Identification of Brain Networks using Time-Varying Spatial Constraints of Neural Activity Reconstruction

J.D. Martínez-Vargas[1] J.S. Castaño-Candamil[1†] G. Castellanos-Dominguez[1]

Abstract— **Electroencephalographic (EEG) data give a direct non-invasive measurement of neural brain activity. Nevertheless, the common assumption about EEG stationarity (time-invariant process) is a strong limitation for understanding real behavior of underlying neural networks. Here, we propose an approach for finding networks of brain regions connected by functional associations (functional connectivity) that vary along the time. To this end, we compute a set of a priori spatial dictionaries that represent brain areas with similar temporal stochastic dynamics, and then, we model relationship between areas as a time-varying process. We test our approach in both simulated and real EEG data where results show that inherent interpretability provided by the time-varying process can be useful to describe underlying neural networks.**

I. Introduction

Nowadays, importance of measuring connectivity between spatially separate, but functionally related brain areas has become of key interest in the study of human neural functions. Although most related works are based on the analysis of functional Magnetic Resonance Imaging techniques [1], some studies have shown that higher temporal resolution, provided by EEG data, allows exploring dynamics and adaptability of different cognitive processes [2], [3]. Thus, the use of EEG-based neuroimaging to identify active brain areas corresponding to resting state or responses to certain stimuli has recently received major attention. Specifically, there is a change from focusing merely on reconstructing brain activity (also known as the *EEG inverse problem*) towards modeling spatio-temporal dynamics of activation patterns, termed *functional connectivity* [4].

Functional connectivity usually comprises two stages: Firstly, the EEG signals are mapped into the source space using an inverse method; secondly, connectivity analysis is performed using predefined regions of interest [5]. About the former stage, provided that brain source reconstruction is a heavily ill-posed problem, its solution implies assumption about some prior information. Most of the state-of-art methods make physiologically meaningful assumptions to improve reconstruction accuracy. One of the physiologically fostered assumptions typically made in the EEG inverse problem solution is that the brain activity can be represented through a small/sparse set of spatial basis functions (termed *spatial blobs* or *patches*), that is, the constrained solution is a linear combination of some predefined spatial patches.

The following patch-based approaches are the most representative: Multiple Sparse Priors (MSP) [6], and Sparse Basis Field Expansion (S-FLEX) [7]. Nevertheless, assessment of brain activity reconstruction, and in turn, evaluation of connectivity networks in most of the cases is limited by the implicit assumption about spatial and temporal stationarity throughout the entire measurement interval [8]. This assumption is far from being totally realistic in many practical scenarios, where brain activity has strong non-stationary spatio-temporal dynamics [9], [8].

This work assumes that the brain activity can be represented by a set of small spatial basis functions or patches enforcing compact and sparse support. Besides, to get a physiologically plausible spatial dictionary, we introduce smooth basis, namely, Gaussian functions. Introduced smooth spatial patches also relax the assumption about EEG data non-stationarity by using time-varying prior knowledge that is introduced as a time varying a-priori covariance matrix. Generally, our method comprises of the following two stages: *i*) Computation of a locally smooth spatial dictionary where each element represents brain areas potentially generating a set of pre-identified dynamics, *ii*) Linear combination of the spatial dictionary elements, which is modeled as a time-varying process. Obtained results on simulated and real EEG databases show that interpretability provided by the time-varying process can be successfully used to encode and describe underlying neural networks.

II. Methods

A. Brain source estimation based on spatial dictionaries

Aiming to estimate brain activity, we consider the following distributed solution, $\boldsymbol{Y}=\boldsymbol{L}\boldsymbol{J}+\boldsymbol{\Xi}$, where $\boldsymbol{Y}\subset\mathbb{R}^{N_c\times N_t}$ is the EEG data measured by N_c sensors at N_t time samples, $\boldsymbol{J}\subset\mathbb{R}^{3N_d\times N_t}$ is the amplitude of the N_d current dipoles in each three-dimensional dimension distributed through cortical surface, and $\boldsymbol{L}\subset\mathbb{R}^{N_c\times 3N_d}$, termed lead field matrix, is the gain matrix representing the relationship between sources and EEG data. Besides, we assume that the EEG measures are affected by the zero mean Gaussian noise $\boldsymbol{\Xi}\subset\mathbb{R}^{N_c\times N_t}$ having the matrix covariance $\boldsymbol{Q}_{\boldsymbol{\Xi}}=\boldsymbol{I}_{N_c}$, where $\boldsymbol{I}_{N_c}\subset\mathbb{R}^{N_c\times N_c}$ is the identity matrix.

Moreover, brain activity may be represented through a linear combination of spatial basis fields (or *Sparse Basis Field Expansions*, S-FLEX), that is, the current amplitude takes the following form $\boldsymbol{J}=\boldsymbol{\Phi}_s\boldsymbol{C}_s$, where each column of

*This work is carried out under grants: Programa Nacional de Formacion de Investigadores "GENERACION DEL BICENTENARIO", 2011, from COLCIENCIAS and a MSc. scholarship funded by Universidad Nacional de Colombia. Also, it is founded by COLCIENCIAS project *Evaluación asistida de potenciales evocados cognitivos como marcador del transtorno por déficit de atensión e hiperactividad (TDAH).*

[1]Authors are working with the Universidad Nacional de Colombia, Sede Manizales. Signal Processing and Recognition Group.

† Corresponding author `jscastanoc at unal.edu.co`

the matrix $\boldsymbol{\Phi}_s \subset \mathbb{R}^{3N_d \times N_s}$ corresponds to a single cortical patch, while the matrix $\boldsymbol{C}_s \subset \mathbb{R}^{N_s \times N_t}$ holds all weighting coefficients, which are assumed to have a Laplacian prior distribution. Consequently, to obtain spatially sparse solutions, the following objective function is derived:

$$\arg\min_{\boldsymbol{C}_s} \left\{ ||\boldsymbol{Y} - \boldsymbol{L}\boldsymbol{\Phi}_s\boldsymbol{C}_s||_F^2 + \lambda||\boldsymbol{C}_s||_{1,2} \right\} \tag{1}$$

where notations $||\cdot||_F$ and $||\cdot||_{1,2}$ stand for the Frobenius and $\mathcal{L}_{1,2}$-norms. The latter one is the $\mathcal{L}_1$-norm grouping each vector dipole component under the $\mathcal{L}_2$-norm to avoid orientation bias. Further explanation of S-FLEX can be addressed in [7][1.1].

B. Time-varying functional connectivity analysis

In the following, we describe the steps for estimating the time-varying Functional Connectivity Matrix (FC-TVAR).

1) Time and spatial dynamics computation: In the beginning, N_q main temporal components are extracted from the original input data by using the first N_q eigenvectors of the temporal covariance matrix obtained from its eigen-decomposition, that is, $\boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^\top = \boldsymbol{Y}^\top\boldsymbol{Y}$, where $\boldsymbol{V} \subset \mathbb{R}^{N_t \times N_t}$ is a matrix where each column corresponds to the right eigenvectors (temporal components) of $\boldsymbol{Y}$, and $\boldsymbol{S} \subset \mathbb{R}^{N_t \times N_t}$ is the matrix holding in the main diagonal its eigenvalues.

Afterward, the introduced compact spatial dictionary, $\boldsymbol{\Phi}_s^*$, is selected from the original set $\boldsymbol{\Phi}_s$ by estimating activity generated by the above temporal decomposition explained. In particular, each element of the reduced spatial set is calculated using the S-FLEX method as follows:

$$\{\boldsymbol{\Phi}_s^*(\cdot, i) = \mathrm{S-FLEX}\left(\boldsymbol{\Phi}_s, \boldsymbol{Y}\boldsymbol{V}_{(\cdot,i)}\right) : \forall i \in N_q\}$$

where $\boldsymbol{V}_{(\cdot,i)}$ holds the i-th main temporal EEG data component and $\mathrm{S-FLEX}\left(\boldsymbol{\Phi}_s, \boldsymbol{Y}\boldsymbol{V}_{(\cdot,i)}\right)$ is the brain activity estimation obtained by using the spatial dictionary $\boldsymbol{\Phi}_s$ and the projected data $\boldsymbol{Y}\boldsymbol{V}_{(\cdot,i)}$. However, each ith element must hold just one well defined spatially coherent generator. So, to avoid elements in $\boldsymbol{\Phi}_s^*$ holding several cortical patches, a k-means clustering algorithm is applied to determine each generator correctly as independent elements of the new dictionary. As a result, we obtain a reduced spatial dictionary of size $N_r \geq N_q$.

2) Computation of time-varying hyperparameters: Given the i-th element of $\boldsymbol{\Phi}_s^*$, its corresponding hyperparameter at time instant k, noted as h_i^k, is recursively calculated using the following EEG covariance, $\boldsymbol{Q}^k \subset \mathbb{R}^{3N_d \times 3N_d}$, estimated within the fixed time window centered at time instant k:

$$\boldsymbol{Q}^k = \sum_{i=1}^{N_s} h_i^k \,\mathrm{diag}(\boldsymbol{\Phi}_s^*(\cdot, i)). \tag{2}$$

Besides, a random walk model is considered to estimate all characterizing temporal hyperparameter dynamics within the following state space framework:

$$h_i^k = h_i^{k-1} + \mu_i^k, \forall i \in N_r \tag{3a}$$

$$\mathrm{vec}\left(\mathrm{cov}\left(\boldsymbol{Y}^k\right)\right) = \mathrm{vec}\left(\boldsymbol{L}\boldsymbol{Q}^k\boldsymbol{L}^\top\right) + \boldsymbol{\gamma}^k \tag{3b}$$

where $\mathrm{vec}(\cdot)$ is the argument vectorization, $\mathrm{cov}\left(\boldsymbol{Y}^k\right)$ is the covariance estimated in the window $\boldsymbol{Y}^k$, centered at time instant k, both μ^k and $\boldsymbol{\gamma}^k \subset \mathbb{R}^{N_c^2 \times 1}$ are measures of noise that are assumed to be normally distributed with scaled identity covariance matrices.

To estimate the introduced hidden states in Eq. (3a) and (3b), the random walk model can be rewritten as to apply the following relationship between the Kronecker product (represented as $\otimes$) and the $\mathrm{vec}(\cdot)$ operator:

$$\boldsymbol{h}^k = \boldsymbol{h}^{k-1} + \boldsymbol{\mu}^k \tag{4a}$$

$$\mathrm{vec}\left(\mathrm{cov}\left(\boldsymbol{Y}^k\right)\right) = \boldsymbol{L} \otimes \boldsymbol{L}\mathrm{vec}\left(\boldsymbol{Q}^k\right) + \boldsymbol{\gamma}^k \tag{4b}$$

Under formulation of Eqs. (4a) and (4b), the hidden states $\boldsymbol{h}^k$ are estimated using the standard Kalman filtering method.Lastly, to compute the FC-TVAR, we first map the time-varying covariance matrix $\boldsymbol{Q}^k$ into a set of $N_b < N_k$ specific brain areas, hereby $\hat{\boldsymbol{Q}}^k \in \mathbb{R}^{N_b \times N_b}$ (see description of Fig. 1). After, we calculate the FC-TVAR at the time instant k ($\boldsymbol{F}^k$) from the mapped time-varying covariance matrix as:

$$\boldsymbol{F}^k = \hat{\boldsymbol{Q}^k}^\top \hat{\boldsymbol{Q}}^k, \quad, \forall k \in= 1, \ldots, T$$

where $T \subset \mathbb{R}^+$ is the time interval analysis.

III. Simulated Data

Generally speaking, to objectively assess the tested brain activity reconstruction algorithms and, by extension, EEG-based connectivity analyses, simulated data is typically used. Bearing this in mind, we simulate two active neural sources placed in the supramarginal and rostral middle-frontal gyri. To this end, The time series simulating active sources are two Morlet wavelets with the central frequency at $10\,Hz$ and time shift of $0.5\,s$ and $1\,s$, respectively. Besides, the head model used to generate the lead field matrix comprises 4000 dipoles placed only on the tessellated cortex surface, as carried out in [6], [7].

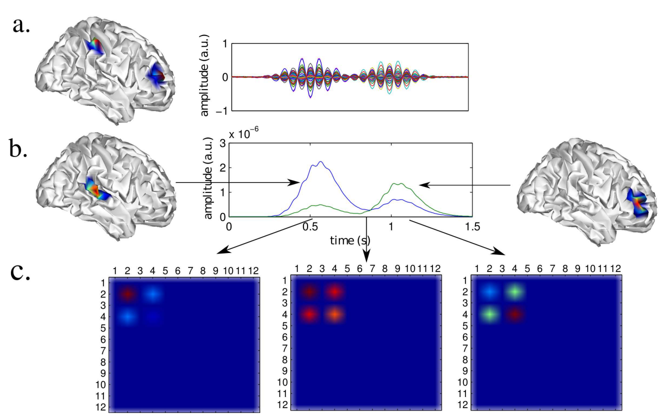
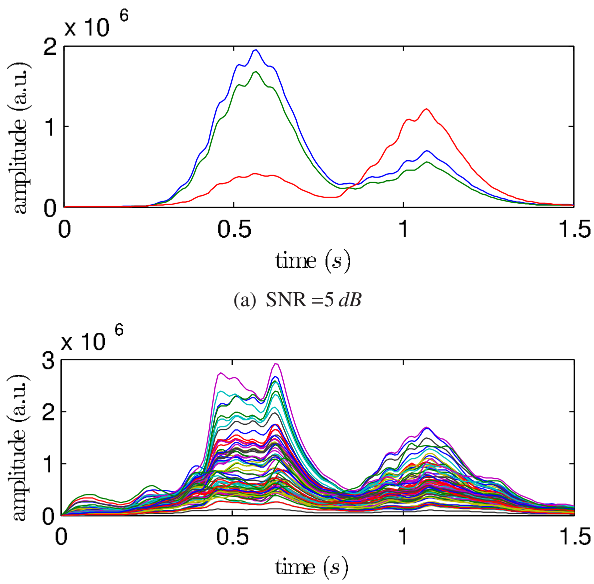


Fig. 1. Simulated brain activity and corresponding EEG (a). spatial dictionary Elements and corresponding time-varying hyperparameters (b). Instantaneous Covariance Matrix at three different time instants (c). The labels stand for: left hemisphere; 1:Inferior Parietal, lateral occipital and Temporal. 2: Post-central and supra-marginal. 3: Pre-central and caudal middle frontal. 4: Rostral middle-frontal. 5: Superior frontal. 6: Superior parietal. From 7 to 12, the same labels are for the right hemisphere.

Fig. 1 shows both the simulated activity and its obtained results where a dipole set with unknown values of amplitude and orientation is considered, yielding a total of 12000 unknowns per time instant: 3 variables per dipole representing activity strength in each one of the three-dimensional directions. Also, the simulated brain activity is measured by 59 electrodes placed according to the international `10-20` EEG system.

The simulated EEG assuming a SNR value of 12 dB is also considered. Accomplished spatial localization of the active dipole corresponding to the supramarginal gyrus has a small localization error. Specifically, the algorithm reconstructed this activity in the superior temporal gyrus. Nevertheless, all dynamics captured by the hyperparameters correspond to the simulated activity. Consequently, the proposed algorithm identified the non-stationary spatial dynamics of the EEG recording. Moreover, the instantaneous covariance matrices shown in Fig. 1.(c) encode spatial relationship among the examined brain areas. Here, we also see that the joint dynamics and transition states are correctly modeled.



(a) SNR =5 dB

(b) SNR =0 dB

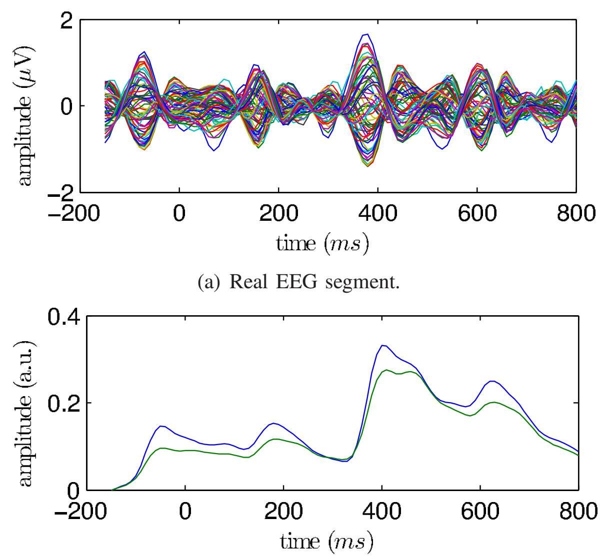
Fig. 2. Influence of noise on hyperparameter estimates of simulated EEG segments. The highest the influence - the smother the time-varying hyperparameter estimation.

Furthermore, to assess performance of the proposed method with respect to Signal-To-Noise Ratio, as seen in Fig. 2, we performed two additional experiments for a simulated EEG with an SNR of 5 dB and 0 dB, respectively. It can be seen in Fig. 2(a) that, unlike simulation with an SNR of 12 dB, an additional hyperparameter (weighting an additional element of the spatial dictionary) is generated for the case of 5 dB value. This fact suggests that for lower SNR ratios, the method generates additional dictionary elements, which should be related to noise. As seen in Fig. 2(b) for 0 dB, interpretability of the method is completely lost, i.e., a lot of spurious hyperparameters are necessary to explain the principal components identified by the SVD.

IV. Real Data

Real EEG data used to validate the proposed algorithm is a standard auditory odd-ball experiment presented in [10]. We select the non target responses of subject #6 from the available dataset. The stimuli lasting 100 ms are separated from each other in 225 ms. Approximately, 450 trials for non-target stimuli were available. EEG signals were recorded monopolarly using 63 wet electrodes placed at those symmetrical positions according to the international `10-20` EEG system. Channels were referenced to the nose. The hardware sampling rate was 1 kHz and the signal was further subsampled at 200 Hz. As the preprocessing stage, data were band-pass filtered between 0.4 and 25 Hz.

Fig. 3 shows the EEG response recording to non-target auditory stimuli and the hyperparameters of the corresponding spatial dictionary. As seen, computed time-varying hyperparameters have peaks about every 225 ms that should be related to the period of the non-target auditory stimuli. Therefore, we hypothesize that the proposed method can extract important dynamics of the underlying neural process.



(a) Real EEG segment.

(b) Hyperparameters

Fig. 3. Original EEG segment extracted from the averaged trials of the studied subject (a). Exemplary of Hyperparameters corresponding to the dictionary shown in Fig. 4.

Also, Figs. 4(a) and 4(b) show that the proposed method can localize activity around the pre-central gyri; an area that is close to the primary auditory cortex, i.e., the area that is expected to be active. Furthermore, Figs. 4(c) and 4(d) show that pairwise spatio-temporal relationship between the two auditory cortices (left and right) is successfully encoded; it can be seen a significant correlation between some neural sources located on the pre-central and post-central gyri of both hemispheres, respectively.

Similarly as in the experiments carried out for simulated data, we add noise to the considered real EEG segment in order to study how the method behaves under realistic noisy conditions. Specifically, we add Gaussian white noise

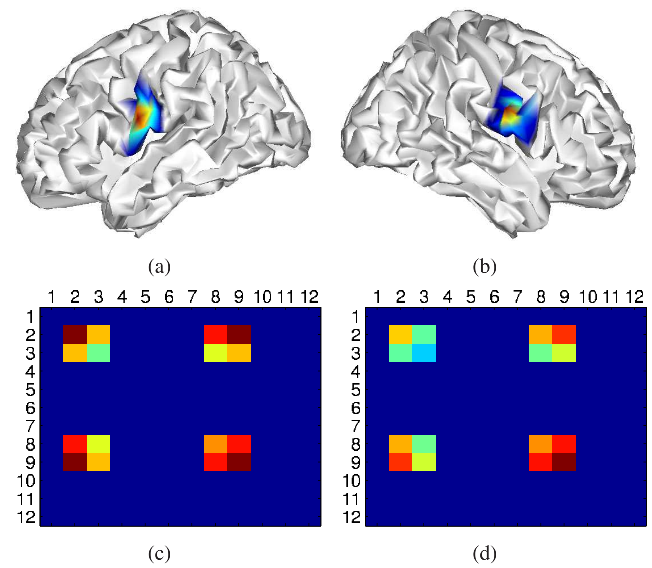


Fig. 4. Obtained results in the real EEG segment: (a)-(b) Reduced spatial dictionaries and (c)-(d) Covariance matrix at two different time instants $t = 175ms$, and $t = 400ms$. The labels are the same as in Fig. 1.

to get a 7 dB-SNR value with respect to the original segment. Obtained results can be seen in Fig. 5 showing the estimated in this case hyperparameters of the spatial dictionary. Once again as in real data, the hyperparameters also are able to capture temporal dynamics of the underlying process (at least, the periodic peaks every 225 ms). Nevertheless, processing of the noisy signal implies that the spatial dictionary has several spurious elements, thus jeopardizing the interpretability of the proposed method.

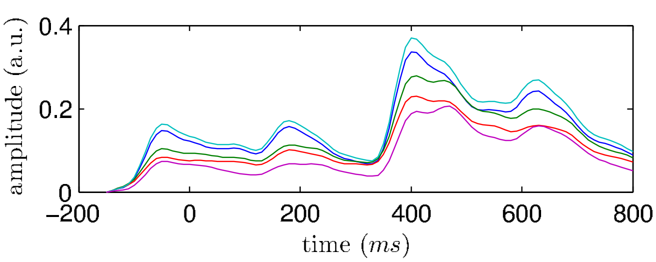


Fig. 5. Influence of noise on time-varying hyperparameters estimates of realistic EEG segments.

V. Discussion and Concluding Remarks

In the present work, we describe a novel solution for the EEG inverse problem and its use in functional connectivity analysis. The presented approach allows to relax widely-used stationarity assumption in several state-of-art inverse solutions. Nonstationarity of EEG signals is clearly assumed by computing a time varying covariance matrix of the brain activity, based on the data covariance. Consequently, relationship among complex dynamics of different brain areas can be identified.

Furthermore, to ensure that the solution has a physiologically significant distribution, the proposed time varying covariance matrix is assumed to be composed of a set of well-defined spatial basis functions as some state of art methods do (S-FLEX or MSP methods). Each one of the introduced spatial basis functions extracts independent dynamics, in this case, orthogonal temporal components given by the SVD. However, the use of SVD to identify dynamics of EEG recordings may be an over-simplifying approach: firstly, an orthogonal constraint in the components may not suffice to describe potentially complex components of the recording. Secondly, using SVD makes the proposed method highly sensitive to noisy signals, as typically seen in EEG, because the principal components identified to build the physiologically meaningful spatial dictionary would correspond to noise and not to the dynamics of interest. This issue may be fixed by time-frequency based priors [11].

On the other hand, the presented method provides interpretability that is usable in the context of functional brain connectivity analysis in real EEG data. In this regard, given that the time-varying covariance matrix is computed using windowed data, its analysis at specific time instants supply information about the neural sources or brain areas that interact with each other, and also allows to examine the temporal dynamics of such interactions. Specifically, we test the proposed method to map the brain activity response to non-target auditory stimuli. Although, within this framework it is difficult to arise strong conclusions about statistical dependencies among anatomically significant brain areas, the relationship between activation of left and right auditory cortex in the presented experiment is successfully identified.

References

[1] K. Li, L. Guo, J. Nie, G. Li, and T. Liu, "Review of methods for functional brain connectivity detection using fmri," *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 131 – 139, 2009.

[2] C. P. Coste, S. Sadaghiani, K. J. Friston, and A. Kleinschmidt, "Ongoing brain activity fluctuations directly account for intertrial and indirectly for intersubject variability in stroop task performance," *Cerebral Cortex*, vol. 21, no. 11, pp. 2612–2619, 2011.

[3] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, 2014.

[4] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity*, vol. 1, no. 1, pp. 13–36, 2011.

[5] J.-M. Schoffelen and J. Gross, "Source connectivity analysis with meg and eeg," *Human brain mapping*, vol. 30, no. 6, pp. 1857–1865, 2009.

[6] 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, no. 3, pp. 1104 – 1120, 2008.

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

[8] A. Gramfort, D. Strohmeier, J. Haueisen, M. Hämäläinen, and M. Kowalski, "Time-frequency mixed-norm estimates: Sparse m/eeg imaging with non-stationary source activations," *NeuroImage*, vol. 70, no. 0, pp. 410 – 422, 2013.

[9] J. Kim and P. Robinson, "Compact dynamical model of brain activity," *Physical Review E*, vol. 75, p. 031907, Mar. 2007.

[10] J. Hoehne, M. Schreuder, B. Blankertz, and M. Tangermann, "A novel 9-class auditory erp paradigm driving a predictive text entry system," *Frontiers in Neuroscience*, vol. 5, no. 99, 2011.

[11] J. CastaÃśo-Candamil, J. Martinez-Vargas, E. Giraldo-Suarez, and G. Castellanos-Dominguez, "Bayesian estimation of neural activity for non stationary sources using time frequency based priors," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, pp. 1521–1524, Nov 2013.