

## The effect of preprocessing choices on classification accuracy using hyperalignment

### 1. Introduction

During the last years, extracting inter - individual correspondence has been one of the most debated topics in current fMRI research. It is not difficult to understand in fact that having a direct match between different participants brains give the possibility of addressing new scientific questions by analyzing between - participants differences and similarities.

This problem, has been firstly addressed by using anatomical alignment algorithms. With these strategies each individual brain is warped to a common reference template (Talarach, MNI) in order to establish spatial correspondence between brains. Therefore, after brain normalization, a point in the common space identified by its x, y, z coordinates is assumed to refer to a similar region in any brain normalized according to the same procedure.

However, brain normalization has several problems. First of all, many algorithms fails on aligning specific part of the brains making really hard to have perfect anatomical matching. Second, it introduce a challenging fundamental problem of neuroscience that goes beyond pure transformation issues since it includes the question about the consistency of *structural-functional relationships*. Since now infact, many neuroimaging studies have demonstrated that on a macro-anatomical scale a functional correspondence between different participants exists for various brain regions. However, this macro-anatomical coherence seems not to be respected on a smaller scale as often a multi-subjects study using anatomical alignment fails on finding those micro-distinctions that help people to properly represent the outside environment and discriminate across millions of complex visual stimuli.

In this context, functional hyperalignment (Haxby, 2011) turned out to be an effective multivariate method to functionally align different participants' brain using brain responses measured with fMRI instead of spatial anatomical landmarks. Hyperalignment thus completely abstracts from spatial information and uses a geometrical transformation, procrustean transform, over useful functional data to build an high dimensional common space that minimize the overall Euclidean distance between these sets of brain responses. Once the data are projected to this commonspace, correspondence between brains is established and the validity of such model can be tested through a between subject classification analysis (BSC). In a BSC the response vector of one subject are classified based on predictions computed over others subjects responses. This procedure allow to built a general valid model that can be valid across subjects and thus that can generalize on response tuning function that are common across brains.

However, because of his novelty, several aspects of this method are still under-investigated. Here, as a first step in our investigation, we propose to study the sensibility of such methods under three different standard preprocessing choices: motion correction, detrending and smoothing. Furthermore, we compared BSC performance for brain response vector that had been transformed in this commonspace to BSC performance of data that were anatomically

aligned and to within-subject classification (WSC), in which response vectors for subject were classified using a model trained on response vectors that comes from the same subject.

## **2. Methods**

### **2.1 Data**

In order to have a solid baseline to use as reference, we decided to use a well-known and well-studied dataset available online (PyMVPA, Hanke et al.) that has already been used for hyperalignment(Haxby, 2011) purposes and was acquired by Haxby, same author of such method. The dataset is a block-design fMRI dataset from a study on face and object representation in human ventral temporal cortex. It consists of 5 subjects with 12 runs per subject. In each run, the subjects passively viewed grey-scale images of eight object categories, grouped in 24s blocks separated by rest periods. Each image was shown for 500ms and was followed by a 1500ms inter-stimulus interval. Full-brain fMRI data were recorded with a volume repetition time of 2.5s, thus, a stimulus block was covered by roughly 9 volumes.

### **2.2 Preprocessing procedures**

The analysis has been done using FSL 5.0 with the integration of in house built python scripts using nipy.org (Taylor) utilities together with numpy, scipy, pandas, pymvpa and matplotlib packages.

We first motion corrected all volumes using as reference the volume that comes in the middle of our time-series. Then, the brain was extracted from the skull and trends were removed independently for each run using a first level linear polynomial. After motion correction, brain extraction and detrending were done, the next step was to create an anatomical mask in order to focalized the analyses only on those areas that carry the 'important' information relative to the task. We highlighted the ventro-temporal cortex because of is well-known (haxby, peelen, hasson, etc.) involvement in the object representation. To do that, our analyses pass through different steps:

- Compute the transformation matrices from the anatomical data of each subject to the functional data (from anatomical to functional space)
- Compute the transformation matrices from this space to a MNI 2mm template's space
- Creating the mask on the reference template
- Bring our masks back to the functional native space

Furthermore, ones we had the transformation matrices from the native space to the normalized MNI space, was straightforward to bring our functional data to this space and run a between subjects MVPA analysis based on anatomical alignment.

### **2.3 Voxel selection**

The voxel selection is the step by which a subset of features is selected to increase the signal-to-noise ratio. Indeed, although we already focused our analysis on the ventro-temporal cortex, many voxels in this region contains irrelevant or redundant information that do not provide more or useful information relative to the task. We thus need to select some voxels within this large region that contain around 10 thousand voxels. There are numerous ways to select this subset of features and all of them required specific theoretical implication. Here, we tried two different ways of selecting feature; one widely used, ANOVA, and one innovative way introduced by Haxby and colleges [Haxby, 2011] that we will refer to as 'correlation based feature selection'.

### **2.3.1 ANOVA**

The analysis of variances (ANOVA) is a statistical test that measure the variation between group means and their associated procedures such as variations in conditions. In simple terms, it calculates the statistical significance of whether or not the means of more groups are equal, thus non-significant, or are not statistically equal and thus significant. The ANOVA is a plane univariate analysis that has thus been performed independently for each voxel and that assign to each voxel a score (F-score) that will refers to the degree of which it is or it is not statistically relevant. After ANOVA is performed, voxels can thus be ranked and a fix number of voxels can be selected. Here, we run one-way ANOVA independently per subjects and a fix (100 and 200) number of top-ranked voxels were selected.

### **2.3.2 Correlation based feature selection**

When hyperalignment was firstly presented, a specific voxel selection was designed. This strategy should have had the property of select those voxels that share the highest synchrony between brains. The idea behind this procedure comes from the theoretical assumption that if the same stimulus is presented to different people, a same or at least similar representation is coded in the brain. However, the pooriness of anatomical alignment on BSC classification problems proved that spatial correspondence between brains alone does not capture response tuning vectors that are common across people. Therefore, if the alignment cannot be anatomical there has to be some other locations where a same, or again similar, encoding process occurs. For this purpose the following feature selection method, correlation-based feature selection, was used.

The correlation-based feature selection (CB) starts with a correlation of each voxel of each subject with all voxels of all others subjects. Each line of the subject's matrix (*voxels x time-series*) is correlated individually with all lines of the other matrix but only the maximum value of this correlation is considered. Thus, when we correlate voxel  $n$  of subject A individually with all other voxels of subject B, and we select only the higher value throughout these correlations values, we are somehow assuming that the same process that is happening in voxel  $n$  is happening also in the other voxel to which voxel  $n$  correlate maximally. Once this step is done for all voxels between all subjects, we end up with a matrix for each subject  $i$  that is no more *voxels x timeseries* but instead is *voxels x number of subjects - 1*, where the rows are all the

original voxels of subject  $i$  while the column (one for each subject) represents only the value of the correlation of that voxel to the voxels to which it is maximum correlated, we will refer to this as “voxel correlation score”. The sum of all columns for each line, this means the sum of all “voxel correlation score” is considered as the “voxel total correlation score”. Given this score, it is again possible to select a fixed number of features. We choose to select the best 100 voxels independently for each hemisphere, this means those voxels that has the higher voxel final score.

## 2.4 Classification

On our data we run a multi-voxel pattern of analysis (MVPA). Respect to the normal univariate analysis that analyzed each feature independently, MVPA analyzed patterns of neural activity as vectors in a high-dimensional space where features are local dimensions of such pattern. However, once the patterns of brain responses to specific categories or conditions are defined throughout MVPA analysis, some sort of dissimilarity measurement across these patterns is required. Usually, a model of brain responses is created using the so called classifiers. The classifiers are algorithms that given a dataset with multiple classes learn it's characteristics to identify on new data if they belongs to one or to others classes. Classifiers can be of several types and their accuracy express their sensibility to discriminate across conditions. Here, we use a linear support vector machine(LSVM) to validate our procedures.

To avoid overfitting [Hawkings] the same data that are used to compute the commonspace cannot be used for others purposes and for the same reason, the same data that are used to train the classifier cannot be used for testing it. Thus, our analysis was done using a leave-one-run-out strategy where one run of each subject is left out and the commonspace is computed over data of all others runs. Once the commonspace is created, the transformation parameters of each subject from their native space to the commonspace are applied to the left-run-out and classification on this new high-dimensional commonspace is performed. For classification we use a leave-one-subject out strategy where the classifier is trained on all subjects with the exception of the left subject out that is used for testing. This analysis was run for all possible partitions with a cross-validation methods.

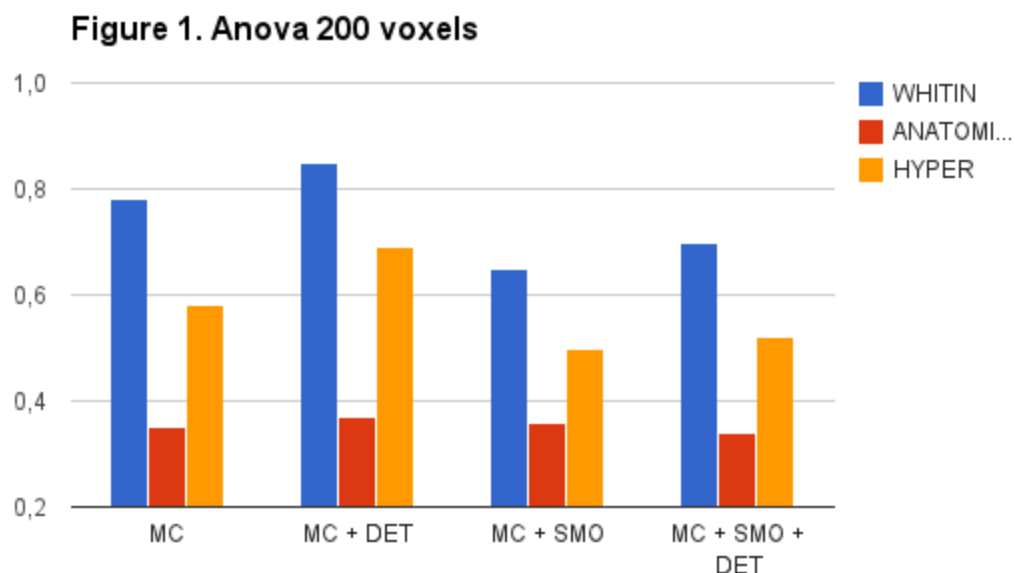
## 3. Results

Here, we show results of classification used the top 200 voxels that were selected using ANOVA feature selection (Figure 1) or correlation based feature selection (Figure 2). Classification has been done within subject (WSC), between subject with anatomical alignment (BSC anatomical) and between subject using hyperalignment (BSC-hyper). Furthermore, all classification has been run on files that come from different preprocessing procedures; motion correction was considered as a baseline and detrending and smoothing were applied both individually on the motion correction files or together as in the common used pipeline.

Overall, ANOVA seems to be more sensitive on this dataset than correlation feature selection. However, while ANOVA strongly increase classification accuracy for all files on WSC and BSC-hyper, it reduce sensibility for anatomical alignment. In fact, regardless of the

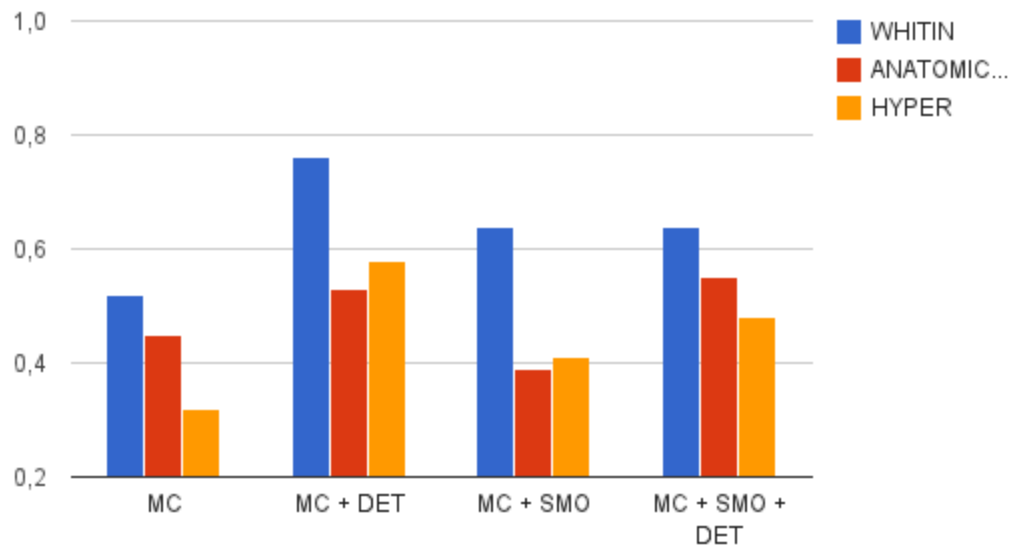
preprocessing analysis applied, it perform always around 30 % of accuracy when voxels were selected using ANOVA while when CB feature selection were used the performance grows until 50 % of accuracy. We were not surprised of these results. Indeed, we expected the correlation feature selection to be more efficient in conditions where data of a continuous stimulus are available rather than a block design where volunteers passively watched prototypical and static images. Infact, it has been demonstrated that a continuous stimulus offers the possibility of capture local brain responses from a wide range of visual stimuli that show an high synchrony across subjects in large regions of the cortex [Hasson, 2004].

Regarding our analysis on preprocess, we expected that adding preprocessing strategies would increase the sensibility of our classifier. We expected that whether all preprocessing steps (motion correction, detrending and smoothing) were used, the classifier would have shown the better performance respect to when only some of those steps were considered. Nevertheless, regardless of feature selection, smoothing seems to strongly reduce the power of the classifier. We assume that while smoothing can be really helpful to increase signal in case where weak brain responses are recorded, it can consistently increase noise, and thus decrease signal, on situation where a sparse and fine-grained signal is required to properly distinguish between conditions. On the other side and as expected, removing trends consistently increase classification accuracy in all conditions.



In general, hyperalignment outperform the plane anatomical alignment. Hyperalignment could reach the accuracy of  $0.74 \pm 0.026$  % (Figure 1) when only motion correction and detrending were applied and remain consistent across conditions. On the other side, anatomical alignment perform at most 56% and was also subjected to a strong variance between conditions that was a proof of his instability. However, even if the sensibility of hyperalignment was incredible high (74% with cache at 12.5%) on distinguish between 8 categories, it never reach performance of WSC ( $85\% \pm 0.080$ ).

**Figure 2. Correlation feature selection, 200 vx**



#### **4. Conclusion**

To summarize, we found that smoothing decrease the quality of our dataset while detrending highly improve it. Furthermore, ANOVA feature selection was more stable respect to correlation-based feature that show his strongness in condition where data of continuous stimulus are available.

Finally, hyperalignment seems to be a strong, stable and promising method that outperform the plane anatomical normalization. It can be used on all studies that want to built a valid model of specific cognitive abilities that can generalize across population. Furthermore, hyperalignment allow to direct match between brains making possible to check for similarities/ dissimilarities between them.

#### **References**

Haxby, James V., et al. "A common, high-dimensional model of the representational space in human ventral temporal cortex." *Neuron* 72.2 (2011): 404-416.

Hanke, Michael, et al. "PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data." *Neuroinformatics* 7.1 (2009): 37-53.

Hawkins, Douglas M. "The problem of overfitting." *Journal of chemical information and computer sciences* 44.1 (2004): 1-12.

Hasson, Uri, et al. "Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding." *Neuron* 57.3 (2008): 452-462.

Taylor, J., and F. Perez. "nipy: Neuroimaging in Python."

Talairach, Jean, and Pierre Tournoux. "Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging." (1988).