Fusion of EEG topograhic features and fMRI using Canonical Partial Least Squares

Kostas Michalopoulos
Wright State University
Assistive Technologies Research Center
Dayton, OH, USA
michalopoulos.2@wright.edu

Nikolaos Bourbakis
Wright State University
Assistive Technologies Research Center
Dayton, OH, USA
nikolaos.bourbakis@wright.edu

***Abstract*—In this paper we present a novel method for describing the EEG as a sequence of topographies, based on the notion of microstates. We use Hidden Markov Models (HMM) to model the temporal evolution of the topography of the average Event Related Potential (ERP) and we calculate the Fisher score of the sequence by taking the gradient of the trained model parameters given the sequence. In this context, the average Event Related Potential (ERP) is described as a sequence of topographies and the Fisher score describes how this sequence deviates from the learned HMM. This alternative modeling of the ERP is used to fuse EEG information, as expressed by the temporal evolution of the topography, and Functional Magnetic Resonance Imaging (fMRI). We use Canonical Partial Least Squares (CPLS) for the fusion of the Fisher score with fMRI features. In order to test the effectiveness of this method, we compare the results of this methodology with the results of CPLS using the average ERP signal of a single channel. Using this methodology we are able to derive components that co-vary between EEG and fMRI and present significant differences between the two tasks. The results indicate that this descriptor effectively characterizes the temporal evolution of the ERP topography and can be used for fusing EEG and fMRI for the discrimination of the brain activity on different tasks.**

Keywords—EEG, fMRI, Partial Least Squares, Fisher score, pattern analysis

I. INTRODUCTION

Combining information from EEG and fMRI has been a topic of increased interest recently. The main advantage of the EEG is its high temporal resolution, in the scale of milliseconds, while the main advantage of fMRI is the detection of functional activity with good spatial resolution [1], [2]. The advantages of each modality seem to complement each other, providing better insight in the neuronal activity of the brain. Although, fMRI provides an indirect measure of neuronal activity different studies have established that there exist common neural generators that explain both EEG activation patterns and the fMRI BOLD response [3]. The main goal of combining information from both modalities is to increase the spatial and the temporal localization of the underlying neuronal activity captured by each modality.

Limitations of the recording technology did not allow for the simultaneous use of both modalities. Recent advancements in the EEG technology though, made possible the recording of EEG inside the fMRI scanner [4]. This technique renewed the interest in the combination of these two modalities and allowed the development of new techniques for their integration and fusion [2], [5], [6]. Decomposition techniques as Canonical correlation analysis [7] and partial Least Squares [8] have been used for the analysis of such datasets. Independent component analysis (ICA) was extended and modified in order to decompose simultaneous recorded datasets, resulting in the Joint ICA [6] and parallel ICA methods [9]. Simultaneous EEG and fMRI recordings allow exploring the direct correspondence between EEG and BOLD variability, either in task related or resting-state experiments. One major drawback of the concurrent recordings is the severe contamination of the EEG recordings with scanner artifacts that seriously degrade the signal quality.

Initial attempts for EEG and fMRI integration, before these technology advancements, involved separate recordings of the same experiment in an effort to identify common sources of the explained brain functionality. The main drawback of separate multimodal recordings is that it is no longer possible to establish direct correspondence between fMRI and EEG trials and therefore it is not possible to directly exploit the response variability in the trial level. It is apparent that some transformation of the data in a common space and the extraction of representative statistics is needed for the effective analysis of such recordings. In the analysis of such datasets the main assumption is that the response to a given stimulus remains the same when recorded in different time points. Different methodologies have been applied for the analysis of separate recordings [10]–[13]. The main idea is to align the two modalities together and work on the estimated response elicited from the task presentation.

On the other hand, the main advantage is that separate recordings do not suffer from signal quality degradation due to scanner interference. The majority of the studies of separate recordings have been focused in solutions in the source space, where source localization techniques are used in order to estimate the brain regions that are more probable to generate

978-1-4799-7502-0/14 $31.00 © 2014 IEEE
DOI 10.1109/BIBE.2014.53

the observed EEG. In this context fMRI is used for evaluation of the result or for restricting the search space of the localization algorithm. In order to study significant features of the EEG and fMRI, analysis of separate recordings exploits the common subject to subject variations in group analysis setting. Canonical Correlation analysis has been used for this purpose with significant results [7]. In this paper we present a novel feature for the characterization of the average ERP that can be used for the decoding EEG activities. We are using EEG and fMRI data that are separately recorded in order to identify common sources of co-variation between the datasets and use the common subspace in order to exploit information from both modalities.

II. Feature Extraction

A. EEG as sequence of field topography maps

Recently, there has been an increased interest in methods that exploit information carried by the topographic configuration of the electrical field in the scalp electrodes [14], [15]. Initially, this approach was introduced in [16], where it was first observed that the topography of the electrical field does not change randomly, but rather follows certain patterns, depending on the task at hand. It was though that these distinct topographies reflect the underlying sequence of brain activation and in a way represent a higher representation of the activation sequence of brain functions needed to complete the task.

In order to recover the dominant topographies from the EEG signal different methodologies have been applied either in the average or the single trial Event Related Potential (ERP) [15], [17]. The main approach in this problem is to first define a measure of similarity between topographies and then apply a clustering algorithm in order to represent the set of topographies using the cluster centroids. In [18], we introduced a new measure of similarity based on the Local Global Graph (LG graph) which was applied for the segmentation of the average ERP. This measure treats the topographic map as an image and uses segmentation in order to extract the LG graph. The segmentation step provides the additional advantage that reduces the dimensionality of the problem and reduces the co-linearity of the measurements of nearby channels, by grouping neighbor channels together. The similarity between topographies is measured in a hierarchical way, starting from the local similarity between nodes of the local graph and then taking under consideration the global relation of the nodes.

1) Modelling using Hidden Markov Models

Different algorithms have been applied for clustering the topographies. A modified k-means algorithm has been applied in [17]. Also hierarchical clustering algorithms and soft clustering algorithms as Gaussian mixture models [15], [18] have been successfully applied for the identification of dominant topographies. Some implementations incorporate temporal filtering of the results in order to remove isolated topographies in time and create a smooth temporal segmentation. In this study, we model the topographic sequence using Hidden Markov Models (HMM).

HMMs are bivariate random processes consisting of a random variable modeling the observed processes and a hidden Markov chain which describes the transition between the

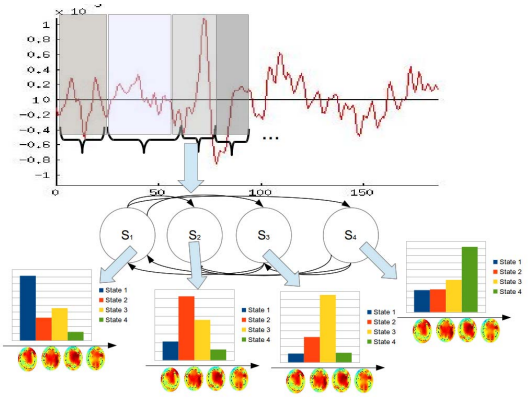

Figure 1: Illustration of the Hidden Markov formulation for the modelling of the ERP topography. Hidden states of the HMM represent periods with quasi-stationary distribution of topographies.

different states. The probability of the current hidden state depends only on the previous state [19]. The distribution of the observations depends only on the current state and is independent of previous observations and states. More formally a HMM is defined as:

$$\lambda = \{\pi, A, B\} \tag{1}$$

$$A_{k,l} = P(S_{t+1} = l | S_t = k) \tag{2}$$

$$B_i = P(O_t^j | S_t = k) \tag{3}$$

$$\sum_{\kappa=1}^{N} \pi_\kappa = 1 \tag{4}$$

where, π represents the initial probabilities of the states, the elements of A the transitions probabilities among the hidden states and B the distribution of the observations under the given state.

In our case, we consider the sequence of the topographies as the observed variable and the different and the underlying hidden states as periods that present stationary distribution of topographies. The HMM seems to be a good fit in the microstate model, since it allows to model both the spatial and temporal relationships of our data. Initial results, in a work to appear, on single trial activations of healthy and Progressive Mild Cognitive Impairment subjects indicate that we were able to distinguish between the activations of the two groups using a generative classifier build from the HMMs.

HMMs can be considered in the more general framework of Dynamic Bayesian Networks (DBN). HMMs are the simplest DBN models we can build. In [20], we extended the modeling of the topographic sequence using Dynamic Bayesian Networks. The trained networks were used for the binary classification between two tasks with good results and confirmed that the topographic sequence carries valuable information about the underlying brain processes complementary to the time-frequency analysis approach. The main problem with this approach is that using these generative models it is difficult to assess the impact of each parameter of the model and therefore it is difficult to associate the model parameters with other features and well known descriptors of

the brain response. In an effort to combine the strengths of generative models with the strengths of discriminative ones, we introduce the Fisher score for the analysis of EEG.

The procedure for building the HMM is to apply a vector quantization step and then learn the model parameters on the encoded sequence. Specifically, we use the k-means algorithm on the multichannel vectors of the average ERP from all subjects in order to extract the observed symbols. The encoded sequence is used to learn the parameters of the discrete HMM using the Baum-Welch algorithm [19]. We apply this methodology on the wideband average ERP signal for extracting the topography codebook and learning the model parameters [20].

B. Mapping to Fisher Score space

The Fisher score space was introduced in [21] in an effort to bridge generative and discriminative models. The main motivation was to map variable length sequences into fixed length feature vectors, a problem often encountered in bioinformatics [21]. Fisher score uses the derivative of the parameters of the HMM given a certain sequence. Using the derivative of each parameter for a sequence we are able to build a feature vector of length equal to the number of parameters of the model. These vectors can be used to form the so-called Fisher kernel which can be used with kernel classifiers and methods [22] .The Fisher score has been very popular recently and have found many applications in classification of protein sequences, text, speech recognition and images for face recognition, shape/ texture recognition and activity recognition.

The Fisher score maps a sequence to fixed length vector using the parameters of the generative model. The gradient of the parameters for a given sequence are used to accomplish this. In the case of HMMs, the gradient of the parameters of the trained HMM is computed for a sequence. The derived gradient describes how the parameters of the model must change in order to adapt to the new sequence. Therefore, the derived features can be used to evaluate how well the given sequence fits the model parameters and we can evaluate the deviation from a given parameter explicitly. The Fisher vector for each parameter is defined as:

$$F_i = \nabla_{\theta_i} log\ P(O|\theta_i) \quad (5)$$

For the case of modeling the EEG topography, the Fisher score of sequence provides useful insight for which parameter deviates most for a given sequence. The gradient of the parameters of the diagonal of the transition matrix A reflect the difference in the mean duration of a given state while the gradient of the off-diagonal entries of the matrix reflect the difference in the mean frequency of transition from a state to another. The same applies for the emission parameters with respect to the topography distribution. For the discrete case the gradient indicates a change in the distribution, where a certain representative topography may appear more often than expected in a given state. The transition parameters are easier to interpret in the continuous case though, were the emission distribution is modeled with a single or a mixture of Gaussian. In this paper we will present results for the discrete case, although the derivation for the continuous is straightforward. The gradient of the transition matrix can be calculated using the sufficient statistics of the HMM. The sufficient statistics of the HMM (the forward and backward probabilities) can be derived by the forward-backward algorithm [19].

III. METHODS

A. Partial Least Squares for multimodal fusion

Our goal is to extract information from the combination of the modalities. We assume that both modalities capture certain aspects of the brain response to the task, from different perspective. In the case of EEG, it captures the neuronal response attributed mainly to large pyramidal cells, while fMRI measures the change in the blood oxygenation in certain areas as a result of the stimulus.

Different methods have been employed for fusing data from different modalities. Joint ICA and parallel ICA are two of the most popular algorithms used for this purpose. The main problem with these approaches is the strong assumption of independence imposed to the latent variables. CCA on the other hand has been used successfully for the fusion EEG and fMRI data [7]. CCA uncovers latent components that are maximally correlated between datasets. The main problem of the CCA method is that it maximizes the correlations between latent variables of the two sets and operates on the cross-correlations matrix. Therefore, CCA is vulnerable to outliers and it is often the case the solution provided by CCA to fail to summarize the variance of the two datasets [26].

On the other hand Partial Least Squares (PLS) methods have been used successfully for the analysis of neuroimaging data [27] and have found applications in different problems [28]. Multiway PLS has been applied in the analysis of simultaneous recorded EEG and fMRI data in [8]. It seems though that PLS methods have been neglected partially due to the success of ICA methods in the analysis of EEG and fMRI. In general, there exist different formulations of the partial least Square method for the analysis of two sets of variables. The one known as Canonical Partial least squares (CPLS) tries to uncover the shared information between two sets of variables. It also known as Canonical Two Block Mode A PLS (PLS-C2A) [29]. The main advantage of CPLS methods over CCA is that CPLS maximizes the covariance between the latent variables of the two datasets and therefore avoids the poor summary of variance problem of CCA, by directly maximizing the cross-covariance between the two sets. Although the formulation of CPLS looks similar to that of CCA, the computational details of CPLS make the solution numerically stable and the results are easier to interpret than CCA [29].

In general, CPLS tries to find two sets of latent variables (one for each set) that maximally co-vary. More formally we assume that we have two sets of variables X, Y, where the columns are the different variables and the rows are the paired samples. We also assume that variables of X and Y have zero mean and are scaled to unit variance. The main assumption is that the variables X are generated by the same number of latent variables as Y. Suppose that we have two paired sets of latent variables L_x, L_y for each set. We are interest to uncover these pairs of latent variables and CPLS does so by modeling the

cross-covariance by such pairs of latent variables. CPLS calculates pairs of latent variables defined as:

$$\{L_{X_i}, L_{Y_i}\}, i \in \{1, .., r\} \quad (6)$$

so that variables $\{L_{X_i}\}$ and $\{L_{Y_i}\}$ represent the most interesting subspaces (in the least square sense) of the cross-covariance matrix$R$. R is defined as:

$$R = X^T Y \quad (7)$$

and we want to find linear combinations $\{L_{X_i}\}$ and $\{L_{Y_i}\}$ such that:

$$Cov(L_{X_1}, L_{Y_1}) = max\{Cov(Xu, Yv)\} \quad (8)$$

where u, v are the coefficients that maximize (3). It is well known that the solution of Eq.(3) can be obtained by solving the singular value decomposition of the cross-covariance matrix as:

$$R = UDV^T \quad (9)$$

Therefore we can recover the linear combination weights for each set, which for set X are the singular vectors U and for the Y set the singular vectors V. The latent variables can now be calculated as:

$$L_X = XV \quad (10)$$

$$L_Y = YU \quad (11)$$

Using this method, we are able to describe, using a linear combination, the features from X that maximally co-vary with features from Y. Working with the covariance matrix guarantees that the latent variables will sufficiently describe the cross-covariance structure of the data, in contrast to CCA that it may provide a poor summary of the datasets variability.

B. Dimensionality reduction using PCA.

The main problem in the analysis of fMRI data is the large number of features compared to the sample size. A single scan of the head usually involves hundreds of thousands of voxels. There has been applied several feature selection techniques in order to reduce the amount of features and allow us to work with a smaller subset of features. Choosing an appropriate region of interest and working with the corresponding voxels is one such technique. Another approach is to first solve the General Linear Model and based on that choose the voxels that differ significantly among tasks. This way we can select only the most informative features for our analysis.

Despite these approaches that significantly reduce the number of features, most of the time it is beneficial to further reduce the dimensionality of the problem. Principal component analysis is a well known method that transforms our data to a lower dimensional space while retaining as much of the original variance as possible. PCA projects the data into the lower subspace as follows:

$$Y = XU \quad (12)$$

where U is the matrix of the eigenvectors of the covariance matrix of X. By choosing the k eigenvectors that correspond to the largest eigenvalue we can map our original variable X to a lower space Y'.as follows:

$$Y' = XU' \quad (13)$$

PCA is well known technique and has been applied extensively for reducing the dimensionality of the initial problem. We apply PCA to the processed EEG and fMRI feature sets separately as a preprocessing step before the analysis of the cross-covariance of the two sets.

IV. APPLICATION

Our goal is to couple information from both modalities in an effort to generate new features that capture the variation between subjects. The main idea is that instead of producing features that summarize the variance of each set independently, we use PLS in order to describe the structure of the cross-covariance matrix in an effort to capture the features that present task related behavior between the two modalities. Therefore, in contrast with PCA which operates in the covariance matrix of each set to reduce the dimensions of the problem, we are taking into consideration the modulation of features across modalities in order to recover task related features.

A. Data Description

The data come from the study [30] and have been made publicly available from the authors. We used the EEG and the fMRI part of the dataset, only. In this dataset 16 subjects were asked to perform a visual task where faces of famous persons, unfamiliar faces and scrambled faces where presented to them. The complete description of the dataset can be found in [30]. The EEG data were recorded in separate sessions several days apart from the fMRI session.

There were 300 faces and 150 scrambled faces in total. The scrambled faces were created by taking the Fourier transform of a group of 150 images of faces. The phases of the transformed images were permuted and then inverse transformed in the original space. Finally, the new scrambled image was masked using the outline of the original face. For the MEG/EEG data each subject completed 900 trials over six sessions, while for the fMRI each subject completed 900 trials over 9 sessions [30].

B. EEG preprocessing and Fisher score calculation

The data were examined for artifacts and Independent component analysis was used in order to remove eye-blink artifacts [31]. Trials that were heavily contaminated were excluded from further analysis. The data were band-pass filtered in the range 0.5-31Hz using a linear Finite Impulse response filter and the average over all trials per task was calculated for all subjects resulting in 2 average ERPs for each subject.

Using the average ERP we extracted the discrete HMM as described in section II. We evaluated different number of states and number of symbols for the HMM. The number of symbols were chosen so that the reconstruction error of each sequence to be less than 30%. The resulting codebook consisted of 15 topographies. We used 6 hidden states for the HMM and these were selected by testing the accuracy performance of a simple maximum likelihood classifier.

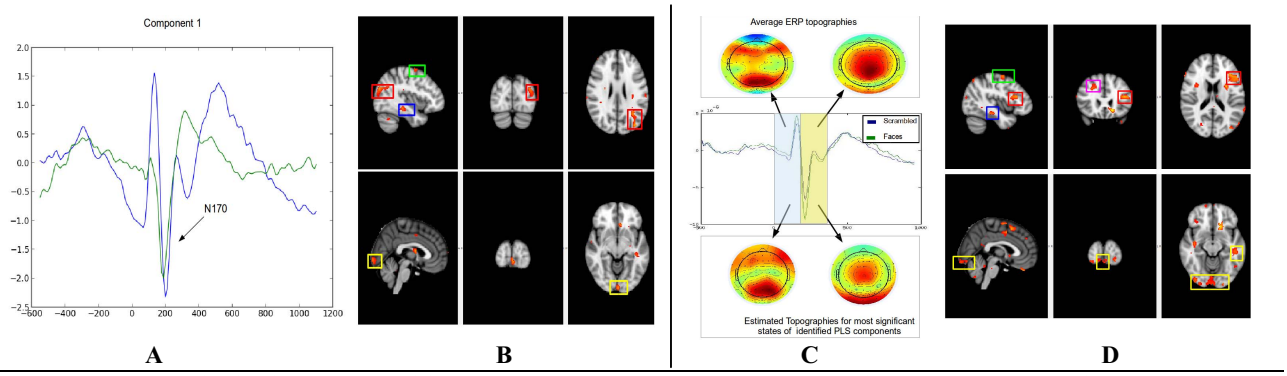

Figure 2: CPLS Componenst related to the N170 peak and the corresponding fMRI activation. **Left** Results from the application of CPLS in the average ERP and fMRI. A component that captures the acivity around the N170 peak is presented (A) and the corresponding component for the fMRI (B). **Right**, Results from the application of CPLS in the EEG Fisher score dataset and FMRI. The most important states recovered from the components that present significantly different activation between the tasks are presented for EEG (C) and fMRI (D). **Figure 2C** top, the mean topographies for the given periods are presented for comparison. At the bottom the estimated topographies from the hidden states with the highest PLS loadings are presented, for the corresponding periods. **Comparison**:We can see that we have common activations for the fMRI activity between the components of CPLS applied in the two datasets. Relevant activity to the N170 is marked with colored rectangles. Yellow rectange marks the area in the Occipital gyrus, while green and blue mark the middle temporal gyrus(blue rectangle) and the precentral gyrus(green rectangle). For the average ERP case with red is marked the lateral superior Occipital cortex. For the Fisher score case red and magenta mark the inferior and superior frontal gyrus, respectivelly.

We followed a simple strategy for calculating the Fisher score for the average sequences. We trained the HMM, on the scrambled faces trials and then the Fisher score of all ERPs was calculated based on this HMM. The Fisher score of each ERP reflects the deviation of the sequence from the HMM parameters learned using the scrambled faces and therefore we expect smaller deviations for ERPs of the same class and larger for the other classes. We treat famous and unfamiliar faces the same and the gradient of the parameters reflect the change in the model depending on the stimulus.

C. *fMRI preprocessing*

The fMRI volumes consisted of 33 T2-weighted transverse echoplanar images. Each session consisted of 210 volumes with a repetition time (TR) of 2000ms. FMRI data were preprocessed as in [30]. We used SPM5 [32] in order to register the slices of each subject together. A T1-weighted image of each subject was segmented to gray matter, white matter and Cerebrospinal fluid and the segments were registered to the corresponding segments of an MNI template in Tailarach space [32]. The slices of each run were first co-registered and then time corrected. The co-registered volumes were then registered and normalized to the processed T1 volume of the corresponding subject.

D. *FeatureExtraction*

Since we want to take advantage of information from both EEG and fMRI using CPLS, we have to work with paired datasets. Trials from EEG and fMRI were recorded at different times and therefore we cannot have a direct correspondence between trials of the same subject. For this reason we worked across subjects. For the EEG we built two datasets for evaluation of our approach. In one, we used the Fisher score vector and in the other the average ERP on a single channel as feature. A channel located in the occipital area was selected that presents strong ERP peaks after the stimulus. The ERP peaks that we anticipate to observe from this type of stimulus, is a positive peak at 100ms (P100) and a negative peak at 170ms (N170) after the stimulus.

For the fMRI data we are going to work with the beta maps of each subject. We model the HRF response of each subject using a canonical HRF, build from two gamma functions [32]. Using the General Linear model we model the activity of each voxel as:

$$Y_i = Xb_i + e \tag{14}$$

where Y is the activity of a given voxel and X is the design matrix of the experiment. We estimate, for each voxel the beta value and use the corresponding maps to build the fMRI dataset. The final fMRI dataset consists of sixteen subjects and 3 parametric maps per subject, one for each task.

V. RESULTS

A. *Evaluation using the average ERP and fMRI.*

We performed CPLS using the features from EEG and fMRI. We constructed two EEG datasets. The first consists of the average of each subject. In this case the samples of the average ERP are the features that we are going to use. The dimensionality of the fMRI data was reduced before applying CPLS. Initially, we selected voxels that differ significantly between tasks, as indicated by an ANOVA test. We set a threshold of 0.05, uncorrected. This liberal threshold was used as an initial data reduction scheme. Our intent is to keep the most informative voxels and reduce the dimensionality rather than to uncover the specific task related voxels. We want to keep enough variability in the data for CPLS to work with. The resulting dataset was further reduced using principal component analysis.

We applied CPLS in the set consisting of the average ERP dataset and the reduced fMRI. The results can be seen in Figure 2. An EEG component that corresponds in the N170 peak is presented along with the corresponding fMRI component. We can see that the EEG component characterizes well the activity

around the N170 peak. On the other hand, the fMRI CPLS component reveals areas that co-vary with the EEG component and are in agreement with other studies regarding the areas involved in the generation of N170 [33]. Significant areas of activations were calculated by thresholding the Z transformed loadings of the fMRI component. We can observe that the N170 component is mostly associated with activations in the occipital and lingual gyrus (marked with yellow rectangle), the left superior Occipital cortex (red rectangle), the middle temporal gyrus (blue rectangle) and the precentral gyrus (green rectangle). These regions match the ones reported by other studies [33] and confirm that results of the decomposition. These initial results indicate that CPLS is able uncover meaningful components that co-vary between the two datasets.

B. Evaluation using the Fisher score and fMRI

In this case we used the EEG dataset containing the features derived by taking the gradient of the HMM parameters. The parameters of interest are the transitions between the states of the HMM model. In order to localize the activity in time, we use three windows around the peaks of interest. The first period starts right after the stimulus onset and includes the P100 peak. The second period includes the negative peak N170 and the last period covers the late period of the trial up to 650ms after the stimulus onset, as can be seen in Figure 2.

We applied CPLS in the two datasets and the results can be observed in Figure 2(C-D). In this case, we searched for components that present significant differences among the stimuli. We used Student's t-test to search for components that displayed significant difference between the two tasks. We used the loadings of these components to map the contribution of the components back to the original variables. The loadings were reformatted into the corresponding transition matrices for each window and the transition weights were Z normalized. In this case, we found components whose loadings significantly differ between tasks, in contrast to the previous one (using the average ERP). In figure 2 right, we present the pair of components that present the greatest difference between tasks.

For each time period, we identified the most important entry in the transition matrix of the components. For all the windows, the features used represent the transition to the same state and therefore can be translated as differences in the mean duration of the state. Since we are working with the average signal, we don't expect significant differences in the rate of transition from one state to another. The transitions in the same state can be explained as the mean duration of the state at the given interval. In figure 2 right the mean topographies, as calculated by the emission probabilities of the identified state, are presented for the P100 and N170 periods along with the corresponding fMRI component.

In this case the results present similarities with the ones for the component for N170 of the average set. We can observe significant fMRI activations located in the occipital lobe (yellow) and the occipital fusiform gyri (yellow) and the post central gyrus (green). In this case, the components project significantly to all periods and therefore is expected to have more active regions than in the case of the average ERP. The additional activations are located mainly in the inferior and superior frontal gyri (red) and are related to ERP components that appear later in time (as the P600) [33].

VI. DISCUSSION

In this study, we introduced a novel feature based on the concept of the Fisher score. We used this novel modeling of the EEG to fuse together information from EEG and fMRI. We use HMMs in order to characterize the topography and the temporal evolution of the average ERP. Based on the trained HMM, for each sequence we calculate the Fisher of each parameter for that sequence. This approach builds on our previous work used to characterize the EEG topography by operating on the HMM parameters. The transformation to the Fisher score space, allows bridging the descriptive power of generative models with the efficiency of discriminative approaches and provides a novel framework for characterizing the EEG activity. The derived score describes how the parameters of the model must change in order to adapt to the new sequence. Using the derived feature vectors we can use standard discriminative techniques for the decoding of the EEG.

We apply this technique in a multimodal dataset of 16 subjects performing a visual task and use the derived results for fusing EEG information with fMRI using CPLS. We evaluated the effectiveness of CPLS in describing the cross-covariance between the EEG and fMRI datasets, by applying it of the average ERP and the fMRI features. CPLS successfully recovered the prominent peaks of the average ERP in different components and the corresponding fMRI component localizes the activation in regions that are related to the corresponding peak, as expected from the literature.

Using the same technique between the Fisher score features of the EEG and the fMRI, the results show that using CPLS it is possible to relate the hidden states described by the HMM model to fMRI activations. Using the Fisher kernel we were able to recover components that are significantly different between states. These results indicate that the Fisher score can be used for the decoding the EEG activation patterns of different tasks. On top of that, we demonstrated that CPLS can be used to combine and fuse information between modalities and its results can be used for multimodal classification and decoding. In future work, we intend to explore the use of the Fisher score and CPLS for the discrimination between tasks in a multimodal multivariate setting.

ACKNOWLEDGMENT

The authors want to thank Dr. Wakeman and Dr. Henson for sharing the dataset by making it publicly available.

REFERENCES

[1] K. Michalopoulos, M. Zervakis, and N. Bourbakis, "Current Trends in ERP Analysis Using EEG and EEG/fMRI Synergistic Methods," Humana Press, 2013, pp. 1–28.

[2] J. Sui, T. Adali, Q. Yu, J. Chen, and V. D. Calhoun, "A review of multivariate methods for multimodal fusion of brain imaging data," J. Neurosci. Methods, vol. 204, no. 1, pp. 68–81, Feb. 2012.

[3] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, A. Oeltermann, and others, "Neurophysiological investigation of the basis of the fMRI signal," Nature, vol. 412, no. 6843, pp. 150–157, 2001.

[4] R. I. Goldman, J. M. Stern, J. Engel Jr, and M. S. Cohen, "Acquiring simultaneous EEG and functional MRI," Clin. Neurophysiol., vol. 111, no. 11, pp. 1974–1980, Nov. 2000.

[5] T. Eichele, V. D. Calhoun, M. Moosmann, K. Specht, M. L. A. Jongsma, R. Q. Quiroga, H. Nordby, and K. Hugdahl, "Unmixing concurrent EEG-fMRI with parallel independent component analysis," Int. J. Psychophysiol., vol. 67, no. 3, pp. 222–234, Mar. 2008.

[6] V. Calhoun, R. Silva, and J. Liu, "Identification of Multimodal MRI and EEG Biomarkers using Joint-ICA and Divergence Criteria," in 2007 IEEE Workshop on Machine Learning for Signal Processing, 2007, pp. 151 –156.

[7] N. M. Correa, Y.-O. Li, T. Adali, and V. D. Calhoun, "Canonical Correlation Analysis for Feature-Based Fusion of Biomedical Imaging Modalities and Its Application to Detection of Associative Networks in Schizophrenia," IEEE J. Sel. Top. Signal Process., vol. 2, no. 6, pp. 998 –1007, Dec. 2008.

[8] E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway Partial Least Squares," NeuroImage, vol. 22, no. 3, pp. 1023–1034, Jul. 2004.

[9] J. Liu and V. Calhoun, "PARALLEL INDEPENDENT COMPONENT ANALYSIS FOR MULTIMODAL ANALYSIS: APPLICATION TO FMRI AND EEG DATA," in 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007, 2007, pp. 1028 –1031.

[10] B. He and Z. Liu, "Multimodal Functional Neuroimaging: Integrating Functional MRI and EEG/MEG," IEEE Rev. Biomed. Eng., vol. 1, no. 2008, pp. 23–40, Nov. 2008.

[11] C. Grova, J. Daunizeau, E. Kobayashi, A. P. Bagshaw, J.-M. Lina, F. Dubeau, and J. Gotman, "Concordance between distributed EEG source localization and simultaneous EEG-fMRI studies of epileptic spikes," NeuroImage, vol. 39, no. 2, pp. 755–774, Jan. 2008.

[12] X. Lei, D. Ostwald, J. Hu, C. Qiu, C. Porcaro, A. P. Bagshaw, and D. Yao, "Multimodal Functional Network Connectivity: An EEG-fMRI Fusion in Network Space," PLoS ONE, vol. 6, no. 9, p. e24642, 2011.

[13] F. Babiloni, F. De Vico Fallani, and F. Cincotti, "Multimodal Integration of EEG, MEG, and Functional MRI in the Study Of Human Brain Activity," in Advanced Methods of Biomedical Signal Processing, S. Cerutti and Carlorchesi, Eds. John Wiley & Sons, Inc., 2011, pp. 153–167.

[14] L. Dipietro, M. Plank, H. Poizner, and H. I. Krebs, "EEG microstate analysis in human motor corrections," in 2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), 2012, pp. 1727–1732.

[15] A. Tzovara, M. M. Murray, G. Plomp, M. H. Herzog, C. M. Michel, and M. De Lucia, "Decoding stimulus-related information from single-trial EEG responses based on voltage topographies," Pattern Recognit., vol. 45, no. 6, pp. 2109–2122, Jun. 2012.

[16] D. Lehmann, H. Ozaki, and I. Pal, "EEG alpha map series: brain micro-states by space-oriented adaptive segmentation," Electroencephalogr. Clin. Neurophysiol., vol. 67, no. 3, pp. 271–288, Sep. 1987.

[17] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Segmentation of brain electrical activity into microstates: model estimation and validation," IEEE Trans. Biomed. Eng., vol. 42, no. 7, pp. 658–665, 1995.

[18] K. Michalopoulos and N. Bourbakis, "Microstate analysis of the EEG using local global graphs," in Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on, 2013, pp. 1–5.

[19] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE, vol. 77, no. 2, pp. 257–286, 1989.

[20] K. Michalopoulos, V. Sakkalis, V. Iordanidou, and M. Zervakis, "Activity detection and causal interaction analysis among independent EEG components from memory related tasks," in Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, 2009, pp. 2070 –2073.

[21] T. Jaakkola, M. Diekhans, and D. Haussler, "A Discriminative Framework for Detecting Remote Protein Homologies," J. Comput. Biol., vol. 7, no. 1–2, pp. 95–114, Feb. 2000.

[22] A. Velivelli, T. Huang, and A. Hauptmann, "Video shot retrieval using a kernel derived from a continuous HMM," Proc SPIE, p. 607311, Jan. 2006.

[23] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, and R. Turner, "Event-Related fMRI: Characterizing Differential Responses," NeuroImage, vol. 7, no. 1, pp. 30–40, Jan. 1998.

[24] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini, "The distributed human neural system for face perception," Trends Cogn. Sci., vol. 4, no. 6, pp. 223–233, Jun. 2000.

[25] M. Hanke, Y. O. Halchenko, P. B. Sederberg, E. Olivetti, I. Fründ, J. W. Rieger, C. S. Herrmann, J. V. Haxby, S. J. Hanson, S. Pollmann, M, "PyMVPA: a unifying approach to the analysis of neuroscientific data," Front. Neuroinformatics, vol. 3, p. 3, 2009.

[26] R. A. Johnson and D. W. Wichern, Applied multivariate statistical analysis. Prentice Hall, 1992.

[27] A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi, "Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review," NeuroImage, vol. 56, no. 2, pp. 455–475, May 2011.

[28] H. Su and G. Zheng, "A Partial Least Squares Regression-Based Fusion Model for Predicting the Trend in Drowsiness," IEEE Trans. Syst. Man Cybern. Part Syst. Hum., vol. 38, no. 5, pp. 1085–1092, Sep. 2008.

[29] J. A. Wegelin, "A Survey of Partial Least Squares (PLS) Methods, with Emphasis on the Two-Block Case," 2000.

[30] R. N. Henson, D. G. Wakeman, V. Litvak, and K. J. Friston, "A Parametric Empirical Bayesian Framework for the EEG/MEG Inverse Problem: Generative Models for Multi-Subject and Multi-Modal Integration," Front. Hum. Neurosci., vol. 5, Aug. 2011.

[31] T.-P. Jung, C. Humphries, T. Lee, S. Makeig, M. J. Mckeown, V. Iragui, and T. J. Sejnowski, Removing Electroencephalographic Artifacts: Comparison between ICA and PCA. 1998.

[32] V. Litvak, J. Mattout, Jrmie, S. Kiebel, C. Phillips, R. Henson, J. Kilner, G. Barnes, R. Oostenveld, J. Daunizeau, G. Flandin, W. Penny, and K. Friston, "EEG and MEG Data Analysis in SPM8," Comput. Intell. Neurosci., vol. 2011, Mar. 2011.

[33] J. Wirsich, C. Bénar, J.-P. Ranjeva, M. Descoins, E. Soulier, A. Le Troter, S. Confort-Gouny, C. Liégeois-Chauvel, and M. Guye, "Single-trial EEG-informed fMRI reveals spatial dependency of BOLD signal on early and late IC-ERP amplitudes during face recognition," NeuroImage, vol. 100, pp. 325–336, Oct. 2014.