Probabilistic Anatomo-Functional Parcellation of the Cortex: How Many Regions?

Alan Tucholka[1,2], Bertrand Thirion[2], Matthieu Perrot[1], Philippe Pinel[3], Jean-François Mangin[1], and Jean-Baptiste Poline[1]

[1] CEA Saclay, Neurospin/LNAO, Bât 145, 91191 Gif-sur-Yvette cedex, France
alan.tucholka@cea.fr
[2] INRIA Saclay-Île-de-France, Parietal, Saclay, France
[3] INSERM UNICOG, Neurospin, Paris, France

Abstract. Understanding brain structure and function entails the inclusion of anatomical and functional information in a common space, in order to study how these different informations relate to each other in a population of subjects. In this paper, we revisit the parcellation model and explicitly combine anatomical features, i.e. a segmentation of the cortex into gyri, with a functional information under the form of several cortical maps, which are used to further subdivide the gyri into functionally consistent regions. A probabilistic model is introduced, and the parcellation model is estimated using a Variational Bayes approach. The number of regions in the model is validated based on cross-validation. It is found that about 250 patches of cortex can be delineated both in the left and right hemisphere based on this procedure.

1 Introduction

One of the main goals in neuroimaging is to study the structure-function relations across subjects in order to better understand and characterize brain regions. Parcelling the cortical surface is a particularly intuitive approach, because it explicitly segments the cortex modules with an anatomical and functional definition that is consistent at the population level.

Anatomical parcellation approaches typically define a coordinate system on the cortical surface that represents the position of the main anatomical features (identified sulci, curvature), and maps the cortex to a sphere [1,2]. Then the cortical maps are further segmented into gyri, which are defined with respect to the main sulci of the brain [3], or directly in the sulcus-based coordinate system [4]. Although the selection and nomenclature of the gyri may vary across publications and softwares, the advantage of these approaches is that they provide a relatively standard division of the cortex into regions; still this description is quite coarse (30 to 60 regions), which limits its usefulness for defining functional modules.

In order to more finely delineate cortical regions, functional information, obtained using e.g. functional Magnetic Resonance Imaging (fMRI) data, provides further insights, and can easily be compared across subjects. A few approaches

have been proposed to identify reproducible functional activity areas among a group of subjects. A method based on replicator dynamics was introduced [5] to subdivide a pre-defined cortical region; coherent subdivisions were found across subjects. In [6] a method based on spectral clustering performed the delineation of homogeneous and connected regions which had similar position and functional activity across subjects. Finally, a clustering of fMRI time courses was performed in [7] to obtain subdivisions of the cortex. This provides a relatively coarse but prior free and consistent parcellation of the cortex based on fMRI data only. These approaches usually do not consider the anatomical information related to the data or simply reduce it to the three-dimensional coordinate systems. By contrast, it is important for interpretation purposes to relate functional information to anatomical structures [8,9]. Moreover, these approaches have not addressed the question of model selection.

In this study, we propose a novel approach to combine anatomical parcellation of the data into gyri and functional informations that further refines the parcellation. The final parcellation procedure is cast into a probabilistic framework and the parameters of the model are identified using a Variational Bayes approach similar to [10]. Finally, cross-validation procedures are used to optimize the number of components as in [11]. Starting from 47 initial anatomical parcels (gyri), 254 and 229 anatomo-functional parcels are found in the left and right hemisphere respectively in our dataset of 25 subjects.

2   Materials and Methods

2.1   Data Description and Pre-processing

Data was acquired from 25 subjects who performed a functional localizer protocol as described in [12]. The Brainvisa package was used to segment different anatomical compartments from the T1 image of the brain of each subject, providing white and gray matter mesh, and segmenting the sulci [13]. This sequence of treatments was applied systematically to all available brains and the quality of resulting segmentation was checked and in some cases corrected to solve inter-subject inconsistencies. Next, we used the method described in [2] to obtain a surface-based coordinate system that takes into account the main sulci, and the methods described in [4] to subdivide the cortical surface into gyri. The coordinate system maps the cortex onto a sphere, and will thus be denoted as (θ, ϕ).

For all subjects a standard preprocessing of fMRI data (distortion correction, correction of differences in slice timing, motion correction and anatomo-functional co-registration) was performed using the SPM5 software. Functional images were then projected onto the gray/white interface using the method described in [14]. Subsequently, on each functional dataset a GLM analysis was carried out to obtain task-related activity maps for different contrasts of experimental conditions. In this work we are using $n_f = 4$ contrasts: *i)* left versus right button presses, *ii)* sentence listening versus sentence reading, *iii)* computation versus reading *iv)* reading versus passive checkerboard viewing. A schematic flowchart of this processing sequence is shown in fig. 1 **A**.
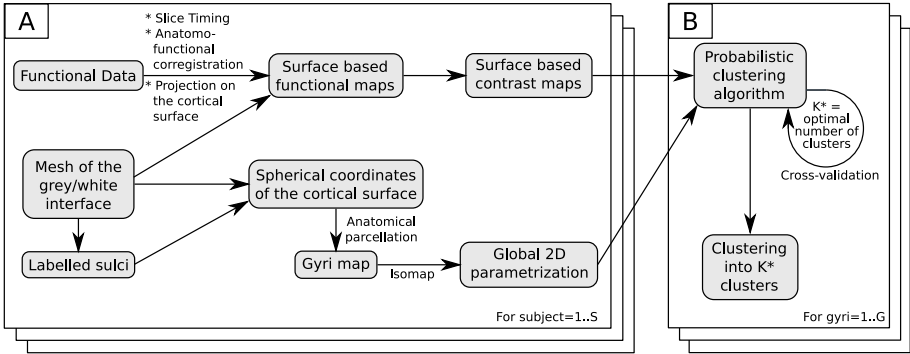
Fig. 1. Flowchart of the method. The part labeled **A** shows steps of pre-treatments up to GLM and global parametrization of the cortical surface. In part **B** sub-division into patches is obtained by probabilistic clustering algorithm.

2.2   Local Coordinates and Distances

In this section we derive a coordinate system adapted to the representation of a gyrus which closely approximates the spherical coordinate system that describes cortical topography. Let (θ, ϕ) be this system of spherical coordinates. The distance between two nodes $p_1(\theta_1, \phi_1)$ and $p_2(\theta_2, \phi_2)$ is given by:

$$d_S(p_1, p_2) = \arccos\left\{\cos(\theta_1 - \theta_2) + (\cos(\phi_1 - \phi_2) - 1)\sin(\theta_1)\sin(\theta_2)\right\} \qquad (1)$$

Since it is convenient to work neither with this non-Euclidean metric nor with three-dimensional coordinates, we create a new local coordinate system that respects the relation (1), by using a multi-dimensional scaling algorithm: Given a matrix $\mathcal{D}_S$ of squared distances between any pairs of points on a gyrus, we find the two-dimensional embedding $x = (x_1, x_2)$ that minimally distorts $\mathcal{D}_S$ in the least-square sense: $x = argmin_{\chi \in G^2}\|\bar{\mathcal{D}}_S - \chi^T\chi\|^2$, where $\bar{\mathcal{D}}_S$ is the matrix $\mathcal{D}_S$ after centering, and G^2 is the set of two-dimensional functions defined on the gyrus. The solution is simply provided by a singular value decomposition of $\mathcal{D}_S$, because the gyrus has a trivial topology. Importantly, this computation can be done across subjects, so that all the subjects are finally in the same local coordinate system, which is quasi isometric to the spherical coordinate system.

2.3   Probabilistic Parcellation of the Data

A probabilistic clustering algorithm is used to parcel the data into functionally homogeneous and spatially coherent regions (flowchart shown in fig. 1 **B**). This algorithm is based on a Bayesian formulation of the problem and the estimation rests on a Variational Bayes approximation similar to [10]. Let $\{x_i\}_{i=1..I}$ be a set of 2D coordinates that represent the position of cortical sites, and let $Y = \{y_i\}_{i=1..I}$ be n_f-dimensional vectors that represent the functional activity related to these sites. Let $K > 0$ be the number of components of the probabilistic model

and let $(w_{ik})_{i=1..I,k=1..K}$ denote the prior probability that the site i belongs to component $k \in [1..K]$. Let $\Theta = (\theta_k)_{k=1..K}$ be the set of parameters of the normal densities that represent the functional information related to the classes. The likelihood of the observation is thus

$$p(Y|\Theta) = \prod_{i=1}^{I} \left(\sum_{k=1}^{K} w_{ik} \mathcal{N}(y_i|\theta_k) \right), \tag{2}$$

which would be a standard Gaussian Mixture Model if the loadings w_{ik} were functions of k only. Here w_{ik} are a function of x_i, a constant scalar γ, and a set of two-dimensional coordinates $\mathcal{T} = (\tau_k)_{k=1..K}$ that describe the position of the clusters on the cortex, which are given a Gaussian prior:

$$w_{ik}(x_i, \mathcal{T}) = \frac{\exp(-\frac{\|x_i - \tau_k\|^2}{2\gamma^2})}{\sum_{l=1}^{K} \exp(-\frac{\|x_i - \tau_l\|^2}{2\gamma^2})} \tag{3}$$

$$p(\mathcal{T}) = \prod_{k=1}^{K} \mathcal{N}(\tau_k; u_k^0, \Lambda_0) \tag{4}$$

The posterior of the parameters is thus

$$p(\Theta, \mathcal{T}|, Y) = \prod_{i=1}^{I} \left(\sum_{k=1}^{K} w_{ik}(x_i, \mathcal{T}) \mathcal{N}(y_i|\theta_k) \right) p(\Theta) p(\mathcal{T}) \tag{5}$$

As in [10], we use an empirical Bayes approach where the spatial and functional parameters are treated independently and a point estimate of Θ is used instead of its full density. We now proceed with the estimation of the posterior $p(\mathcal{T}|Y, \Theta)$, which is the complex part of the problem; to simplify notations, the dependence on Θ will be omitted. The complex dependence of the posterior on $\mathcal{T}$ prevents straightforward estimation of the parameters, hence we use a variational Bayes (VB) approximation. The VB approach relies on the observation that

$$\log p(Y) = F + KL = \int q(\mathcal{T}|Y) \log \frac{p(Y, \mathcal{T})}{q(\mathcal{T}|Y)} d\mathcal{T} + \int q(\mathcal{T}|Y) \log \frac{q(\mathcal{T}|Y)}{p(\mathcal{T}|Y)} d\mathcal{T} \tag{6}$$

where F stands for the variational free energy and KL is the Kullback-Leibler divergence between the approximate and the true posterior and $q(\mathcal{T}|Y)$ is called the variational density. The minimization of KL is thus equivalent to the maximization of F. As a classical approximation, the positions τ_k are assumed to be jointly independent and normally distributed.

$$q(\mathcal{T}|Y) = \prod_{k=1}^{K} \mathcal{N}(\tau_k; u_k, \Lambda_k) \tag{7}$$

The maximization of the free energy of the probabilistic model with respect to the variables τ_k yields

$$q(\tau_k|Y) = \exp\left(\langle \log(p(\mathcal{T}, Y|\gamma)) \rangle_{q(\tau_l|Y), l \neq k} \right) \tag{8}$$

Due to its complex dependence on $\mathcal{T}$, the joint distribution $\log(p(\mathcal{T}, Y)) = \log(p(\mathcal{T})) + \log(p(Y|\mathcal{T}))$ is approximated using a second-order Taylor expansion of $-\log(p(Y|\mathcal{T}))$ around its mode:

$$-\log(p(Y|\mathcal{T})) = f(\mathcal{T}) \approx f(u) + (\mathcal{T} - u)^T h + \frac{1}{2}(\mathcal{T} - u)^T J(\mathcal{T} - u) \qquad (9)$$

where f represents the dependence of the model on $\mathcal{T}$ implied by Eqs. (2,3), h and J represent the gradient and Hessian of f, and are derived from Eq. (2). Given this approximation, it is possible to perform the update in Eq. (8). After some algebra, the update rules of the parameters in Eq. (7) are (at iteration j):

$$\Lambda_k^{(j)} \leftarrow (\Lambda_0^{-1} + J_{k,k}^{-1})^{-1}, u_k^{(j)} \leftarrow \Lambda_k^{(j)}\left(\Lambda_0^{-1}u_k^0 + J_{k,k}^{-1}u_k^{(j-1)} - h_k\right) \qquad (10)$$

where $J_{k,k}$ is the sub-matrix of J related to τ_k and h_k is the kth component of the gradient h. Using a standard empirical Bayes approach, the parameters of the classes Θ are optimized using the current estimate of $\mathcal{T}$ (this is simply the M-step of an EM algorithm), $\Theta = argmax_\theta \, p(Y|\theta, \mathcal{T})$, see Eq. (2). The estimations of $\mathcal{T}$ and Θ are alternated. Finally, the prior mean and variance on the position $((u_k^0)_{k \in [1..K]}, \Lambda_0)$ are obtained through an initial clustering of the data that does not take into account the functional information.

2.4   Optimizing the Model

Two main parameters of the model need to be optimized, namely K, which represents the number of components needed in the mixture model, and γ, which controls the strength of the spatial information (see Eq. (3)). In order to have an interpretable procedure, and to avoid any confounding effect of the choice of the prior parameters, we proceed by cross-validation on the subjects data: For each subject $s \in \{1, .., S\}$, we estimate the parameters $(\mathcal{T}, \Theta)$ by pulling the data from all the subjects of the cohort but s, and then assess the goodness of fit of the resulting model on the data of s using Eq. (2). We do so for various choices of K and γ and select the best one according the maximal average cross-validated likelihood criterion. Let $K^\star$ be the best value for K.

  We noticed that the dependence on γ -at least within a reasonable range of values- is not dramatic i.e. it is smooth and has a mild impact on $K^\star$, so that the results on the optimality of $K^\star$ are not artefactual in that sense.

3   Results

To test our algorithm, we simulated different sets of five constant patches added with noise, within a square region, with a learning and a test set. γ and $K^\star$ were optimized using our method and the algorithm used in [11]. This was repeated 100 times. Row **A** in fig. 2 shows $i)$ a typical input of the simulation experiment, $ii)$ a typical and approximately correct partition into five regions, $iii)$ the values of $K^\star$ across 100 simulations, the corresponding values of $K^\star$ with the algorithm
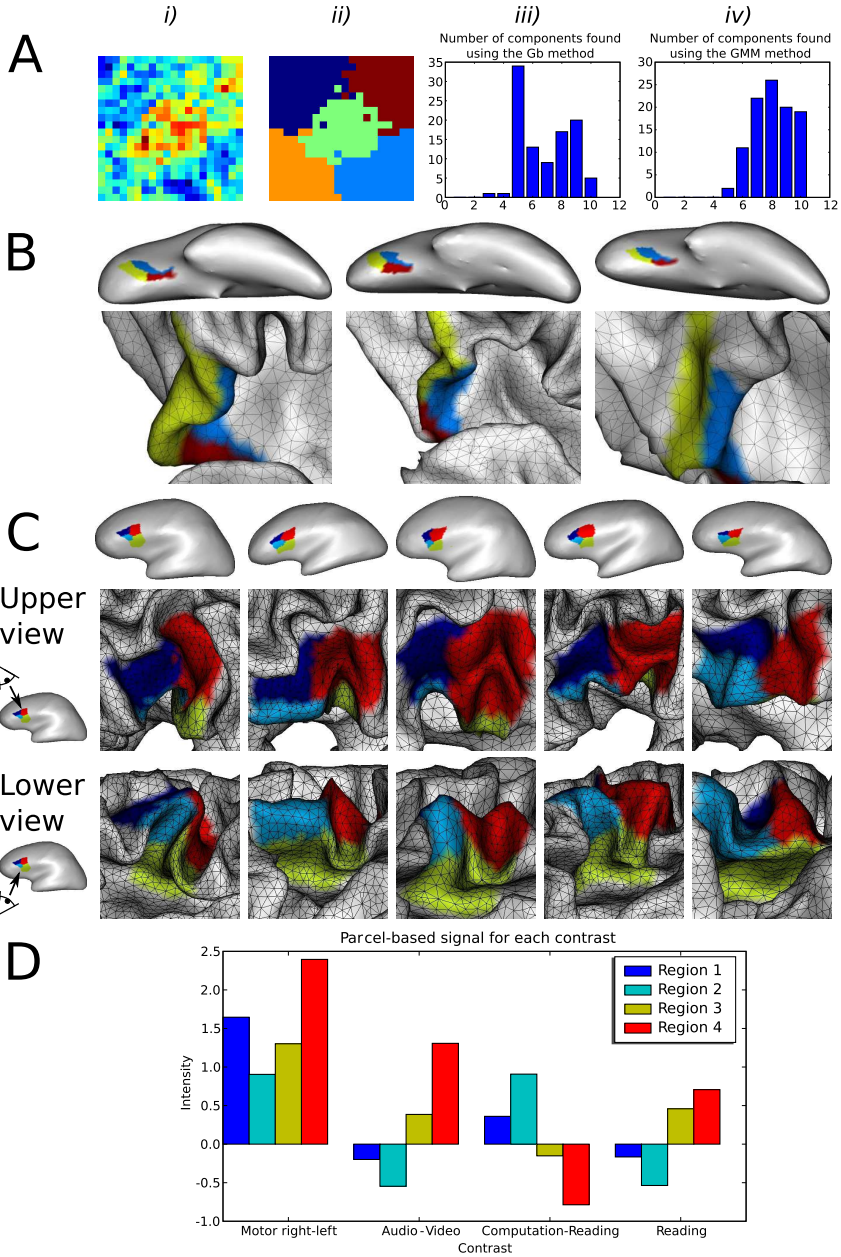
Fig. 2. A Validation of our approach on simulated data: *i)* example of a simulated dataset *ii)* resulting parcellation *iii)* histogram of $K^\star$ for 100 random draws *iv)* idem with the method of [11]. **B** Sub-division on the frontal-ventral gyrus. Two larges patches have characteristic limit on the crest of the gyrus. **C** Pattern of patches on the opercular section of the Broca's area. In these five subjects, the relative position of the regions is remarkably stable. **D** Functional information related to the different parcels of **C**.

[11]. Our algorithm finds much more frequently the correct solution than the other, which typically overestimates K.

Using a maximum a posteriori partition of the data, the individual meshes can be labelled according to the spatial model. The resulting parcels are clearly interpretable through both the gyrus that contains them and the functional activity that characterizes the parcel.

The parcellation is also geometrically consistent across subjects: On the frontal-ventral gyrus (in fig. 2 **B**) we identify the border between two large patches clearly on the crest of the gyrus and the small third patch is always on the same end of the gyrus. On the opercular section of the Broca's area (fig. 2 **C**), four distinct patches are positioned in the same patchwork pattern, and are characterized functionally in fig 2. For instance, subregion 4 has a more asymmetric activity in the motor task, and is more involved in the auditory task, while subregion 2 is more involved in the computation task, and minimally involved in reading and listening sentences; subregions 1 and 3 have intermediate behaviour between these two cases. Altogether, the number of regions found in the model is 254 in the left cortex and 229 in the right cortex.

4   Discussion

We have shown that by combining anatomical and functional information on the cortical surface it is possible to identify the structure of the brain anatomo-functional partitioning on three hierarchical levels: Lobe $\Rightarrow$ Gyrus $\Rightarrow$ Functional patch/region. With respect to previous anatomo-functional parcellation [5,6,7,11], this clarifies the definition of the entities found by this approach. The last level division is provided by functional data, which means that the parcellation ultimately describes some local gradient of functional information within a given anatomical gyrus. This approach is intended to provide useful insights in the organisation of brain functional topography. A related observation is that in most of the cases the limits between patches occur at the crest line of the gyri (typically on the top of the gyri). Deciding whether this is an intrinsic feature of functional anatomy or an effect of functional data projection deserves further investigation.

Ultimately, the optimal number of subdivisions per gyri is related to the across-subjects consistency of the functional features used to perform the subdivision, hence functional variability at the population-level. A compromise is thus found between a large number of small parcels that do not generalize well across subjects, and a coarse parcellation into a few subregions that is not sufficient to describe the available functional information. Based on 25 subjects, our algorithm provided a subdivision into 254 regions on the left hemisphere, and 229 on the right hemisphere. It is comparable, and somewhat finer than what was found using 3D parcellation [11] (about 500 regions for the entire brain). However, the present approach is based on a more appropriate probabilistic model, and the former did not use explicit anatomical information. Moreover, the number of regions clearly depends on the available functional information, and will probably be further refined in the future. These results also depends strongly on the correction of the EPI distortion and precise anatomo-functional co-registration. Misfits shift functional

data in individual subjects and might introduce significant bias. Interestingly, the left/right asymmetry in this number indicates that more information is available to segment brain regions in the left hemisphere, which coincides with the fact that reading is known to yield a much wider activation network on the left cortex in average. Future improvements of the method include a hierarchical model for the functional information related to the components in Eq. (2), which will better account for the inter-individual variability. Moreover, some simplifying hypotheses in our estimation procedure may be bypassed in the future.

Accumulating knowledge on the spatial localization of functional activity in various experimental contexts is an important challenge for neuroimaging. In this work we have merged different functional and anatomical informations into a coherent probabilistic framework and inferred an optimized parcellation of the cortex, which could thus be a basis for future anatomo-functional atlases.

References

1. Fischl, B., Sereno, M.I., Dale, A.M.: Cortical surface-based analysis. ii: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207 (1999)
2. Clouchoux, C., Coulon, O., Rivière, D., Cachia, A., et al.: Anatomically constrained surface parameterization for cortical localization. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 344–351. Springer, Heidelberg (2005)
3. Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., et al.: Automatically parcellating the human cerebral cortex. Cereb Cortex 14, 11–22 (2004)
4. Clouchoux, C., Coulon, O., Anton, J.L., Mangin, J.F., et al.: A new cortical surface parcellation model and its automatic implementation. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 193–200. Springer, Heidelberg (2006)
5. Neumann, J., von Cramon, D.Y., Forstmann, B.U., et al.: The parcellation of cortical areas using replicator dynamics in fMRI. Neuroimage 32, 208–219 (2006)
6. Thirion, B., et al.: Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27, 678–693 (2006)
7. Golland, P., Golland, Y., Malach, R.: Detection of spatial activation patterns as unsupervised segmentation of fMRI data. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 110–118. Springer, Heidelberg (2007)
8. Brett, M., Johnsrude, I.S., Owen, A.M.: The problem of functional localization in the human brain. Nat. Rev. Neurosci. 3, 243–249 (2002)
9. Devlin, J.T., Poldrack, R.A.: In praise of tedious anatomy. Neuroimage 37, 1033–41; discussion 1050–8 (2007)
10. Woolrich, M.W., Behrens, T.E.: Variational bayes inference of spatial mixture models for segmentation. Trans. Med. Imaging 25, 1380–1391 (2006)
11. Thyreau, B., Thirion, B., Flandin, G., et al.: Anatomo-functional description of the brain: A probabilistic approach. In: ICASSP, pp. V–V (2006)
12. Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Bihan, D.L., Poline, J.B., Dehaene, S.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci. 8, 91 (2007)
13. Perrot, M., Rivière, D., Mangin, J.F.: Identifying cortical sulci from localizations, shape and local organization. In: 5th ISBI, Paris, France, pp. 420–423 (2008)
14. Operto, G., Bulot, R., et al.: Projection of fMRI data onto the cortical surface using anatomically-informed convolution kernels. Neuroimage 39, 127–135 (2008)