

Detection of Spatial Activation Patterns as Unsupervised Segmentation of fMRI Data

Polina Golland¹, Yulia Golland², and Rafael Malach³

¹ Computer Science and Artificial Intelligence Laboratory, MIT, USA

² Department of Psychology, Hebrew University of Jerusalem, Israel

³ Department of Neurobiology, Weizmann Institute of Science, Israel

Abstract. In functional connectivity analysis, networks of interest are defined based on correlation with the mean time course of a user-selected ‘seed’ region. In this work we propose to simultaneously estimate the optimal representative time courses that summarize the fMRI data well and the partition of the volume into a set of disjoint regions that are best explained by these representative time courses. Our approach offers two advantages. First, it removes the sensitivity of the analysis to the details of the seed selection. Second, it substantially simplifies group analysis by eliminating the need for a subject-specific threshold at which correlation values are deemed significant. This unsupervised technique generalizes connectivity analysis to situations where candidate seeds are difficult to identify reliably or are unknown. Our experimental results indicate that the functional segmentation provides a robust, anatomically meaningful and consistent model for functional connectivity in fMRI.

1 Introduction and Motivation

In this paper we propose and demonstrate a new approach to detection and analysis of spatial patterns of activation from fMRI data. Functional connectivity analysis [4] is widely used in fMRI studies for detection and analysis of large networks that co-activate with a user-selected ‘seed’ region of interest. Time course correlation typically serves as a measure of similarity with the mean time course of the selected seed region. Since no alternative hypothesis for correlation values is formulated, the user must select a threshold, or significance level, used in rejecting the null hypothesis that assumes zero correlation. This approach is highly useful for analyzing experiment-specific fMRI data, but integration across different runs or across subjects is challenging due to high variability of inter-voxel correlation values across scans. Furthermore, in some studies it is unclear how to select the seed region, and we would instead prefer to discover the interesting ‘seeds’ and the associated networks in an unsupervised way. To eliminate the sensitivity of the analysis to the seed and threshold selection and to generalize the method to situations when candidate seeds are not immediately obvious, we propose to simultaneously estimate an optimal partition of the volume into a set of disjoint networks and the representative time courses associated with these networks. This formulation gives rise to an unsupervised segmentation algorithm

that, very much like EM-segmentation of anatomical scans [19], estimates segmentation labels by fitting a mixture density to the image data. The algorithm adaptively determines the threshold for assigning a voxel to a network based on the similarity of that voxel to the representative time courses, eliminating the need for subject-specific threshold selection.

Our approach is based on a model that parcelates the brain into disjoint sub-regions. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [3] provide an alternative model of functional connectivity that treats the data as a linear combination of components, i.e., spatial maps with associated time courses. The physiological evidence for either model is yet to be established, but we find the parcellation model more appealing in explaining functional organization, in particular when we extend the model to multiple scales.

ICA, PCA and clustering have been extensively explored in the contexts of regression-based detection [1, 2, 7, 8, 9, 13, 14, 16, 17]. Application of clustering in fMRI analysis has traditionally focused on grouping voxels into small, functionally homogeneous regions in paradigm-based studies [8, 13, 17]. In contrast, we aim to construct a top-down representation of global patterns of activation spanning the entire brain. Recently, clustering was also demonstrated in application to full-brain scans in rest state [5, 18], revealing anatomically meaningful regions of high functional coherency. Unlike prior work in clustering of fMRI data [8, 13, 17, 18], we do not aim to determine the optimal number of systems in the decomposition. Instead, our experience shows that active browsing of the segmentation results across several levels of resolution (system size) in the anatomical region of interest is substantially more instructive than considering flat parcellations generated for a fixed, large number of clusters. In addition, we perform clustering on the original time courses, replacing the dimensionality reduction step used in [18] by the constrained signal model that effectively fold the dimensionality reduction into the estimation process. The resulting algorithm is simple to implement and analyze, yet it produces highly stable results across runs and subjects.

ICA offers an unsupervised component-based decomposition of the spatio-temporal fMRI data, but the interpretation of the resulting component maps remains challenging. In particular, the method produces a flat decomposition into a large number of spatially sparse, typically non-overlapping, components which are often treated as a segmentation of the volume. We find it natural to explicitly formulate the problem of characterizing the spatial patterns of co-activation as segmentation of the fMRI volume. This model is well matched to the questions of interest in exploratory analysis of fMRI data, in addition to producing anatomically meaningful results that are easy to interpret as partitions of the cortex into systems.

2 Unsupervised Segmentation of fMRI Data

Classical correlation-based connectivity analysis assumes a user-specified hypothesis, for example through a selection of a seed region. In contrast, we

formulate the problem of characterizing connectivity as a partition of voxels into subsets that are well characterized by N_s representative hypotheses, or time courses, $\mathbf{m}_1, \dots, \mathbf{m}_{N_s}$ based on the similarity of their time courses to each hypothesis. We model the fMRI signal \mathbf{Y} at each voxel as generated by the mixture $p_{\mathbf{Y}}(\mathbf{y}) = \sum_{s=1}^{N_s} \lambda_s p_{\mathbf{Y}|S}(\mathbf{y}|s)$ over N_s conditional likelihoods $p_{\mathbf{Y}|S}(\mathbf{y}|s)$ [15]. λ_s is the prior probability that a voxel belongs to system $s \in \{1, \dots, N_s\}$. Following a commonly used approach in fMRI analysis, we model the class-conditional densities as normal distributions centered around the system mean time course, i.e., $p_{\mathbf{Y}|S}(\mathbf{y}|s) = \mathcal{N}(\mathbf{y}; \mathbf{m}_s, \Sigma_s)$. The high dimensionality of the fMRI data makes modeling a full covariance matrix impractical. Instead, most methods either limit the modeling to estimating variance elements, or model the time course dynamics as an auto-regressive (AR) process. At this stage, we take the simpler approach of modeling variance and note that the mixture model estimation can be straightforwardly extended to include an AR model. Unlike separate dimensionality reduction procedures, this approach follows closely the notions of functional similarity used by the detection methods in fMRI. In other words, we keep the notion of co-activation consistent with the standard analysis and instead redefine how the co-activation patterns are represented and extracted from images.

We employ the EM algorithm [6] to fit the mixture model to the fMRI signals from a set of voxels, leading to a familiar set of update rules:

$$\begin{aligned} \tilde{p}^n(s|\mathbf{y}_v) &= \frac{\lambda_s^n \mathcal{N}(\mathbf{y}_v; \mathbf{m}_s^n, \Sigma_s^n)}{\sum_{s'} \lambda_{s'}^n \mathcal{N}(\mathbf{y}_v; \mathbf{m}_{s'}^n, \Sigma_{s'}^n)}, & \lambda_s^{n+1} &= \frac{\sum_v \tilde{p}^n(s|\mathbf{y}_v)}{V} \\ \mathbf{m}_s^{n+1} &= \frac{\sum_v \tilde{p}^n(s|\mathbf{y}_v) \mathbf{y}_v}{\sum_v \tilde{p}^n(s|\mathbf{y}_v)}, & \Sigma_s^{n+1}(t, t) &= \frac{\sum_v \tilde{p}^n(s|\mathbf{y}_v) (\mathbf{y}_v(t) - \mathbf{m}_s^{n+1}(t))^2}{\sum_v \tilde{p}^n(s|\mathbf{y}_v)} \end{aligned}$$

where $\tilde{p}^n(s|\mathbf{y}_v)$ is the estimate of the posterior probability that voxel v belongs to system s , and $\{\lambda_s^n, \mathbf{m}_s^n, \Sigma_s^n\}_{s=1}^{N_s}$ are the model parameter estimates at step n of the algorithm. As mentioned above, we model the covariance matrix as a diagonal matrix. To ensure that we properly explore the non-convex space of the solutions, we perform multiple runs (10 in our experiments) of the algorithm using different random initializations and select the solution that corresponds to the maximum likelihood of the data. We initialize each run by randomly selecting N_s voxels and using their time courses as an initial guess for the cluster means.

When the algorithm converges, $\tilde{p}(\cdot|\mathbf{y}_v)$ represents probabilistic segmentation. The exponential form of class-conditional densities, combined with high dimensionality of the input space, leads to essentially binary posterior probabilities (in our experience, fewer than 1% of voxels is assigned a posterior probability that is more than 10^{-3} away from 0 or 1). Unlike anatomical segmentation, atlas-based approaches are not applicable to this problem, since the instantaneous properties of fMRI signals vary substantially across runs, and little is known about spatial organization of the functional activity we should expect to see as a result of segmentation. Consequently, we perform the segmentation in a fully unsupervised fashion.

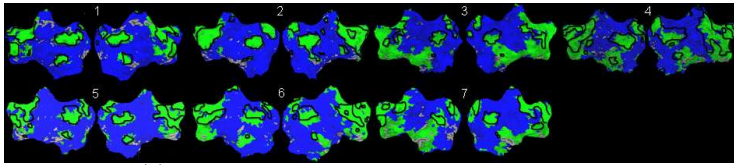
To summarize the results of the segmentation across subjects, we must make sure that the labels assigned to the same anatomical system agree across subjects. We employ an approximate algorithm that matches the label assignments in pairs of subjects with the goal of maximizing the number of voxels with the same label in both subjects until convergence. In practice, this algorithm quickly (1-2 passes over all subjects) finds the correct label permutation in each subject.

3 Experimental Results

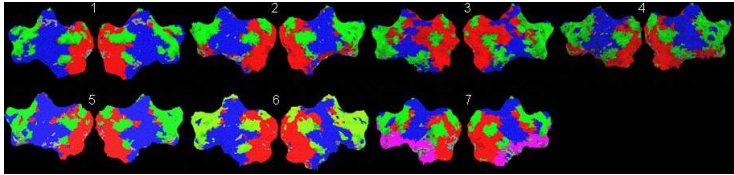
We demonstrate our approach on a study of functional connectivity that included 7 subjects. We used previously collected fMRI scans in a large set of visual experiments, from simple localizer tasks to viewing continuous stimuli (movies), as well as a rest scan. The total amount of fMRI data per subject was close to one hour. In the movie viewing experiments, the functional connectivity analysis revealed two systems: the stimulus-dependent system that contained sensory-motor cortexes and was strongly correlated with the seed region in the visual cortex and the ‘intrinsic’ system that showed little correlation with the visual seed, but exhibited high intra-system correlation [11].

The functional scans were pre-processed for motion artifacts, manually aligned into the Talairach coordinate system, detrended (removing linear trends in the baseline activation) and smoothed (8mm kernel). We experimented with different amount of smoothing and observed that it had very little effect on the resulting decompositions. We restricted the analysis to the voxels in the cortical segmentation mask in the corresponding anatomical scans and chose to visualize the resulting decompositions on the inflated surface of the cortex, as well as using a flattened representation of both hemispheres. Functional segmentations extracted for the same subject in different visual experiments varied little in the anatomical boundaries of the detected systems even though the details of the systems’ time course dynamics changed substantially across experiments. This is in line with the current theories of the functional organization of the brain that postulate anatomically stationary regions whose changing activation dynamics drives the cognitive processes. Using all the data in a single segmentation resulted in a more repeatable segmentation when compared across subjects, suggesting that increasing the amount of fMRI data stabilizes the estimation process. The experimental results reported in the remainder of this section are based on all available data for each subject.

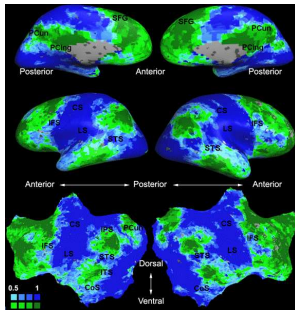
Fig. 1a shows the 2-system partition extracted in each subject independently of all others. It also displays the boundaries of the intrinsic system determined through the traditional seed selection, showing good agreement between the two partitions. In contrast to the difficulties associated with the subject-specific threshold selection in group analysis within the standard functional connectivity framework, the clustering-based decomposition produces highly repeatable maps that do not involve subject-specific adjustments. Fig. 1c shows a group-level label map that summarizes the maps from Fig. 1a, further illustrating the stability of the decomposition. We emphasize that no sophisticated group-wise



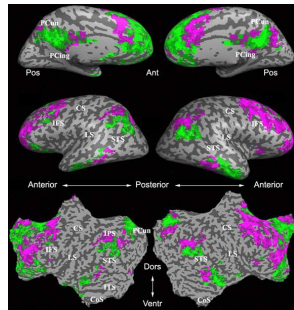
(a) 2-system segmentation, subject-specific maps



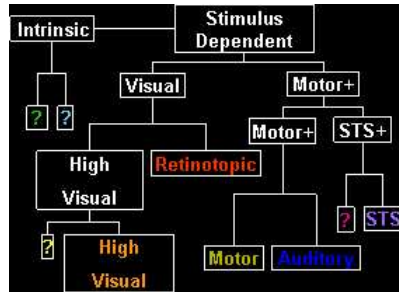
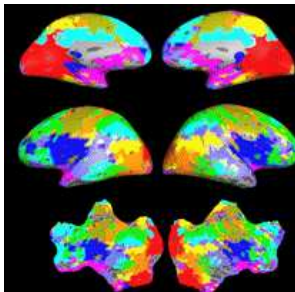
(b) 3-system segmentation, subject-specific maps



(c) group average of 2-system subject-specific maps in (a)



(d) group average of 2-system partition of the intrinsic system



(e) Hierarchical interpretation of 8-system segmentation for subject 1

Fig. 1. Functional segmentation examples. (a,b) Subject-specific segmentation results for two and three systems respectively (flattened view). Green: intrinsic system, blue: stimulus-driven cortex, red: visual cortex. Solid lines show the boundaries of the intrinsic system determined through seed selection. (c) Group average of the subject-specific 2-system maps. Color shading shows the proportion of subjects whose clustering agreed with the majority label. (d) Group average of the subject-specific segmentation of the intrinsic system into two sub-systems. Only voxels consistently labeled across subjects are shown. (e) Subject-specific segmentation into a large number of systems. Browsing of all preceding levels (not shown here) revealed the hierarchy displayed on the right. Colors show matching systems in the image (left) and labels in the hierarchy (right).

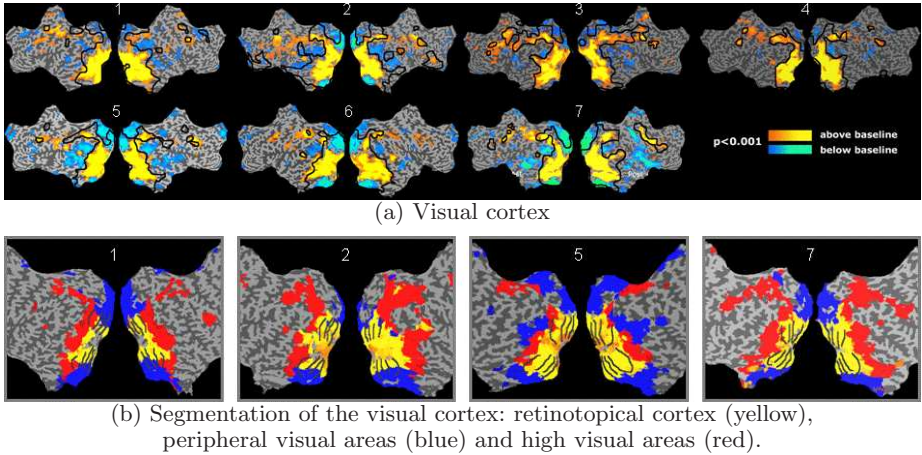


Fig. 2. Comparison with the regression-based detection. (a) Color shows the statistical parametric map; solid lines indicate the boundaries of the visual system obtained through clustering. (b) 3-system segmentation of the visual cortex for subjects 1,2,5,7. Only the posterior half of the flattened view is shown for each subject. The black lines indicate the boundaries of V1-V4 regions.

registration was performed; no information was shared across subjects during segmentation. Subsequent subdivision of the cortical gray matter into three systems produced the results in Fig. 1b. With the exception of one subject, the 3-system segmentation reveals visual cortex. In subject 7, the visual cortex separated in segmentation into 4 systems (shown in the figure). Fig. 1e shows an example of subject-specific segmentation into 8 systems and its hierarchical interpretation that was constructed by a neuroscientist through interactive browsing of the increasingly detailed segmentations (not shown here). While the final segmentation map in Fig. 1e would be difficult to interpret if it were considered as a stand-alone flat parcellation, a collection of segmentations into an increasingly large number of systems makes interpretation exceedingly easy. The nested nature of the segmentation results suggests future work in hierarchical representations that capture and exploit this property to improve detection and interpretation.

Since the original study aimed to characterize the intrinsic system, we also performed a subdivision of just that system in each subject. Interestingly, this subdivision produced a stable partition across subjects; the corresponding group-level map is shown in Fig. 1d. The overlap of the smaller sub-systems is weaker than that of the intrinsic system, but it clearly represents a coherent division of the intrinsic system. We are currently investigating neuroanatomical and functional characteristics of the two new sub-areas.

We also provide preliminary validation of the method by comparing the resulting parcellations with well known partitions in the visual cortex. Fig. 2a compares the boundaries of the visual system identified through clustering (red

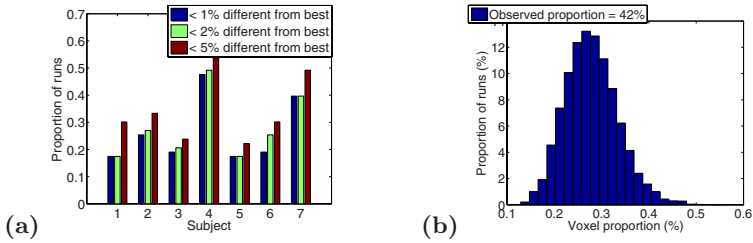


Fig. 3. Performance statistics. (a) Proportion of runs that resulted in the segmentation that was close to the best (max likelihood) segmentation. (b) Null hypothesis distribution for the number of voxels that showed perfect agreement across all subjects.

cluster in Fig. 1b) with the statistical parametric map (SPM) from a block-design visual localizer experiment. The regions of reduced activation in the SPM correspond to the well known suppression of activation in the intrinsic system (anterior of the visual cortex; outside the visual system) and the reduction in the signal often observed in the peripheral visual areas (posterior cortex; included in the visual system). We can see that the two methods agree on the location of the boundary between the visual cortex and the adjacent areas. We emphasize that the segmentation method had no access to the protocol regressor from the visual experiments. Further subdivision of the visual system revealed the central-peripheral partition that separates retinotopic areas (V1 through V4) from the peripheral visual areas and the retinotopic-high partition that separates the retinotopic cortex from the high visual areas. These two fundamental organizational principles have been extensively studied using a set of techniques, including fMRI [12]. Fig. 2b compares the results of segmentation with the retinotopic mapping available for four subjects in our study. The solid lines indicate the extent and the boundaries of the retinotopic areas obtained in a separate fMRI experiment [12]. We can see that the segmentation accurately estimated the boundaries between the retinotopic areas (yellow), the high visual areas (red) and the peripheral areas (blue) using fMRI data from a diverse set of experiments which were not specifically tailored for retinotopic mapping.

To test the robustness of the EM algorithm in this application, we ran the segmentation for 100 different random iterations and examined the resulting maps. We sorted the resulting segmentations by the value of the corresponding data likelihood. Using the best result as a reference, Fig. 3a shows the number of resulting segmentations that varied from the best segmentation by less than 1%, 2% and 5% respectively. We observe that a reasonable proportion of the runs (from 15% to 50%) produced segmentations that are very close to the best one. Similar to other hill-climbing optimization problems, our goal is not to ensure that all runs result in a good solution, but rather than sufficiently high proportion of random initializations leads to a good solution. To quantify the significance of the agreement across subjects, we ran permutation tests. In each iteration of the test, the voxel locations were permuted, the best relabeling across subjects was estimated and the proportion of voxels that achieved perfect agreement across

subjects was recorded. When comparing the observed proportion in the real data with the histograms created in the permutation test for 10^5 iterations (Fig. 3b), we note that the result is dramatically significant under this null hypothesis. Our future work includes developing more realistic null hypotheses that maintain spatial statistics of labels observed in the estimated segmentations.

4 Conclusions

We proposed and demonstrated a novel approach to characterizing global spatial patterns of co-activation in fMRI. The analysis produces hierarchical decompositions of the gray matter into a set of regions with increasingly consistent functional activity. By explicitly decoupling inter-subject variability in the spatial pattern of activation from the time course variability, our approach overcomes the need for subject-specific threshold selection often necessary in the standard methods for group analysis of fMRI data. In contrast to component-based analysis, the proposed method provides an intuitive model of cortical parcellation into systems and leads naturally to a hierarchical formulation that we plan to explore in the future. We provide initial validation of the method by comparing the detected systems with the previously known divisions of cortical areas. We also demonstrate an application of the method to detect novel functional partitions.

Acknowledgements. This research was supported in part by the NIH NIBIB NAMIC U54-EB005149, NCRN NAC P41-RR13218 and NCRN mBIRN U24-RR021382 grants, by the NSF CAREER grant 0642971 to P. Golland and by Morris and Barbara Levinson Professorial Chair and ISF Center of Excellence the Dominique Center, the Benozio Center for Neuro-Degeneration to R. Malach.

References

- [1] Baumgartner, R., et al.: Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. *J. Magnetic Resonance Imaging* 7(6), 1094–1101 (1997)
- [2] Beckmann, C.F., Smith, S.M.: Tensorial Extensions of Independent Component Analysis for Group FMRI Data Analysis. *NeuroImage* 25(1), 294–311 (2005)
- [3] Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. *Neural Computation* 7, 1129–1159 (1995)
- [4] Biswal, B., et al.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine* 34, 537–541 (1995)
- [5] Cordes, D., et al.: Hierarchical clustering to measure connectivity in fMRI resting-state data. *Magnetic Resonance Imaging* 20(4), 305–317 (2002)
- [6] Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *Series B* 39(1), 1–38 (1977)
- [7] Fadili, M.J., et al.: A multistep Unsupervised Fuzzy Clustering Analysis of fMRI time series. *Human Brain Mapping* 10(4), 160–178 (2000)
- [8] Filzmoser, P., Baumgartner, R., Moser, E.: A hierarchical clustering method for analyzing functional MR images. *Magnetic Resonance Imaging* 10, 817–826 (1999)

- [9] Friston, K.J., et al.: Functional connectivity: the principle component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* 13, 5–14 (1993)
- [10] Friston, K.J.: Functional and effective connectivity: a synthesis. *Human Brain Mapping* 2, 56–78 (1994)
- [11] Golland, Y., et al.: Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. *Cereb. Cortex* 17, 766–777 (2007)
- [12] Grill-Spector, K., et al.: A sequence of early object processing stages revealed by fMRI in human occipital lobe. *Human Brain Mapping* 6(4), 316–328 (1998)
- [13] Goutte, C., et al.: On clustering fMRI time series. *Neuroimage* 9(3), 298–310 (1999)
- [14] McKeown, M.J., et al.: Analysis of fMRI data by blind separation into spatial independent components. *Human Brain Mapping* 6, 160–188 (1998)
- [15] McLachlan, G.J., Basford, K.E.: *Mixture models. Inference and applications to clustering.* Dekker (1988)
- [16] Moser, E., Diemling, M., Baumgartner, R.: Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. *J. Magnetic Resonance Imaging* 7(6), 1102–1108 (1997)
- [17] Thirion, B., Faugeras, O.: Feature Detection in fMRI Data: The Information Bottleneck Approach. *Medical Image Analysis* 8, 403–419 (2004)
- [18] Thirion, B., Dodel, S., Poline, J.B.: Detection of signal synchronizations in resting-state fMRI datasets. *NeuroImage* 29, 321–327 (2006)
- [19] Van Leemput, K., et al.: Automated model-based tissue classification of MR images of the brain. *IEEE TMI* 18(10), 897–908 (1999)