of the challenges an inverse solver has to deal with, we select 3 patches on the cortical surface randomly as having same transient signatures. The oscillatory part of the artificial signal is made up of 20 Hz oscillations and three different anatomical regions being active simulatenously with random sampled phase difference, Fig. 1. All methods were tested at different SNRs (-10.0,-5.2, -3.0, -1.0, 0, 0.8, 3.0 dB) using white noise. At all SNRs the Aquavit performs best in AUC measure, e.g. at SNR=0dB AUC is, 0.9996 (Aquavit), 0.8451 (STE) and 0.4791 (Champagne). Fig. 1 demonstrates the root-mean-square (RMS) source activation during the transient and oscillatory period for the two best performing inverse methods based on the AUC measures.

Real EEG Data To validate the model on real EEG data we applied it to a study of face perception [14]. In this contribution we reconstruct the average event related potential (ERP) of trials involving real faces as stimuli. We demonstrate the usage of Brodmann and graph Laplacian as spatial basis set and temporal functions obtained from discrete cosine transform using SPM8. Fig. 2 illustrates the root-mean-square (RMS) source activation $\mathbf{S}$ of the post-stimulus period 0-600ms. Visual region V1 has a significant source and moreover ventral occipito-temporal appears as well a frontal responses in accordance with previous face studies, [15].

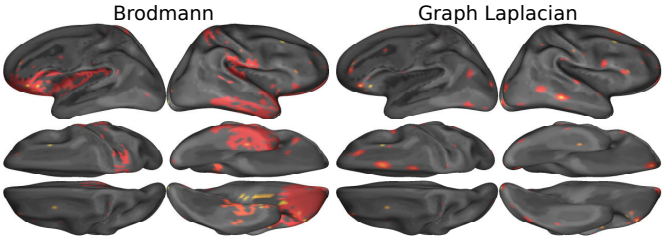

Fig. 2. Aquavit, RMS-power of poststim 0-600ms.

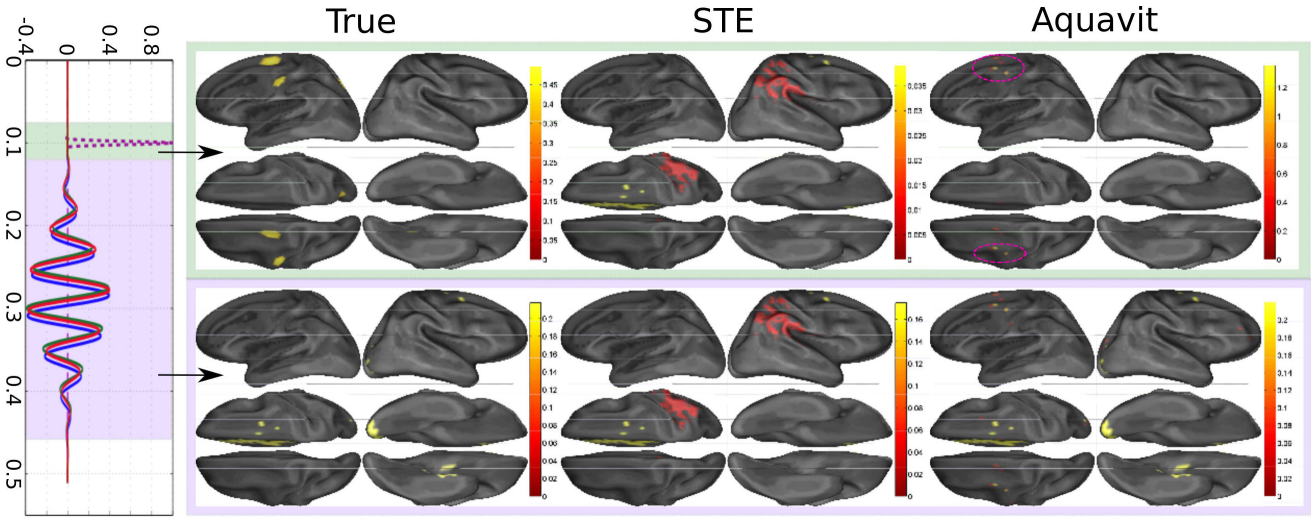

Fig. 1. Simulated sources time series and locations. Reconstructed RMS power within the window of interest.

4. CONCLUSION

We have presented a hierarchical Bayesian model facilitating spatio-temporal source configurations using spatial and temporal basis functions. In case the spatio-temporal basis functions proves inadeuate the model benefits of a source correction term being able to capture source activity outside the spatio-temporal space apriori specified in the prior. The model provides tractable variational posterior distributions, which is a significant benefit as it allows us to provide estimates of the uncertainties associated with the source estimates. Our results demonstrate that the model is capable in balancing spatio-temporal prior guidance and source correction estimation to obtain superior estimates relative to current inverse methods.

5. REFERENCES

[1] S. Baillet and L. Garnero, "A bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 374–385, 1997.

[2] C. Phillips, M.D. Rugg, and K.J. Friston, "Anatomically Informed Basis Functions for EEG Source Localisation: Combining Functional and Anatomical Constraints," *NeuroImage*, vol. 16, no. 3, pp. 678–695, 2002.

[3] S. Haufe, R. Tomioka, T. Dickhaus, C. Sannelli, B. Blankertz, G. Nolte, and K.R. Müller, "Large-scale eeg/meg source localization with spatial flexibility," *NeuroImage*, vol. 54, no. 2, pp. 851–859, 2011.

[4] M. Sahani and S. S. Nagarajan, "Reconstructing meg sources with unknown correlations," in *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[5] N.J. Trujillo-Barreto, E. Aubert-Vazquez, and W.D. Penny, "Bayesian M/EEG source reconstruction with spatio-temporal priors," *NeuroImage*, vol. 39, pp. 318–335, 2008.

[6] A. Bolstad, B. Van Veen, and R. Nowak, "Space-time event sparse penalization for magneto-/electroencephalography," *NeuroImage*, vol. 46, no. 4, pp. 1066–1081, 2009.

[7] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," *Neuroimage*, vol. 49, no. 1, pp. 641–655, 2010.

[8] L.K. Hansen and C.E. Rasmussen, "Pruning from adaptive regularization," *Neural Computation*, vol. 6, no. 6, pp. 1223–1232, 1994.

[9] C. Stahlhut, H.T. Attias, D. Wipf, L.K. Hansen, and S.S. Nagarajan, "Sparse spatio-temporal inference of electromagnetic brain sources," in *MLMI'10*, vol. 6357 of *LNCS*, pp. 157–164. Springer Berlin / Heidelberg, 2010.

[10] H. Attias, "A variational Bayesian framework for graphical models," *Advances in neural information processing systems*, vol. 12, no. 1-2, pp. 209–215, 2000.

[11] D.J.C. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.

[12] A. Gramfort, T. Papadopoulo, E. Olivi, M. Clerc, et al., "Openmeeg: opensource software for quasistatic bioelectromagnetics," *Biomed. Eng. Online*, vol. 9, no. 1, pp. 45, 2010.

[13] N. Jmail, M. Gavaret, F. Wendling, A. Kachouri, G. Hamadi, J.-M. Badier, and C.-G. Bnar, "A comparison of methods for separation of transient and oscillatory signals in eeg," *Journal of Neuroscience Methods*, vol. 199, no. 2, pp. 273 – 289, 2011.

[14] R.N.A. Henson, Y. Goshen-Gottstein, T. Ganel, L.J. Otten, A. Quayle, and M.D. Rugg, "Electrophysiological and hemodynamic correlates of face perception, recognition and priming," *Cerebral Cortex*, vol. 13, pp. 793–805, 2003.

[15] K. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. Trujillo-Barreto, R. Henson, G. Flandin, and J. Mattout, "Multiple sparse priors for the M/EEG inverse problem," *NeuroImage*, vol. 39, pp. 1104–1120, 2008.