

Clustering Inter-Subject Correlation Matrices in Functional Magnetic Resonance Imaging

Jukka-Pekka Kauppi, Iiro P. Jääskeläinen, Mikko Sams, and Jussi Tohka

Abstract—We present a novel clustering method to probe inter-subject variability in functional magnetic resonance imaging (fMRI) data acquired in complex audiovisual stimulus environments, such as during watching movies. We calculate voxel-wise inter-subject correlation matrices across individual subject fMRI time-series and cluster them over the cerebral cortex. We address correlation matrix clustering problem and modify a standard K -means algorithm to cope better with spurious observations. We investigate suitability of the modified K -means with hierarchical clustering based postprocessing to correlation matrix clustering with several artificially generated data sets. We also present clustering of fMRI movie data. Preliminary results suggest that our methodology can be a valuable tool to investigate inter-subject variability in brain activity in different brain regions, such as prefrontal cortex.

I. INTRODUCTION

One of the central questions in neuroscience is how similarly individuals process and experience real-life events. Although individual experiences are never identical, some people evidently share experiences more than other ones. Functional magnetic resonance (fMRI) studies, where several subjects are exposed to the same natural stimulus while their brain activity is measured, may enable the investigation of variability in brain function probably underlying subjective experiences.

There is an increasing number of studies wherein a movie is shown to several subjects while their brain activity is measured using fMRI [1], [2]. A successful approach to analyze this multidimensional data is to compare inter-subject blood-oxygen level dependent (BOLD) signal activity by calculating voxel-wise n by n correlation matrices between fMRI time-series of n subjects. Voxel-wise statistical testing of average of the correlation matrix elements provide information whether responses between subjects are similar or not [2].

The drawback of the ISC analysis is that it does not take into account the variability of the subjects – whereas it is

common that many subjects have highly similar BOLD signal activity in the sensory brain areas such as visual and auditory cortex while they are exposed to a complex stimulus such as a movie, it is also typical that the same subjects respond differently to these stimuli in brain areas processing higher-order information, such as prefrontal cortex. Due to high variation in brain activity among subjects in these brain areas, investigation of average ISCs among subject pairs becomes inconvenient. However, it would be highly beneficial to investigate ISCs also in higher-order areas because high ISC values may indicate that subjects share aspects of, e.g., emotional experiences, social cognition or moral decision-making. In addition, ISC analysis provides a data-driven way for localizing neuronally active brain regions, since high ISC may imply stimulus-related neuronal processing.

In this work, we present a novel concept of clustering ISC matrices across the brain voxels. Unlike the average ISC, correlation matrix preserves the structure of variability among all subject pairs, allowing more sophisticated analysis of inter-subject similarities. We use clustering across the brain voxels to highlight those brain regions where the structure of ISC matrix is relatively constant. This makes analysis results easier to interpret and increases robustness of the method against random effects in the data.

One difficulty in clustering complex fMRI data is that all brain regions are not active during an fMRI measurement. In our context, this means that the number of spurious ISC matrices (i.e., spurious observations) is very high. Errors and inaccuracies in stereotactic registration increase the number of spurious observations further, suggesting the use of highly robust clustering techniques to alleviate ISC matrix clustering problem. Because typical outlier detection techniques (see, for example [3]) are designed to handle a reasonably low number of outliers, we here propose a modified K -means clustering algorithm to better deal with a very high number of spurious observations, and further enhance clustering with appropriate postprocessing. We investigated the performance of our method first with simulations and then clustered the real fMRI data set of 12 subjects watching a drama movie (“Crash”, Lions Gate Films, 2004).

Previously, clustering of fMRI data has been presented in the context of functional connectivity analysis aiming to parcellate the brain into several functional units (see, for example [4]-[7]). The main difference between these studies and our approach is that we perform clustering based on similarity in ISC matrices generated from n subject’s time-

Manuscript received July 15, 2010. This work was supported in part by the Academy of Finland (grants 130275 and 129657, The Finnish Centre of Excellence Program 2006 - 2011), and in part by Tampere Graduate School in Information Science and Engineering (TISE).

Jukka-Pekka Kauppi (e-mail: jukka-pekka.kauppi@tut.fi) and Jussi Tohka (jussi.tohka@tut.fi) are with the Department of Signal Processing, Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland.

Iiro P. Jääskeläinen (e-mail: iirojaaskelainen@yahoo.com) and Mikko Sams (e-mail: mikkosams@mac.com) are with the Department of Biomedical Engineering and Computational Science (BECS), The Aalto University School of Science and Technology, Espoo, Finland.

series instead of similarity in time-series of a single subject.

II. CLUSTERING METHOD

A. *K-means clustering of correlation matrices*

Our task is to cluster correlation matrices $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p\} \subset \mathbf{R}^{n \times n}$ originating from fMRI time-series of p brain voxels calculated voxel-by-voxel basis across n subjects. We can use K -means to partition p observations into K clusters D_1, \dots, D_K by minimizing the following criterion function:

$$Q(D) = \sum_{j=1}^K \sum_{\mathbf{x}_i \in D_j} \|\mathbf{x}_i - \mathbf{M}_j\|_F^2, \quad (1)$$

where $\|\cdot\|_F$ denotes a Frobenius norm and $\mathbf{M}_j \in \mathbf{R}^{n \times n}$ is the mean of the j 'th cluster:

$$\mathbf{M}_j = \frac{1}{|D_j|} \sum_{\mathbf{x}_i \in D_j} \mathbf{x}_i. \quad (2)$$

Let $\mathbf{x}_i = \text{vec}(\mathbf{X}_i)$ and $\mathbf{m}_j = \text{vec}(\mathbf{M}_j)$ denote vectorized matrices where the elements of $\mathbf{x}_i, \mathbf{m}_j \in \mathbf{R}^{n(n-1)/2}$ consist of the upper (or lower) triangular part of the symmetric matrices \mathbf{X}_i and \mathbf{M}_j , respectively. Minimizing (1) leads to the same solution as minimizing the within-cluster sum-of-the-squares (WCSS) criterion using the standard K -means approach (for details of the K -means algorithm, see for example [8], [9]) after vectorizing correlation matrices. The mean vectors \mathbf{m}_j can be converted into correlation matrices after clustering because \mathbf{M}_j is always a valid correlation matrix: a symmetric matrix is a correlation matrix if and only if it is positive semi-definite and has 1's in the diagonal [10]. Clearly, \mathbf{M}_j has 1's in the diagonal. Moreover, \mathbf{M}_j is symmetric and positive semi-definite as a (scaled) sum of symmetric and positive semi-definite matrices. Note that when using common data transformations to enhance clustering quality, cluster centers remain symmetric but, depending on the transformation, may not remain positive semi-definite. Therefore, after clustering the transformed data, we calculate final means based on original correlation matrices.

B. *Modified K-means*

The standard K -means is not well-suited to our problem because high number of spurious observations in the data – not belonging to any cluster – would bias the mean estimates. For this reason, we propose modification of the standard algorithm that utilizes only the densest regions in each cluster to update the mean estimates. There are several techniques that can be used to find densest regions in the data (see, for example [11]). We use Euclidean distance to the C 'th nearest neighbor as a measure of density because it does not require any distributional assumptions about the cluster structure. More specifically, we define density of an observation $\mathbf{x}_i \in D_j$ to be the radius of the smallest \mathbf{x}_i centered sphere containing C samples $\mathbf{x}_q \in D_j$. There are no strict rules how to select C and the number of densest data points M

used to update cluster means, but it is natural to select these parameters according to the smallest cluster size of interest. Throughout this work, we use $M = C = 30$ as we assume that clusters smaller than this size are not very interesting to analyze. If some cluster contains less than M data points during the K -means iterations, we update the corresponding mean using all samples in that cluster. The modified K -means algorithm is summarized below. Because correlation matrix clustering problem is equivalent to the problem of clustering vectorized matrices using the WCSS criterion, we use standard vector notation for simplicity.

Step 1. Select K initial cluster centers: $\mathbf{m}_1^{(1)}, \mathbf{m}_2^{(1)}, \dots, \mathbf{m}_K^{(1)}$.

Step 2. Generate a new partition $D^{(t)} = \{D_1^{(t)}, D_2^{(t)}, \dots, D_K^{(t)}\}$ at iteration t by assigning each observation to its closest cluster center:

$$D_j^{(t)} = \{\mathbf{x}_i : \|\mathbf{x}_i - \mathbf{m}_j^{(t)}\|_2 \leq \|\mathbf{x}_i - \mathbf{m}_{j_0}^{(t)}\|_2, j_0 = 1, \dots, K\}.$$

Step 3. Select a subset $\hat{D}_j^{(t)} \subseteq D_j^{(t)}$ of M densest observations for $j = 1, \dots, K$. If $|D_j^{(t)}| \leq M$, select $\hat{D}_j^{(t)} = D_j^{(t)}$.

Step 4. Compute new cluster centers as the means of the densest data points in each cluster:

$$\mathbf{m}_j^{(t+1)} = \frac{1}{|\hat{D}_j^{(t)}|} \sum_{\mathbf{x}_i \in \hat{D}_j^{(t)}} \mathbf{x}_i$$

Step 5. Evaluate the WCSS criterion with respect to dense observations:

$$J(\hat{D}^{(t)}) = \sum_{j=1}^K \sum_{\mathbf{x}_i \in \hat{D}_j^{(t)}} \|\mathbf{x}_i - \mathbf{m}_j^{(t)}\|^2.$$

Stop if $|J(\hat{D}^{(t)}) - J(\hat{D}^{(t-1)})| < \varepsilon$, where ε is a small positive constant. Otherwise, increase t by one and go to step 2.

In order to avoid computation of a full distance matrix (with $p(p-1)/2$ computations) when evaluating densities, we compute densities within each region D_j as clustering evolves. With this approach, the number of distances that need to be computed in the beginning of the algorithm is $p(p/K-1)/2$, assuming that all K regions are initially same size. At every successive K -means iteration, additional distance computations are required to update the densities. If distance matrices for all regions are stored in memory after each iteration, the number of distances that have to be calculated in successive iterations depends on the amount of change in the partition. More specifically, the number of distances that need to be computed to update the densities in a region $D^{(t)}$ is $(ab + b(b-1))/2$, where $a = |D^{(t-1)} \cap D^{(t)}|$ and $b = |D^{(t)} - D^{(t-1)}|$. Naturally, as the algorithm evolves, the

update process gets faster as b becomes smaller.

Note that although our modified K -means procedure requires calculation of point densities, steps 2, 4 and 5 in the algorithm are very fast even for very large data sets, since typically $M \ll p$. For the same reason, considerably fewer iterations are required for the modified algorithm to converge compared to the standard K -means. These properties make the modified K -means applicable to cluster even very large fMRI data sets.

C. Data preprocessing

We investigated the effects of different data transformations to clustering quality with synthetic data sets and then utilized this knowledge to select a suitable data transformation to real data. Fisher's z transformation [12] has a property of stabilizing the variance of the correlation coefficients. Under the assumption that individual correlation coefficients arise from normally distributed samples, transformed coefficients are asymptotically distributed according to a normal distribution with a variance independent of the true correlation value. We assumed that this transformation is useful because the K -means works best for spherical Gaussian clusters [13]. Moreover, we investigated the effect of standardizing samples by subtracting sample-wise means from each element and then dividing each element by the sample-wise standard deviation. We also investigated the effect of both transformations, where coefficients were first transformed using the Fisher's z and then standardized.

D. Postprocessing

One of the major questions with iterative partitive clustering techniques such as K -means is how to choose the number of clusters. A typical approach is to run K -means several times using different K and use some criterion to select the best value. However, this approach is less prominent if the data contains high number of spurious samples inherently belonging to no cluster. Therefore, instead of trying to optimize the model for specific selection of K , we run K -means using reasonably high value of K and then reduce the number of clusters using appropriate postprocessing. No exact rules how to select the value for K can be given, but the value should be high enough so that the true clusters are well represented in the initial partition. After applying the modified K -means, we merge highly similar clusters using single linkage agglomerative hierarchical clustering (HC) [9]. We adopt a technique proposed in [7] and use correlation coefficient similarity between the elements of two cluster centers as a merging criterion.

After merging stage, we sharpen the cluster centers and enhance spatial localization of the clusters by removing those observations from each cluster, which are highly dissimilar to the corresponding centers. Also here, we use correlation similarity to define spurious observations. As a final postprocessing step, we remove clusters with very low number (less than M) of samples.

III. MATERIAL

We first validated the proposed clustering approach with artificially generated data sets and then used our technique to

cluster real fMRI experiment data. We generated two types of artificial data sets to investigate the performance of the method: 1) data sets containing Gaussian clusters contaminated with uniformly distributed random noise to evaluate performance of the modified K -means against the standard K -means algorithm, and 2) artificial correlation matrix clusters to investigate the suitability of the technique to clustering ISC matrices. To ensure that our synthetic correlation matrix data sets were realistic, we selected center for each correlation matrix cluster from the real fMRI experiment data. In the following, we start by describing the experimental fMRI data in order to better explain our artificial correlation matrix generation approach.

A. fMRI experimental data

Functional MRI was measured from $n = 12$ subjects while they watched a drama movie "Crash" (directed by Paul Haggis, Lions Gate Films, 2005; for a detailed description of the experiment, see [14]). Preprocessing of the data included motion correction, fMRI time-series detrending, spatial smoothing, and registration of the functional images to a common stereotactic template (a detailed description of the preprocessing operations can be found in [2]). ISC matrices of the size 12 by 12 were calculated across 14-min individual fMRI recordings (time-series consisted of $l = 244$ samples) across the whole cerebral cortex. The cerebral cortex was defined according to the Harvard-Oxford probabilistic cortical atlas leading to $p = 54,866$ correlation matrices to be clustered (the atlas is provided together with the FSL software package; for more details of the FSL, see [15]). The modified K -means algorithm was run 40 times from random initial conditions using $K = 100$. The best solution was selected according to the lowest WCSS across the whole data set. In the postprocessing stage, we used correlation coefficients 0.7 and 0.8 for cluster merging and for discarding spurious samples, respectively. Validity of the parameters was verified by visual inspection – with the selected parameters, obtained clusters were spatially compact and element-wise dispersion of the correlation coefficients within the clusters was low. In this initial study, we strictly wanted to avoid assigning spurious matrices in any of the clusters to get the overview where the most compact ISC matrix clusters are located. After postprocessing, the final number of clusters was 78 with the median cluster size of 61 voxels. Altogether 48440 samples were discarded.

B. Synthetic correlation matrix data

To validate our method with artificial correlation matrix data sets, we randomly selected correlation matrices originating from real fMRI time-series and used them as cluster centers in our simulation experiments. The observations corresponding to each cluster center (\mathbf{M}) were generated by a 3-step procedure: 1) drawing $n = 12$ l -dimensional random vectors, each modeling an fMRI time course, independently from the standard normal distribution resulting in a random 12 by l matrix \mathbf{T} ; 2) computing the product $\mathbf{L}\mathbf{T}$, where the

upper triangular matrix \mathbf{L} was obtained by calculating the Cholesky factorization of \mathbf{M} as $\mathbf{M} = \mathbf{L}^T \mathbf{L}$, to produce the desired ISCs between the time series; 3) calculating the 12 by 12 correlation matrix of these artificially generated and correlated l -dimensional time-series. To investigate clustering performance in a variety of situations, we randomly selected cluster centers across the cortex several times and generated artificial clusters based on them as described above. We generated altogether 100 data sets each containing 10 clusters with 50 samples using several time-series lengths. Note that there is an inverse relationship between the time series length and dispersion of the cluster, because short time-series introduces more uncertainty to individual correlation values. In the simulations, we expressed clustering results with respect to the length of the time-series rather than cluster dispersion because it was more interesting to investigate the effect of time-series length to the clustering quality. Such analysis can be especially useful if one wishes to cluster fMRI recordings with reasonably short duration.

We used the adjusted rand (AR) index [16] to evaluate the clustering results with simulated correlation matrix data sets. AR index assesses the degree of agreement between the two partitions. If the agreement between the true partition and the obtained partition after clustering is perfect, AR index equals 1. The expected value of the AR index is 0 if there is no agreement between the two partitions. Several empirical studies have suggested the use of AR index for external cluster validation, see for example [17].

C. Synthetic Gaussian data

To compare the performance of the modified K -means against the standard K -means algorithm, we generated 100 two-dimensional data sets each containing 10 Gaussian clusters with random means selected uniformly from the range from -5 to 5 and variance $\sigma^2 = 0.08$. To allow reliable evaluation of the clustering (see our performance measures below), we ensured that any two cluster centers were at least 4σ apart from each other in all data sets. The number of samples in each cluster was selected uniformly from the range from 30 to 120. We added uniformly distributed random samples to the data sets to investigate robustness of both algorithms against spurious observations. Both algorithms were run 10 times from random initial conditions using $K = 10$, and the best estimates according to the lowest WCSS were assessed. For each true mean, the error distance was calculated with respect to its closest mean estimate, and the average of these distances was reported. We also reported the number of found clusters by associating estimated means to their closest clusters. For each estimated mean, closest cluster was determined by finding its 5 nearest observations and choosing the cluster containing most of these points. If the same number of data points was found for more than one cluster, the number of nearest neighbors was increased by one as long as a single cluster was found. If most of the nearest data points were noise points, the

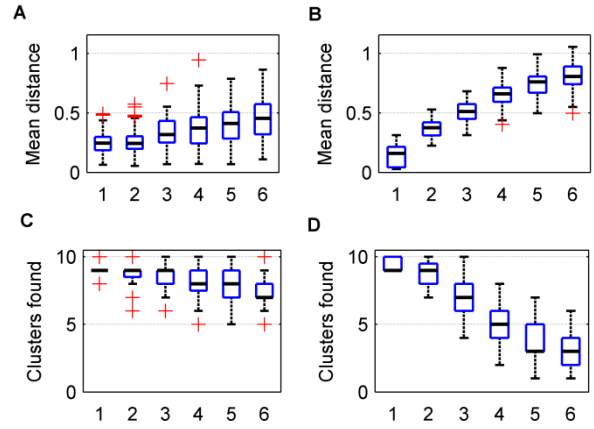


Fig. 1. Clustering results of the data sets containing 10 Gaussian clusters: (a) average of the distances between true cluster centers and their closest estimates using the modified K -means, (b) the corresponding results using the standard K -means, (c) the number of true clusters found by using the modified K -means, and (d) the corresponding result by using the standard K -means. Each box plot represents results of 100 data sets generated with different numbers of spurious samples. The numbering corresponds to the following fractions of spurious samples w.r.t. the total number of samples in clusters: 1) 0, 2) 0.5, 3) 1.0, 4) 2.0, 5) 3.0, and 6) 4.0.

corresponding mean estimate was not assigned to any cluster.

IV. RESULTS

A. Synthetic data

Fig. 1 shows clustering results for the Gaussian data sets using both the modified and the standard K -means. Results are presented using the standard box plot notation, where the central horizontal line in the middle of the boxes shows the median, and the edges of the boxes show the 25th and 75th percentiles of the selected performance measure. The whiskers show the range of the performance measure excluding outliers, and outliers are denoted individually using the cross-markers. Except for the ideal case without any spurious samples in the data (box plot 1), the mean estimates using the modified K -means (Fig. 1(a)) are consistently better when compared to the standard K -means (Fig. 1(b)). Note much faster degradation in the mean estimates in the case of standard K -means when the number of spurious samples increases. The number of true clusters found by the modified and the standard algorithm are shown in Figs. 1(c) and 1(d), respectively. Unlike the standard method, the modified K -means found most of the clusters even in the presence of very high number spurious observations. Because K -means is in general sensitive to the selection of the initial mean estimates, also the modified algorithm lost some of the clusters.

Fig. 2 presents the effect of preprocessing and time-series length to the clustering quality for artificial correlation matrix data sets. For $l \geq 40$, clustering was successful for most of the data sets (Figs. 2(a)-2(d)). Comparison of different preprocessing options reveals that Fisher's z transformation alone (Fig. 2(b)) did not improve clustering results compared to untransformed data (Fig. 2(a)) – instead, results were slightly worse when the transformation was

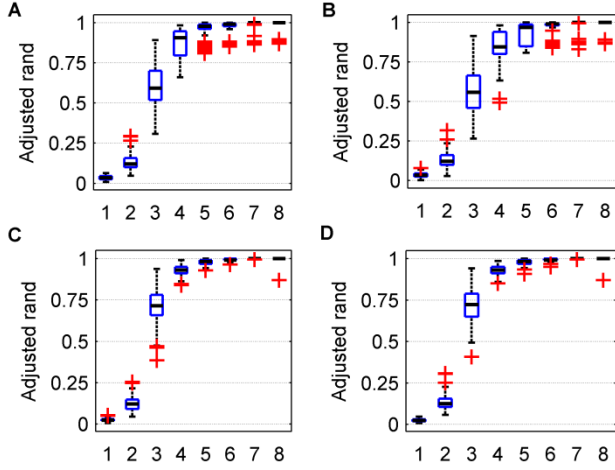


Fig. 2. The effect of different data transformations to the clustering performance for 100 artificial ISC matrix data sets (10 clusters, $K = 10$) using four different preprocessing options: (a) no preprocessing, (b) Fisher's z transform, (c) standardization, and (d) Fisher's z transform followed by standardization. Box plots correspond to data sets generated using different time-series lengths: 1) $l = 5$, 2) $l = 10$, 3) $l = 20$, 4) $l = 30$, 5) $l = 40$, 6) $l = 50$, 7) $l = 100$, and 8) $l = 200$. Standardization of the samples before clustering improved the clustering quality of some data sets notably.

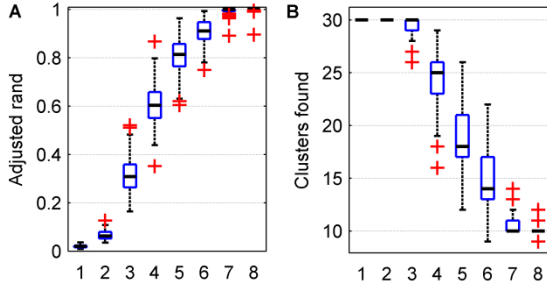


Fig. 3. Results for the simulated ISC matrix data sets using the modified K -means and cluster merging: (a) clustering performance, and (b) the number of clusters found. Each box plot corresponds to 100 data sets generated using time-series with different lengths. True number of clusters in each data set was 10. The modified K -means was run with $K = 30$ and cluster merging using the threshold 0.7 was applied. When $l \geq 100$, clustering was successful for most synthetic data sets.

used. Sample-wise standardization (Fig. 2(c)) improved the results considerably compared to the untransformed data, and the combined transformation (Fig. 2(d)) provided even slightly better result in some cases. For $l = 200$, perfect clustering was obtained for nearly all generated data sets after standardizing the samples.

Fig. 3 displays clustering results for the synthetic correlation matrix data sets after combining the modified K -means with cluster merging. The modified K -means was run using $K = 30$ (true number of clusters in each data set was 10) and cluster merging was performed using the threshold 0.7. Samples were preprocessed using Fisher's z transformation followed by standardization as suggested by the results shown in Fig. 2. For shorter time-series lengths, within-cluster dispersion of the clusters was so high that the cluster merging failed (see Figs. 3(a) and 3(b) for $l \leq 30$). When $l \geq 100$, the correct number of clusters was found in most data sets with perfect clustering quality.

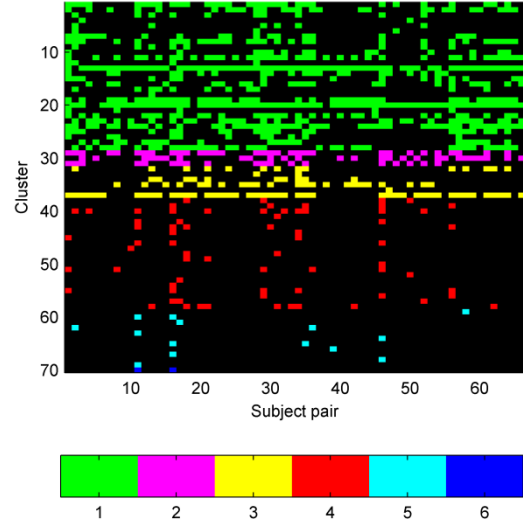


Fig. 4. The summary of ISC matrix clusters found in real fMRI experiment across the cerebral cortex. Statistically significant ($p < 0.001$) ISCs are shown in colors for all 66 subject pairs (non-significant ISCs are presented in black). The color coding corresponds to cortical regions where clusters are located: 1 = occipital lobe, 2 = parietal lobe, 3 = temporal lobe, 4 = frontal lobe, 5 = cingulate gyrus, and 6 = insular lobe. When moving from sensory areas towards higher order areas, the number of significant ISCs decreases drastically, but some of the subjects show high similarity also across these areas. Eight of the found 78 clusters did not show statistically significant ISC and they are excluded from the figure.

A. Real data

Fig. 4 shows the summary of the ISC matrix clusters found across the cerebral cortex for the real fMRI recordings. Each row represents elements of one correlation matrix cluster. The elements within each cluster are colored if the sample-wise median of the cluster shows statistically significant ISC ($p < 0.001$) under the null hypothesis that the correlation coefficient is zero. Approximate location of the clusters over the cortex is shown with the color coding. In visual, parietal and temporal cortex, many subject pairs showed statistically significant ISC. This finding is in line with the previous studies which have reported significant voxel-wise ISC in these areas during a movie stimulus [1], [2]. As assumed, there was much less ISC present as one moves from sensory areas towards higher order areas. Interestingly, however, certain subjects showed similar responses also across many higher order regions, including several areas in frontal cortex and cingulate gyrus (see for example subject pairs 11, 16, 33, and 47). The spatial organization of some of the clusters is shown in Fig. 5. The found clusters were spatially compact, possibly indicating that the corresponding brain areas of some subjects were active during the movie watching. Especially large cluster was found in precuneus. Note also that anterior cingulate gyrus is parcellated into subregions, corroborating the results of recent diffusion MRI studies [18]. Although more careful analysis and neuroscientific discussion of the results is out of the scope of this work, results in Figs. 4 and 5 demonstrate the potential of the presented ISC matrix clustering concept. For example, in the presence of some external information, it might be

possible to investigate brain mechanisms that are characteristic of certain group of subjects.

V. CONCLUSION

We presented a new correlation matrix clustering concept to investigate inter-subject variability of the brain activity in fMRI-based studies aiming to better understand how the human brain functions in natural real-world environments. We modified the standard K -means algorithm to better cope with spurious observations and showed that suitable data transformations and hierarchical clustering can be used to improve clustering quality. Results obtained by using real fMRI data indicated that the proposed methodology is suitable to investigate inter-subject variability in brain activity across fMRI recordings. Although we obtained compact clusters, important remaining methodological issue is how to improve initial mean estimates before applying clustering.

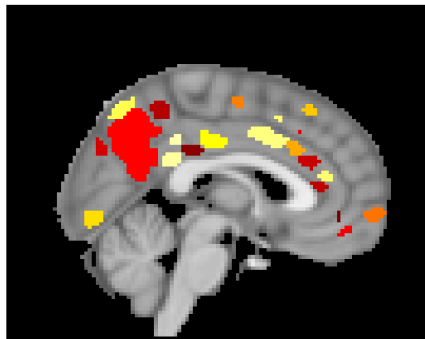


Fig. 5. Spatial organization of some ISC matrix clusters across the cerebral cortex. The sagittal slice is presented in MNI coordinate system ($x = -2$ mm). Clusters are visualized with arbitrary colors.

REFERENCES

- [1] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach, "Intersubject synchronization of cortical activity during natural vision," *Science*, vol. 303, no. 5664, pp. 1634-1640, March 2004.
- [2] J.-P. Kauppi, I. P. Jääskeläinen, M. Sams, and J. Tohka, "Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency," *Frontiers in Neuroinformatics* 4:5, 2010.
- [3] V. Barnett, and T. Lewis, *Outliers in Statistical Data*. 3rd ed. Wiley Series in Probability and Statistics, 1998.
- [4] R. Baumgartner, G. Scarth, C. Teichtmeister, R. Somorjai, and E. Moser, "Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: reproducibility," *Journal of Magnetic Resonance Imaging*, vol. 7, no. 6, pp. 1094-1101, 1997.
- [5] C. Goutte, P. Toft, E. Rostrup, F. Nielsen, and L. K. Hansen, "On clustering fMRI time series," *NeuroImage*, vol. 9, no. 3, pp. 298-310, March 1999.
- [6] Y. Golland, P. Golland, S. Bentin, and R. Malach, "Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems," *Neuropsychologia*, vol. 46, no. 2, pp. 540-553, January 2008.
- [7] W. Liao, H. Chen, Q. Yang, and X. Lei, "Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering," *IEEE Trans. on Medical Imaging*, vol. 27, no. 10, pp. 1472-1483, October 2008.
- [8] J. B. Macqueen, "Some methods of classification and analysis of multivariate observations," in *Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281-297.
- [9] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323, 1999.
- [10] G. Marsaglia and I. Olkin, "Generating correlation matrices," *SIAM Journal on Scientific and Statistical Computing*, vol. 5, no. 2, pp. 470-475, 1984.
- [11] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison-Wesley, March 2005.
- [12] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507-521, 1915.
- [13] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. 2nd ed. Wiley-Interscience, November 2000.
- [14] I. P. Jääskeläinen, K. Koskentalo, M. H. Balk, T. Autti, J. Kauramäki, C. Pomren, and M. Sams, "Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing," *Open Neuroimaging Journal*, vol. 2, pp. 14-19, 2008.
- [15] S. Smith, M. Jenkinson, M. Woolrich, C. Beckmann, T. Behrens, H. Johansenberg, P. Bannister, M. Deluca, I. Drobnjak, and D. Flitney, "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, pp. S208-S219, 2004.
- [16] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193-218, December 1985.
- [17] G. W. Milligan and M. C. Cooper, "A study of the comparability of external criteria for hierarchical cluster analysis," *Multivariate Behavioral Research*, vol. 21, no. 4, pp. 441-458, 1986.
- [18] M. Beckmann, H. Johansen-Berg, and M. F. Rushworth, "Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization," *Journal of Neuroscience*, vol. 28:29, no. 4, pp. 1175-90, January 2009.