Analyzing the Information Distribution in the fMRI Measurements by Estimating the Degree of Locality

Itir Onal, Mete Ozay, *Student Member, IEEE*, Orhan Firat, İlke Öztekin, Fatoş T. Yarman Vural, *Member, IEEE*

Abstract— **In this study, we propose a new method for analyzing and representing the distribution of discriminative information for data acquired via functional Magnetic Resonance Imaging (fMRI). For this purpose, we form a spatially local mesh with varying size, around each voxel, called the seed voxel. The relationship among each seed voxel and its neighbors is estimated using a linear regression model by minimizing the square error. Then, we estimate the optimal mesh size that represents the connections among each seed voxel and its surroundings by minimizing Akaike's Final Prediction Error (FPE) with respect to the mesh size. The degree of locality is represented by the optimum mesh size. Our results indicate that the local mesh size with the highest discriminative power varies across individual participants. The proposed method was tested on an fMRI study consisting of item recognition (IR) and judgment of recency (JOR) tasks. For each participant, the estimated arc weights of each local mesh with different mesh size are used to classify the type of memory judgment (i.e.IR or JOR). Classification accuracy for each participant was derived using k-Nearest Neighbor (k-NN) method. The results indicate that the proposed local mesh model with optimal mesh size can successfully represent discriminative information for neuroimaging data.**

I. INTRODUCTION

How information is represented and distributed in the brain is a fundamental question in cognitive neuroscience. In order to address this question, functional Magnetic Resonance Imaging (fMRI) has been widely used as a powerful tool. A recently growing method [1-7], Multivariate pattern analysis (MVPA) aims to extract discriminative information based on distributed patterns of activation. In this approach, neural activation of multiple voxels is used as a feature vector to train a machine learning algorithm. The performance of this algorithm indicates the accuracy of the model to represent the underlying cognitive process. This approach is also referred to as decoding or classification [8-10].

Generally, in MVPA approaches, after the raw data is preprocessed, the voxel intensity values are fed to one of the well-known classifiers or clustering algorithms, such as Neural Networks, Bayesian classifiers, Kernel machines or Ensemble classifiers [11]. Since MVPA methods focus on high spatial-frequency patterns of response, the methods are conducted within individual participant's data [12]. The general trend is to feed the multiple voxel intensity values to a classifier for each participant, where the voxel intensity values are acquired after the same pre-processing steps.

In this study, we propose a new method to investigate the distributed nature of discriminative information in the human brain. Instead of using a vector of multiple voxel intensity values, we use a local mesh model that represents the spatial relationship among voxels, which has been previously shown to have more discriminative power than voxel intensities in [11]. A local mesh is formed around each voxel (called seed voxel of the local mesh) with its spatially closest voxels. The arc weights of the local mesh, that represent the relationship between the seed voxel and its neighboring voxels, are estimated using a linear regression model for each mesh. The error coming from this regression varies as a function of the selected number of neighboring voxels in the local mesh (i.e. the mesh size) for each participant. This error plays an important role in determining the optimal mesh size across individual participants. In this study, we determine the mesh size by employing the error variance in an information theoretic criterion [13]. This criterion, namely the Final Prediction Error (FPE), is a function of the error variance, model order and number of samples. By adopting the Akaike's Final Prediction Error (FPE) criterion to our local mesh model, we compute the optimum mesh size, which corresponds to the order of the linear regression model [13, 14]. In the proposed study, the mesh size is estimated by minimizing the FPE with respect to the order of a regressor. The minimum of the FPE that determines the optimum local mesh size is unique for each individual participant and does not differ for experimental categories. Furthermore, the suggested local mesh model with varying mesh size provides an effective tool to represent the voxel connectivity of the fMRI measurements.

In the current study, 8 participants completed two memory tasks, namely item recognition (IR) and judgment of recency (JOR). In each task, participants studied five consonants and made a memory decision to two consonants probes. In IR, participants were asked to indicate the letter that was in the study list. In JOR, task they were asked to indicates the letter that was presented more recently in the study list. During these operations, neural activation was recorded using fMRI. The local mesh model was employed on the acquired data, and classification accuracy for discriminating the IR and JOR judgments was used to test the feasibility of the proposed method.

I. Onal, is with the Department of Computer Engineering, Middle East Technical University, 06800, Ankara-Turkey (corresponding author to provide phone: (90) 312-2105597; e-mail: itir@ ceng.metu.edu.tr).

M. Ozay, O. Firat, and F. T. Yarman Vural are with the Department of Computer Engineering, Middle East Technical University, 06800, Ankara-Turkey (e-mail: mozay@metu.edu.tr, orhan.firat@ceng.metu.edu.tr, vural@ceng.metu.edu.tr).

İ. Oztekin is with the Department of Psychology, Koç University, 06800, İstanbul-Turkey (e-mail: ioztekin@ku.edu.tr).

To our knowledge, this is the first study of using final prediction errors in a space-domain with the local relational structures. The classification accuracy corresponding to optimum mesh size is compared with both the average accuracy of different mesh sizes and the accuracy of classical MVPA methods, suggested in [1], [2] and [3]. The simplicity and efficiency of the proposed method has the potential to be used in future brain decoding research.

II. MATERIALS AND METHODS

A. *fMRI experiment and pre-processing*

In the fMRI experiment, neural activation is examined in two working memory tasks, namely item recognition and judgment of recency (JOR) [15]. For both tasks, participants studied a list of five consonants, presented sequentially for 500 ms each. Following the study list, a visual mask/ task cue was shown for 750 ms, indicating the upcoming memory judgment (IR or JOR). Then, participants were presented with two probe consonants. In JOR trials, both of these consonants are from the current list and the participants are expected to select the consonant which was presented more recently in the study list. In IR trials, one of these consonants are from the current list where the other one is new, and the participants are expected to indicate the one belonging to the study list (see; Fig.1). Participants had 3 sec to respond.

Image processing and data analysis were performed using SPM2 (www.fil.ion.ucl.ac.uk/spm/). In the preprocessing phase first, slice acquisition timing across slices are corrected. Next, images are realigned to the first volume in each run in order to correct for head movement. Functional and anatomical images are then normalized to a standard template EPI. Finally, images are smoothened using a 6-mm full-width half-maximum isotropic Gaussian kernel.

B. *Representation of the local connectivity of Voxels: Local Relational Features (LRF)*

In this study, at each time instant t_i, $i = 1,2,\ldots,N$, the intensity values of voxels $v(t_i,\bar{s}_j)$ at locations $\bar{s}_j$, $j = 1,2,\ldots,M$ are measured and each t_i is associated with a task label c_i where N represents the number of time samples and M represents the number of voxels. In the current experiment, the task label is either IR or JOR. fMRI measurements are represented in an NxM matrix, where N is the number of time samples and M is the number of voxels. Since the voxels are distributed in the brain in three dimensions, their location $\bar{s}_j$ is a three dimensional vector, where $\bar{s}_j = (x_j, y_j, z_j)$.

In the proposed local mesh model, there are two major approaches to select the *p-neighborhood* η_p of each voxel $(t_i,\bar{s}_j)$. In the first approach a neighborhood system is defined spatially [11] where the *p*-nearest neighbors are selected as the ones whose Euclidean distance of voxel coordinates are the smallest to the seed voxel. The second approach is to define functional neighborhood [16], where the *p*-nearest neighbors are selected based on the functional connectivity between the surrounding voxels and the seed voxel. A popular functional connectivity measure is Pearson correlation as defined in [16].

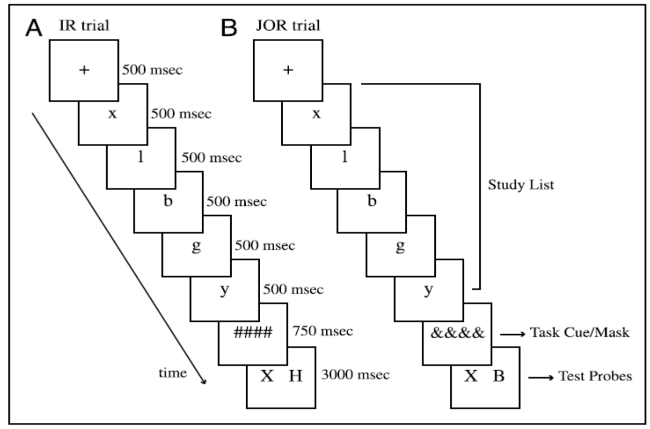

Fig. 1, taken from Oztekin et al., 2009. A sample sequence for an experimental trial composed of item recognition (IR) and judgment of recency (JOR) tasks. Each trial starts with a fixation point and then a study list of five letters is presented for both tasks. After the presentation of study list, the participants are shown a visual mask that cued either an IR or JOR trial. Then two test probes are presented, and participants executed either an IR (shown on panel A) or a JOR (shown on panel B) judgment.

In the local mesh model, a seed voxel is connected to its p-neighbors $\{v(t_i,\bar{s}_k)\}_{k=1}^{p}$ with arc weights $a_{i,j,k}$ (Fig. 2) and the p value is called the mesh size.

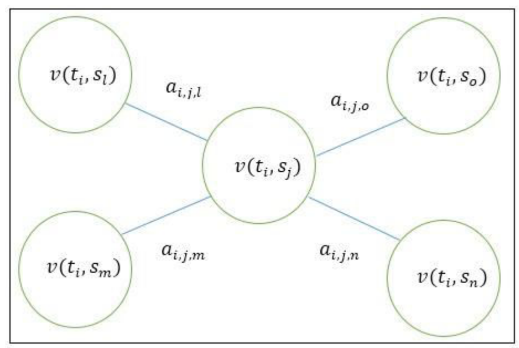

Fig. 2. Mesh diagram which represents a seed voxel $v(t_i,\bar{s}_j)$ and its p-nearest neighbors $\{v(t_i,\bar{s}_k)\}_{k=1}^{p}$ at a time instant t_i. The estimated arc weights $a_{i,j,k}$ represent the relationship between the seed voxel and its neighbors.

These arc weights are estimated using the linear regression equation:

$$v(t_i,\bar{s}_j) = \sum_{\bar{s}_k \in \eta_p} a_{i,j,k}\, v(t_i,\bar{s}_k) + \varepsilon_{i,j,p} \,. \qquad (1)$$

In (1), arc weights are estimated by minimizing the squared error ${\varepsilon_{i,j,p}}^2$ with Levinson – Durbin recursion [17]. These arc weights $a_{i,j,k}$ represent the relationship between a voxel $v(t_i,\bar{s}_j)$ and the voxels in η_p. For each voxel, at each t_i, arc weights $a_{i,j,k}$ are estimated and form a $1xp$ mesh arc vector $\bar{a}_{i,j} = [a_{i,j,1}\; a_{i,j,2}\; \ldots\; a_{i,j,p}]$. Then, each voxel $v(t_i,\bar{s}_j)$ is represented with this mesh arc vector $\bar{a}_{i,j}$ and to represent a voxel, its relationships between its neighboring voxels are used instead of its own intensity value. During this operation, all the voxels are used as seed voxels once and around each of them a local mesh is created. Combining the mesh arc vectors for all time instants, an Nxp mesh arc vector $A_j = [\bar{a}_{1,j}\; \bar{a}_{2,j}\; \ldots\; \bar{a}_{N,j}]^T$ is constructed. Finally, by combining A_j

of each voxel, an N by pM feature matrix $F = [A_1\ A_2\ ...\ A_M]$ is constructed.

C. Akaike's Final Prediction Error (FPE)

One of the crucial steps in designing the proposed local mesh model is to determine an optimal mesh size p which maximizes the accuracy of the classifier. The size of the local mesh also represents the degree of connections of a voxel with its neighbors. In this study, we adopt an information theoretic criterion suggested by Akaike in [13].

In our proposed method, first we compute the squared error at each time instant t_i, for each mesh and for each mesh size p by using :

$$\varepsilon_{i,j,p}{}^2 = \left(v(t_i, \bar{s}_j) - \sum_{\bar{s}_k \epsilon \eta_p(\bar{s}_j)} a_{i,j,k}\, v(t_i, \bar{s}_k)\right)^2 . \quad (2)$$

By taking the average of squared errors for all time instants t_i and for all seed voxels $v(t_i, \bar{s}_j)$, we approximate the variance of the error for the mesh size p as follows :

$$E(\bar{\varepsilon_p}{}^2) \cong \frac{1}{N}\frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{M} \varepsilon_{i,j,p}{}^2 , \quad (3)$$

where $E(\cdot)$ is the expectation operator. This expected squared error is used to determine the Akaike's Final Prediction Error, FPE, in space-domain such that:

$$FPE_p = E(\bar{\varepsilon_p}{}^2)\left(\frac{M+p+1}{M-p-1}\right). \quad (4)$$

Note that in the above formulation, the first term $E(\bar{\varepsilon_p}{}^2)$ is a monotonically decreasing function of p, whereas the second term is a monotonically increasing function. Therefore, FPE is a convex function in terms of p and has a unique minimum. According to Akaike's pioneering work [13], this function gives us a measure about the model quality. Minimizing FPE with respect to the model order p gives us the optimum mesh size. In this study, FPE_p is computed for various mesh sizes and the p value which minimizes FPE_p criterion is selected.

III. Results

The proposed method was tested on each of the eight participants' data. Each participant's data consists of 240 training samples (120 samples for both IR task and JOR task) and 80 test samples (40 samples for both IR task and JOR task). Our region of interest (ROI) consists of 2030 voxels that were identified from a whole-brain voxel-wise contrast assessing the active voxels during the experiment, using a threshold of $p < .001$, uncorrected. Dimension of features increases linearly with the increase in the mesh size p. Accordingly, for each mesh size p, the number of features used in classification is p x 2030.

The optimum mesh size for each participant is estimated by computing the FPE criterion with the varying p value in the interval [2-25] in both training and test data. It is observed that the minimum FPE value takes place in this interval. Moreover for each p value, extracted arc weight vectors are used in classification and the classification accuracies are computed using the k-NN method. The value of k is estimated via cross validation in the training set for each participant. Table 1 presents the results for one participant. Classification performances presented in Table 1 indicate that the mesh sizes in the interval [2-25], FPE decreases to some point (where p = 23) and then starts to increase. As a result, our method estimates the mesh size as 23. The corresponding classification accuracy is 67% which represents the average performance computed across a range of mesh sizes. The average accuracy of the classifiers for the mesh sizes [2-25] is 61%. These accuracies might seem low for a 2-class classification task. However, it is important to note that training (i.e. encoding) phases are identical across the IR and JOR tasks, and participants were only cued about which retrieval operation they should perform after the study phase.

TABLE I. FPE and Classification Accuracy for Different Mesh Sizes

Mesh Size	***FPE***	***Classification accuracy with arc vectors***	**Mesh Size**	***FPE***	***Classification accuracy with arc vectors***
2	95,57	63%	14	78,78	63%
3	91,99	63%	15	78,83	61%
4	86,76	63%	16	78,40	58%
5	86,25	62%	17	78,59	58%
6	85,55	57%	18	78,33	59%
7	82,96	65%	19	78,24	58%
8	81,87	65%	20	77,97	59%
9	80,59	65%	21	78,02	57%
10	80,05	61%	22	77,94	57%
11	79,94	62%	**23**	**77,63**	**67%**
12	79,34	58%	24	77,76	61%
13	78,99	62%	25	77,82	57%

For each participant, the average accuracy is computed in an interval [2-25] of mesh size with the same manner. Furthermore, the features obtained after the pre-processing step are directly fed to the classifier and the classification accuracies are computed to show that using arc vectors improves the performance on the data compared to the classical MVPA methods.

TABLE II. Classification Accuracies for Estimated Mesh Size and Classical MVPA Method

Participants	***Estimated optimum mesh size***	***k-NN Accuracy***		
		Classical MVPA method	***Classification with arc vectors using estimated mesh size***	***Average accuracy of the classifiers for p ϵ [2-25]***
Participant 1	17	58%	66%	61%
Participant 2	23	58%	67%	61%
Participant 3	24	62%	60%	61%
Participant 4	25	53%	58%	57%
Participant 5	23	54%	59%	57%
Participant 6	16	53%	59%	57%
Participant 7	25	57%	56%	55%
Participant 8	17	57%	58%	57%

Table II provides a comparison across the classification accuracy of the proposed method and classical MVPA methods. For each participant, the estimated mesh size differs in an interval [16 - 25] where these values correspond to the minimum of FPE on test data. When the arc vectors of the estimated mesh size are used in classification, 6 participants among 8 give 1%-9% increase in the classification performance, but for two of them, the performance decreases 1%-2% compared to the classical MVPA method.

In addition, the average accuracies for different mesh sizes selected in the interval [2-25], are provided in Table II for each participant. We observe that the accuracy corresponding to the estimated mesh sizes are always higher than that of the average accuracy over all mesh sizes. Note that the proposed method cannot detect the mesh size with the highest accuracy for participant 3 and participant 7. In Table III it can be seen that selecting the mesh size as 25 for participant 3 and 22 for participant 7 gives the highest accuracy and is better than the accuracy of classical MVPA approach.

TABLE III. CLASSIFICATION ACCURACIES FOR MESH SIZES WITH BEST ACCURACIES AND CLASSICAL MVPA METHOD

Participants	***Mesh size that gives best accuracy***	***k-NN Accuracy***	
		Classical MVPA method	***Classification with arc vectors using mesh size that gives the best accuracy***
Participant 3	25	62%	65%
Participant 7	22	57%	58%

Finally, we investigate the effect of the mesh sizes on the classification accuracy of a specific task. For this purpose, we estimate the optimum mesh size for each task, minimizing the FPE measure of (4), using only the samples which belong to the same class. Surprisingly, we observe that for this particular experimental set-up, the optimal p value for both classes remain the same as the optimal value computed for each participant. Therefore, we conclude that although distribution of information changes from participant to participant, it does not change from class to class.

IV. CONCLUSION

In this paper we propose a new machine learning approach to explore how information is represented in the brain for during cognitive processing. Our method introduces a local mesh model of varying size to represent each voxel by its linear relations with the neighboring voxels. The size of the local meshes is computed by minimizing an information theoretic criterion, namely Final Prediction Error (FPE). Since the optimum mesh size greatly differs from participant to participant, our approach provides a generic method to select the optimum mesh size. We showed that FPE, which depends on the number of available samples, error variance and model order that comes from the linear regression function, can be used to determine the mesh size, and serve as an effective tool to model neural activity during cognitive processing.

Our focus in this study was to find the optimum mesh sizes for different participants to investigate the participant dependency on the underlying cognitive task. We observed that the connectivity degrees of voxels highly depend on the individual participants and it is not possible to use a generic p value which is valid for all the participants. The results on 8 participants show that the FPE criterion is quite promising, where for 6 of 8 participants (75%) it is able to detect the optimum mesh size. Although it fails to detect the optimum mesh sizes in 2 participants, the classification accuracies differ only 1%-2% from the classical MVPA method. In this study, the proposed method is tested on item recognition (IR) and judgment of recency (JOR) tasks. In order to reach a more generic success using this method, further research will focus on implementing the same method for different cognitive tasks among multiple participants.

REFERENCES

[1] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten and P. Pietrini, “Distributed and overlapping representations of faces and objects in ventral temporal cortex,” *Science*, 293: 2425–2430, 2001.

[2] D. D. Cox, and R. L. Savoy. “Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex,” *Neuroimage*, 19: 261–270, 2003

[3] J.-D. Haynes and G. Rees, “Decoding mental states from brain activity in humans,” *Nature reviews. Neuroscience*, vol. 7, no. 7, pp. 523-34, Jul. 2006

[4] J.-D. Haynes, “Decoding and predicting intentions,” *Annals of the New York Academy of Sciences*, 1224: 9–2, 2011

[5] A. J. O’Toole, F. Jiang, H. Abdi, N. Penard, J. P. Dunlop and M. A. Parent, “Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data,” *Journal of Cognitive Neuroscience,* 19: 1735–1752, 2007.

[6] N. Kriegeskorte, “Pattern-information analysis: from stimulus decoding to computational-model testing,” *Neuroimage*, 56:411–421, 2011

[7] N. Kriegeskorte, R. Goebel and P. Bandettini, “Information-based functional brain mapping,” *Proceedings of the National Academy of Sciences of the United States of America*, 103: 3863–3868, 2006

[8] P. A. Bandettini, “What's new in neuroimaging methods?” *Annals of the New York Academy of Sciences* , 1156: 260–293, 2009

[9] M. Mur, P. A. Bandettini and N. Kriegeskorte, “Revealing representational content with pattern-information fMRI--an introductory guide.” Social Cognitive and Affective Neuroscience, 4: 101–109, 2009

[10] Z. Yang, F. Fang and X. Weng, “Recent developments in multivariate pattern analysis for functional MRI ”, *Neuroscience Bulletin,* 28(4): 399–408, August 2012

[11] M. Ozay, I. Oztekin, U. Oztekin and F. T. Yarman Vural, and “A Mesh learning model for pattern analysis of brain activity” , *Human Brain Mapping,* under review.

[12] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, “Beyond mind-reading: multi-voxel pattern analysis of fMRI data.,” *Trends in cognitive sciences*, vol. 10, no. 9, pp. 424-30, Sep. 2006.

[13] H. Akaike, “Fitting autoregressive models for prediction,” *Ann. Inst. Stat. Math.,* vol.21 pp. 243-247, 1969.

[14] H. Akaike, “Statistical predictor identification,” *Ann. Inst. Stat. Math.,* vol.22 pp. 203-217, 1970.

[15] I. Oztekin, B. McElree, B. P. Staresina and L. Davachi, “Working Memory Retrieval: Contributions of the Left Prefrontal Cortex, the Left Posterior Parietal Cortex, and the Hippocampus,” *Journal of Cognitive Neuroscience* vol. 21 no. 3, pp. 581-593

[16] O. Firat, M.Ozay, I. Onal, I. Oztekin, and F. T. Yarman Vural, “A mesh learning approach for brain data modeling,” *IEEE 20th Conference on Signal Processing and Communications Applications (SIU)*, 2012

[17] P. P. Vaidyanathan, “The Theory of Linear Prediction”, *Morgan & Claypool Publishers*, 2008