

# DIFFUSION MAP FOR CLUSTERING FMRI SPATIAL MAPS EXTRACTED BY INDEPENDENT COMPONENT ANALYSIS

*Tuomo Sipola<sup>1,\*</sup>, Fengyu Cong<sup>1,†</sup>, Tapani Ristaniemi<sup>1</sup>, Vinoo Alluri<sup>1,2</sup>,  
Petri Toivainen<sup>2</sup>, Elvira Brattico<sup>2,4,5</sup>, Asoke K. Nandi<sup>3,1,‡</sup>*

<sup>1</sup>Department of Mathematical Information Technology,  
University of Jyväskylä, Finland

<sup>2</sup>Finnish Centre of Excellence in Interdisciplinary Music Research,  
University of Jyväskylä, Finland

<sup>3</sup>Department of Electronic and Computer Engineering,  
Brunel University, United Kingdom

<sup>4</sup>Cognitive Brain Research Unit, Institute of Behavioral Sciences,  
University of Helsinki, Finland

<sup>5</sup>Brain & Mind Laboratory, Biomedical Engineering and Computational Science,  
Aalto University, Espoo, Finland

## ABSTRACT

Functional magnetic resonance imaging (fMRI) produces data about activity inside the brain, from which spatial maps can be extracted by independent component analysis (ICA). In datasets, there are  $n$  spatial maps that contain  $p$  voxels. The number of voxels is very high compared to the number of analyzed spatial maps. Clustering of the spatial maps is usually based on correlation matrices. This usually works well, although such a similarity matrix inherently can explain only a certain amount of the total variance contained in the high-dimensional data where  $n$  is relatively small but  $p$  is large. For high-dimensional space, it is reasonable to perform dimensionality reduction before clustering. In this research, we used the recently developed diffusion map for dimensionality reduction in conjunction with spectral clustering. This research revealed that the diffusion map based clustering worked as well as the more traditional methods, and produced more compact clusters when needed.

**Index Terms**— clustering, diffusion map, dimensionality reduction, functional magnetic resonance imaging (fMRI), independent component analysis, spatial maps

---

\*Tuomo Sipola's work was supported by the Foundation of the Nokia Corporation and the Finnish Foundation for Technology Promotion.

†This work was partially supported by TEKES (Finland) grant 40334/10 Machine Learning for Future Music and Learning Technologies.

‡Asoke K. Nandi would like to thank TEKES for the award of the Finland Distinguished Professorship.

## 1. INTRODUCTION

In order to gain understanding about the human brain, various technologies have recently been introduced, such as electroencephalography (EEG), tomography, magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI). They provide scientists with data about the temporal and spatial activity inside the brain. Functional magnetic resonance imaging is a brain imaging method that measures blood oxygenation level. It detects changes in this level, that are believed to be related to neurotransmitter activity. This enables the study of brain functioning, pathological trait detection and treatment response monitoring. The method localises brain function well, and thus is useful in detecting differences in subject brain responses [1, 2].

Deeper understanding about the simultaneous activities in the brain begins with a decomposition of the data. Independent component analysis (ICA) has been extensively used to analyze fMRI data. It tries to decompose the data into multiple components that are mixed in the original data. Basically, there are two ways to perform ICA: group ICA and individual ICA [3]. Group ICA is performed on the data matrix including all the participants' fMRI data, and individual ICA is applied on each dataset of each participant. Among datasets of different participants, group ICA tends to need more assumptions which are not required by individual ICA [4]. For individual ICA, if the components for each participant are known, it is expected to find the most common components among the participants. Therefore, clustering spatial maps extracted by ICA is a necessary step for the individual ICA approach to

find common spatial information across different participants in fMRI research.

ICA decomposes the individual datasets and creates components that can be presented with spatial maps. After ICA has been applied, a data matrix of size  $n$  by  $p$  is produced, where  $n$  is the number of spatial maps and  $p$  is the number of voxels of each spatial map. The  $n$  spatial maps come from different participants, and  $n$  is much smaller than  $p$  in fMRI research. Clustering the spatial maps is mostly done using the  $n$  by  $n$  similarity matrix of the  $n$  by  $p$  data matrix [3, 5]. Surprisingly, it usually works well although such a similarity matrix inherently can just explain a certain amount of the total variance contained in the high-dimensional  $n$  by  $p$  data matrix [5].

New mathematical approaches for functional brain data analysis should take into account the characteristics of the data analyzed. As stated, spatial maps have high dimensionality  $p$ . In machine learning, dimensionality reduction is usually performed on such datasets before clustering. In the small- $n$ -large- $p$  clustering problem, the conventional dimensionality reduction methods, for example, principal component analysis (PCA) [6], might not be suitable for the non-linear properties of the data. In this research, we apply a recently developed non-linear method called diffusion map [7, 8] for dimensionality reduction. The probabilistic background of the diffusion distance metric will give an alternative angle to this dataset by facilitating the clustering task and providing visualization. This paper explores the possibility of using the diffusion map approach for fMRI ICA component clustering.

## 2. METHODOLOGY

This paper considers a dimensionality reduction approach to clustering of high-dimensional data. The clustering procedure flows as follows:

1. Data normalization with logarithm
2. Neighborhood estimation
3. Dimensionality reduction with diffusion map
4. Spectral clustering

Data normalization should be done if the features are on differing scales. This ensures that the distances between the data points are meaningful. Neighborhood estimation for diffusion map creates the neighborhood where connections between data points are considered. Dimensionality reduction creates a new set of fewer features that still retain most information. Spectral clustering groups similar points together.

We assume that our dataset consists of vectors of real numbers:  $X = \{x_1, x_2, \dots, x_n\}, x_i \in \mathbb{R}^p$ . In practice the dataset is a data matrix of size  $n \times p$ , whose rows represent the samples and columns the features. In this study each row vector is a spatial map and column vector contains the corresponding voxels in different spatial maps.

### 2.1. Diffusion map

Diffusion map is a dimensionality reduction method that embeds the high-dimensional data to a low-dimensional space. It is part of the manifold learning method family and can be characterized with its use of diffusion distance as the preserved metric [7].

The initial step of the diffusion map algorithm itself calculates the affinity matrix  $W$ , which has data vector distances as its elements. Here Gaussian kernel with Euclidean distance metric is used [7, 9]. For  $\epsilon$  selection, see below. The affinity matrix is defined as

$$W_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\epsilon}\right),$$

where  $x_i$  is the  $p$ -dimensional data point. The neighborhood size parameter  $\epsilon$  is determined by finding the linear region in the sum of all weights in  $W$ , while trying different values of  $\epsilon$  [10, 11]. The sum is

$$L = \sum_{i=1}^n \sum_{j=1}^n W_{i,j},$$

From the affinity matrix  $W$  the row sum diagonal matrix  $D_{ii} = \sum_{j=1}^n W_{ij}, i \in 1 \dots n$  is calculated. The  $W$  matrix is then normalized as  $P = D^{-1}W$ . This matrix represents the transition probabilities between the data points, which are the samples for clustering and classification. The conjugate matrix  $\tilde{P} = D^{\frac{1}{2}}PD^{-\frac{1}{2}}$  is created in order to find the eigenvalues of  $P$ . In practice, substituting  $P$ , we get

$$\tilde{P} = D^{-\frac{1}{2}}WD^{-\frac{1}{2}}.$$

This so-called normalized graph Laplacian [12] preserves the eigenvalues [9]. Singular value decomposition (SVD)  $\tilde{P} = U\Lambda U^*$  yields the eigenvalues  $\Lambda = \text{diag}([\lambda_1, \lambda_2, \dots, \lambda_n])$  and eigenvectors in matrix  $U = [u_1, u_2, \dots, u_n]$ . The eigenvalues of  $P$  and  $\tilde{P}$  stay the same. It is now possible to find the eigenvectors of  $P$  with  $V = D^{-\frac{1}{2}}U$  [9].

The low-dimensional coordinates in the embedded space  $\Psi$  are created using  $\Lambda$  and  $V$ :

$$\Psi = V\Lambda.$$

Now, for each  $p$ -dimensional point  $x_i$ , there is a corresponding  $d$ -dimensional coordinate, where  $d \ll p$ . The number of selected dimensions depends on how fast the eigenvalues decay. The coordinates for a single point can be expressed as

$$\Psi_d : x_i \rightarrow [\lambda_2 v_2(x_i), \lambda_3 v_3(x_i), \dots, \lambda_{d+1} v_{d+1}(x_i)]. \quad (1)$$

The diffusion map now embeds the data points  $x_i$  while preserving the diffusion distance to a certain bound given that enough eigenvalues are taken into account [7].

## 2.2. Spectral clustering

Spectral clustering is a method to group samples into clusters by benefitting from the results of spectral methods that reveal the manifold, such as the diffusion map. Spectrum here is understood in the mathematical sense of spectrum of an operator on the matrix  $P$ . The main idea is that the dimensionality reduction has already simplified the clustering problem so that the clustering itself in the low-dimensional space is an easy task. This leaves the actual clustering for any clustering method that can work with real numbers [13, 14, 15].

The first few dimensions from the diffusion map represent the data up to a relative precision, and thus contain most of the distance differences in the data [7]. Therefore, some of the first dimensions will be used to represent the data. Threshold at 0 in the embedded space divides the space between the possible clusters, which means that a linear classification can be used. With the linear threshold, the second eigenvector separates the data into two clusters in the low-dimensional space. This eigenvector solves the normalized cut problem, which means that there are small weights between clusters but the internal connections between the members inside the cluster are strong. Clustering in this manner happens through similarity of transition probabilities between clusters [13, 14, 16, 17].

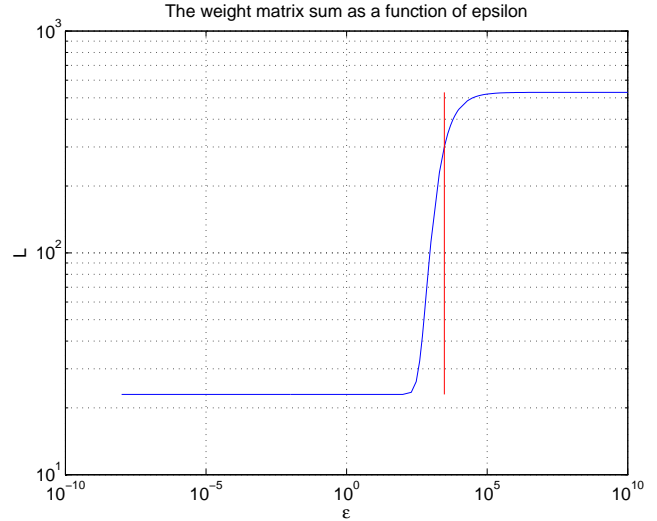
## 3. RESULTS

The data comes from experiments where participants listened to music. The data analysis was performed on a collection of spatial maps of brain activity. After dimensionality reduction and spectral clustering, the results are presented and compared to more traditional methods.

### 3.1. Data description

In this research the fMRI data are based on the data sets used by Alluri et al. [18]. Eleven musicians listened to a 512-second modern tango music piece during the experiment. In the free-listening experiment the expectation was to find relevant brain activity significantly correlating with the music stimulus. The stimuli were represented by musical features used in music information retrieval (MIR) [18].

After preprocessing, PCA and ICA were performed on each dataset of each participant, and 46 ICA components (i.e., spatial maps) were extracted for each dataset [19, 20]. Then, temporal courses of the spatial maps were correlated with one musical feature, Brightness [18]. As long as the correlation coefficient was significant (statistical  $p$ -value  $< 0.05$ ), the spatial maps were selected for further analysis. Altogether,  $n = 23$  spatial maps were selected for 11 participants. The number of voxels for each spatial map was  $p = 209,633$ . So, the 23 by 209,633 data matrix was used for the clustering to find the common spatial map across the 11 participants.



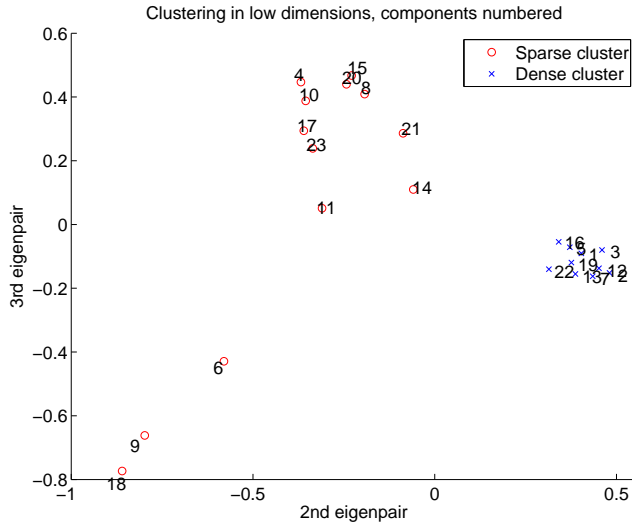
**Fig. 1.** Selecting  $\epsilon$  for diffusion map. The red line shows the selected value.

### 3.2. Data analysis

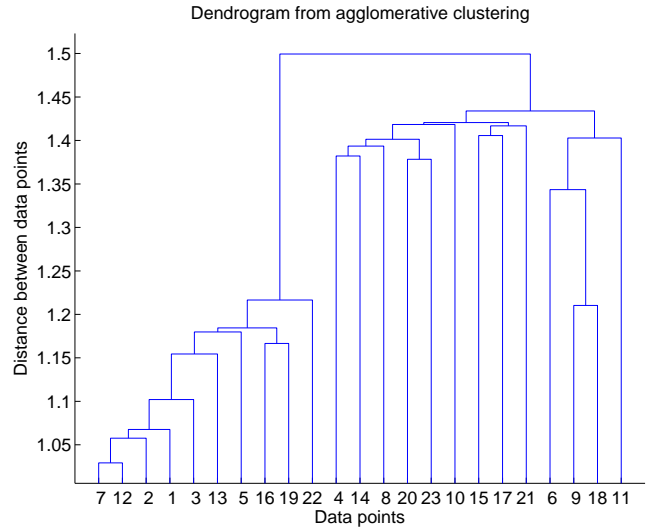
The data matrix was analyzed using the methodology explained in Section 2. The dimensionality of the dataset was reduced and then the spectral clustering was carried out. The weight matrix sum for  $\epsilon$  selection is in Figure 1; the used value is in the middle region, highlighted with straight vertical line. Clustering was performed with only one dimension in the low-dimensional space. To compare the results with more traditional clustering methods, the high-dimensional data was clustered with agglomerative hierarchical clustering [21] with Euclidean distances using the similarity matrix [5] and  $k$ -means algorithms [21]. The clustering results for two clusters were identical using all the methods.

Figure 2 shows the resulting clustering from the diffusion map. The figure uses the first two eigenpairs for low-dimensional presentation, for these two clusters even one dimension is enough. The spatial maps are numbered and the two clusters are marked with different symbols. The dividing spectral clustering line is at 0 along the horizontal axis, so the point to the right of 0 are in one cluster and to the left another. Two clusters, dense and sparse, are detected using this threshold. The dense cluster, marked with crosses, contains components that are considered to be similar according to this clustering. The traditional PCA and kernel PCA with Gaussian kernel for spectral clustering are compared to the diffusion map [22, 23]. In Figure 3 diffusion map with correct  $\epsilon$  creates more firm connections, which eases the clustering task. The effect of diffusion distance metric is also seen.

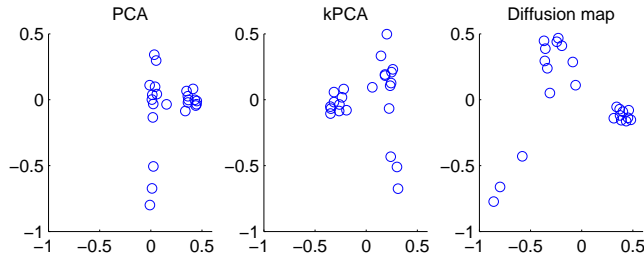
In Figure 4 the dendrogram produced by the agglomerative clustering is shown. The clustering results are the same as with the dimensionality reduction approach. The separation is visible at the highest level and the structure corresponds



**Fig. 2.** Diffusion map clustering results.



**Fig. 4.** Dendrogram from the agglomerative clustering.



**Fig. 3.** Dimensionality reduction method comparison. The coordinates have been scaled.

to the distances seen in Figure 2. All the points in, e.g., the dense cluster in Figure 2 are in the left cluster of Figure 4. This comparison shows the evident separation between the two clusters and also validates the results from diffusion map methodology.

Figure 5 illustrates the kind of spatial maps that are found in the dense cluster. Dark areas along the lateral sides is used to highlight those voxels whose values differed more than three standard deviations from the mean. The numbers marking the slices are their Z-coordinates. The corresponding low-dimensional point is in Figure 2 numbered as 3. It is now possible to inspect the clusters more closely with domain experts.

Figure 6 shows the correlation matrix of all the 23 spatial maps. This is a way to inspect the similarity of the brain activity. The correlation matrix is also the basis of analysis for the hierarchical clustering [5]. In the figure it can be seen that there is some correlation between some of the spatial maps, but not so much between others.

Figures 7 and 8 illustrate the internal structure of the clusters by showing the correlation matrices for the individual clusters. The members in the dense cluster have higher cor-

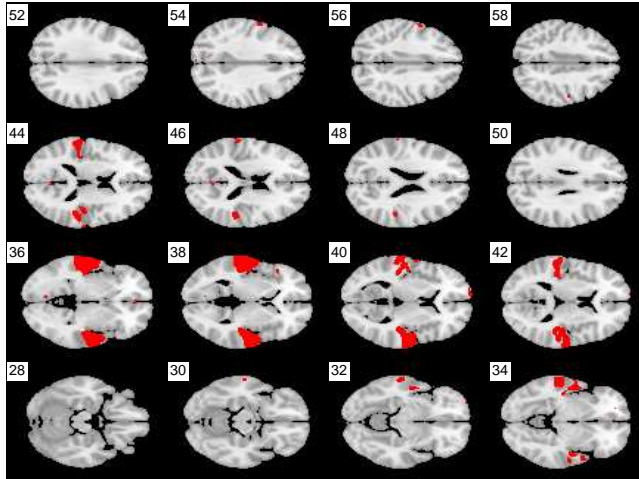
relation among themselves than the members in the sparse cluster. This information is also seen in Figure 2 where the diffusion distances inside the dense cluster are smaller.

#### 4. DISCUSSION

In this paper we have proposed a theoretically sound non-linear analysis method for clustering ICA components of fMRI imaging. The clustering is based on diffusion map manifold learning, which reduces the dimensionality of the data and enables clustering algorithms to perform their task. This approach is more suitable for high-dimensional data than just applying clustering methods that are designed for low-dimensional data. The assumption of non-linear nature of brain activity also promotes the use of methods designed for such problems. Particularly, the advantage of diffusion map is in visualizing the distribution of all data samples ( $n$  spatial maps with  $p$  voxels in each) by using only two coordinates. As seen in the visualization, it becomes more straightforward to determine the compact cluster from the two-dimensional plot derived from the 209,633-dimensional feature space than from the similarity matrix.

The results show that the proposed methodology separates groups of similarly behaving spatial maps. Results from diffusion map spectral clustering are similar to hierarchical agglomerative clustering and  $k$ -means clustering. Small sample size and good separation of clusters makes the clustering problem rather simple to solve. Moreover, the visualization obtained from diffusion map offers an interpretation for clustering.

The proposed methodology should be useful for analyzing the function of the brain and understanding which stimuli create similar spatial responses in which group of participants.



**Fig. 5.** Example spatial map in the dense cluster, this is data point number 3. Dark lateral areas mark more than three standard deviations from the mean, e.g. in slices 36 and 38.

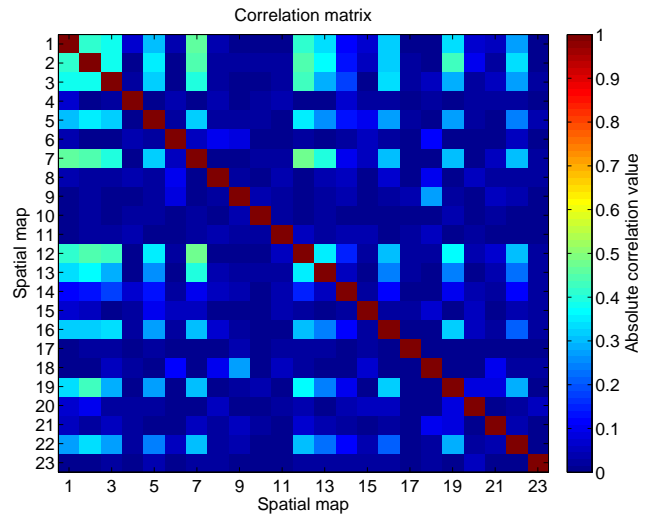
The domain experts can gain more basis for the interpretation of brain activity when similar activities are already clustered using automated processes suitable for the task. Furthermore, visualization helps to identify the relationships of the clusters.

Diffusion map execution times become increasingly larger if the number of samples goes very high. This can be overcome to a certain degree with out-of-sample extension. Big sample sizes are also a problem with traditional clustering methods. However, diffusion map offers a non-linear approach, and is suitable for high-dimensional data. Both properties are true for fMRI imaging data.

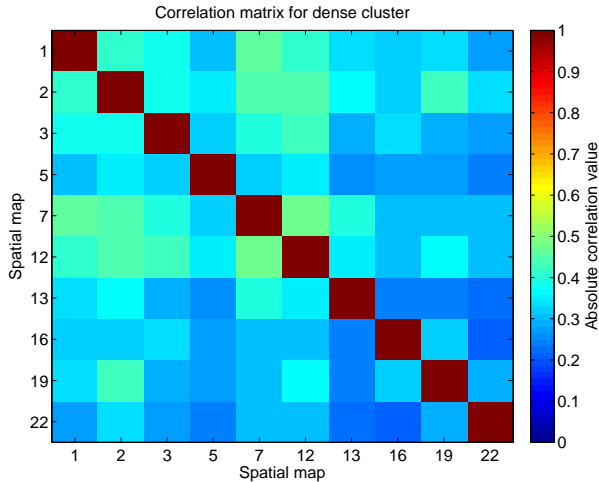
The analysis could be expanded to more musical features and to bigger datasets in order to further validate its usefulness in understanding the human brain during listening to music. The method is not restricted only to certain kind of stimulus, so it is usable with diverse fMRI experimental setups. Furthermore, situations where traditional clustering fails when processing spatial maps, the proposed methodology might give more reasonable results.

## 5. REFERENCES

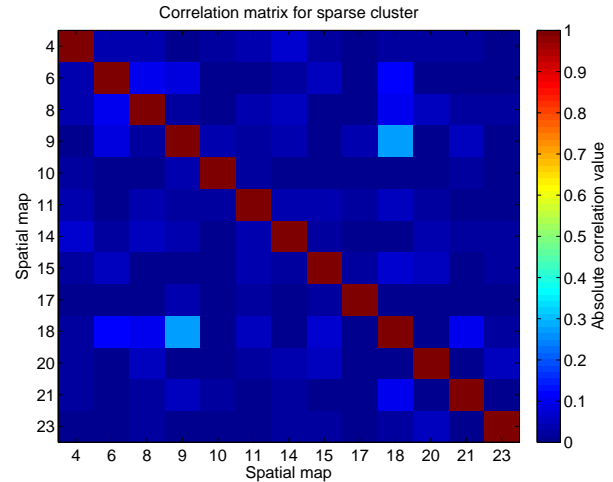
- [1] P M Matthews and P Jezzard, “Functional magnetic resonance imaging,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 1, pp. 6–12, 2004.
- [2] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional Magnetic Resonance Imaging*, Sinauer, Massachusetts, 2nd ed. edition, 2009.
- [3] Vince D Calhoun, Jingyu Liu, and Tülay Adalı, “A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data,” *Neuroimage*, vol. 45, no. 1 Suppl, pp. S163, 2009.
- [4] Fengyu Cong, Zhaoshui He, Jarmo Hämäläinen, Paavo H.T. Leppnen, Heikki Lyytinen, Andrzej Cichocki, and Tapani Ristaniemi, “Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection,” *Journal of Neuroscience Methods*, vol. 212, no. 1, pp. 165–172, 2013.
- [5] Fabrizio Esposito, Tommaso Scarabino, Aapo Hyvarinen, Johan Himberg, Elia Formisano, Silvia Comani, Gioacchino Tedeschi, Rainer Goebel, Erich Seifritz, Francesco Di Salle, et al., “Independent component analysis of fMRI group studies by self-organizing clustering,” *Neuroimage*, vol. 25, no. 1, pp. 193–205, 2005.
- [6] Ian Jolliffe, *Principal component analysis*, Springer Verlag, 2002.
- [7] Ronald R. Coifman and Stéphane Lafon, “Diffusion maps,” *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.
- [8] B. Nadler, S. Lafon, R.R. Coifman, and I.G. Kevrekidis, “Diffusion maps, spectral clustering and reaction coordinates of dynamical systems,” *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 113–127, 2006.
- [9] Boaz Nadler, Stéphane Lafon, Ronald Coifman, and Ioannis G. Kevrekidis, “Diffusion maps – a probabilistic interpretation for spectral embedding and clustering algorithms,” in *Principal Manifolds for Data Visualization and Dimension Reduction*, Timothy J. Barth, Michael Griebel, David E. Keyes, Risto M. Nieminen, Dirk Roose, Tamar Schlick, Alexander N. Gorban, Balázs Kégl, Donald C. Wunsch, and Andrei Y.



**Fig. 6.** Absolute correlation values between the spatial maps.



**Fig. 7.** Absolute correlation values between the spatial maps that belong to the dense cluster.



**Fig. 8.** Absolute correlation values between the spatial maps that belong to the sparse cluster.

Zinovyev, Eds., vol. 58 of *Lecture Notes in Computational Science and Engineering*, pp. 238–260. Springer Berlin Heidelberg, 2008.

- [10] R.R. Coifman, Y. Shkolnisky, F.J. Sigworth, and A. Singer, “Graph laplacian tomography from unknown random projections,” *Image Processing, IEEE Transactions on*, vol. 17, no. 10, pp. 1891–1899, oct. 2008.
- [11] Amit Singer, Radek Erban, Ioannis G. Kevrekidis, and Ronald R. Coifman, “Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 38, pp. 16090–16095, 2009.
- [12] F. R. K. Chung, *Spectral Graph Theory*, p. 2, AMS Press, Providence, R.I, 1997.
- [13] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss, “On spectral clustering: Analysis and an algorithm,” in *Advances in Neural Information Processing Systems 14*, 2001, pp. 849–856, MIT Press.
- [14] Ravi Kannan, Santosh Vempala, and Adrian Vetta, “On clusterings: Good, bad and spectral,” *J. ACM*, vol. 51, pp. 497–515, May 2004.
- [15] Ulrike von Luxburg, “A tutorial on spectral clustering,” *Statistics and Computing*, vol. 17, pp. 395–416, 2007.
- [16] Jianbo Shi and J. Malik, “Normalized cuts and image segmentation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.
- [17] Marina Meila and Jianbo Shi, “Learning segmentation by random walks,” in *NIPS*, 2000, pp. 873–879.
- [18] Vinoo Alluri, Petri Toiviainen, Iiro P Jääskeläinen, Enrico Glerean, Mikko Sams, and Elvira Brattico, “Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm,” *Neuroimage*, vol. 59, no. 4, pp. 3677–3689, 2012.
- [19] T. Puoliväli, F. Cong, V. Alluri, Q. Lin, P. Toiviainen, A. K. Nandi, E. Brattico, and T. Ristaniemi, “Semi-blind independent component analysis of functional MRI elicited by continuous listening to music,” in *International Conference on Acoustics, Speech, and Signal Processing 2013 (ICASSP2013)*, Vancouver, Canada, May 2013.
- [20] Valeri Tsatsishvili, Fengyu Cong, Tuomas Puoliväli, Vinoo Alluri, Petri Toiviainen, Asoke K Nandi, Elvira Brattico, and Tapani Ristaniemi, “Dimension reduction for individual ICA to decompose fMRI during real-world experiences: Principal component analysis vs. canonical correlation analysis,” in *European Symposium on Artificial Neural Networks 2013*, Bruges, Belgium, April 2013.
- [21] Rui Xu, Donald Wunsch, et al., “Survey of clustering algorithms,” *Neural Networks, IEEE Transactions on*, vol. 16, no. 3, pp. 645–678, 2005.
- [22] K-R Müller, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, and Bernhard Schölkopf, “An introduction to kernel-based learning algorithms,” *Neural Networks, IEEE Transactions on*, vol. 12, no. 2, pp. 181–201, 2001.
- [23] Quan Wang, “Kernel principal component analysis and its applications in face recognition and active shape models,” *arXiv preprint arXiv:1207.3538*, 2012.