REPRODUCIBILITY OF VISUAL BRAIN MAPPING USING fMRI

Pedro B. Cruz

Instituto Superior Técnico

ABSTRACT

Functional magnetic resonance imaging (fMRI) techniques are extensively used today to assess brain activation upon the application of specific stimuli or tasks. In particular, retinotopic mapping of the human visual cortex can be used to evaluate cortical re-organization in visual field loss disorders or to assist the pre-surgical planning for tumour resection. With such clinical responsibilities, retinotopic mapping demands a robust protocol with minimal variability. The objective of this work was to develop such a protocol and to determine its reproducibility. We implemented a quadrant stimulation paradigm and tested a group of 5 subjects on two separate sessions, comprising four runs each. We then analyzed the results using two different dataset sizes (including 2 or 4 runs each), as well as a number of combinations of pre- and post-processing options. We finally assessed the within- and between-session reproducibility of the protocol using both dataset sizes, by calculating a number of variability measures of different activation parameters. In general, we found that reproducibility was higher for within-session relative to between-session comparisons. Moreover, increasing the dataset size considerably increased the reproducibility of the results. However, no significant systematic differences in reproducibility were found by using different processing parameter combinations.

With the optimal processing choices, the median distance between activation peaks was ~3 mm for within-session comparisons of datasets with 2 runs and between-session comparisons of datasets with 4 runs, rising to ~5 mm for between-session comparisons of datasets with 2 runs. In terms of the extent and intensity of the activation, the median coefficients of variation varied between 11 and 33% for the activation volume and between 2 and 4 % for the mean signal change, for between-session comparisons of datasets with 4 and 2 runs, respectively. As an overall measure of reproducibility, the median percentage of overlapping voxels varied between ~44 and 57 % for between-session comparisons of datasets with 4 and 2 runs, respectively.

In conclusion, we have developed and characterized the reproducibility of an fMRI protocol to detect and localize the retinotopic activation of visual cortex associated with quadrant stimulation of the visual field.

Keywords **-** fMRI, Retinotopy, Visual Mapping, Brain Activity, Reproducibility

1. INTRODUCTION

Functional brain assessment with fMRI has several potential applications, not only in basic neuroscience and research, but also in the clinical setting. The human visual cortex is one of the most studied brain regions. As it is a sensory system, the possibility of systematically manipulating the input in order to study the outputs makes it a very "trackable" brain function [1].

Retinotopic organization of the visual cortex

The human visual cortex is divided into several functional areas that have different local neuronal properties [2]. It is widely accepted that there is a hierarchy in these functional areas. Some of these areas possess a very interesting property known as retinotopy. Retinotopy consists of having a continuity relation between retinal adjacent neurons and the respective cortical ones. This means that neighbour neurons in the retina will keep this relation upon the respective brain cortex areas. This property is known as homeomorphism, which establishes that both domains are locally bijective and continuous. In other words, the visual field map (cortex) preserves the spatial layout off the visual scene (retina).

Although being a homeomorphic relation, the spatial layout preservation involves a symmetry transformation in both directions, left is right and right is left, and up is down and down is up. A magnification factor is also present in the retinotopic map. Retinal neurons near the fovea will have a bigger corresponding area in the visual cortex than on more eccentric retinal neurons.

Mapping the Human Virtual Cortex

The concept of retinotopic mapping only gained form when visual field sign mapping (VFS) was developed [3]. This mapping is used to determine the borders between visual cortical areas, V1, V2, V3, V3a and hV4 (visual cortical

areas classification and segmentation varies from author to author, although some, the lowest in the hierarchy such as V1 and V2, are consensually accepted by the scientific community). The stimulation paradigm consisted of a rotating wedge, composed of a contrast-reversing checkerboard pattern that rotates in a specific direction around a fixation target for several cycles. The borders between brain areas could therefore be assessed by reversals in the sign of the phase of the fMRI response to the stimulus. Before this paradigm another one had already been created and consisted of an expanding and contracting ring (centred on the fixation point) which permitted retinotopic mapping in a phase-locked way, in opposition to the rotating wedge (Fig.1.1), that allowed mapping from the foveal visual area to the retinal periphery [4].

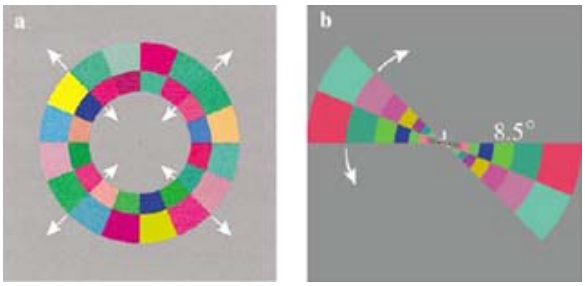

Fig. 1.1 – Expanding eccentricity ring (a) and rotating wedge (b) mapping paradigms [2]

Clinical Applications

Functional brain assessment with fMRI has several potential applications in the clinical setting. In cases of brain abnormalities, fMRI can play an active role in presurgical assessment. Presurgical fMRI has essentially three goals [5]: first, to assess the risk of postoperative undesired loss of brain function; second, to select patients for intraoperative mapping; third, to act as a neuronavigation intermediate in the surgical procedure itself. This way fMRI will be of central importance in integrating the whole clinical process of surgery.

Previous studies suggest, as a golden rule, that the minimum distance between tumour margin and essential structure for effective nearly risk-free surgical resection is of 10 to 15 mm [5]. More recently, authors have shown that this minimum distance can be of 10 mm, using fMRI, in risk-free surgical resection. However, further studies are needed to corroborate this hypothesis. fMRI can have an important role in identifying anatomical landmarks associated with important functional areas, in particular on brain distorted pathological cases. Another potentially important role is assessing brain cortical reorganization (plasticity), which has already been shown in some studies [5].

The second goal of presurgical fMRI is to select patients for intraoperative cortical stimulation (ICS) which allows the surgeon to accurately assess the tumour's resection margin.

Finally, the third goal is to aid surgeons in neuronavigation by using functional map pre-surgical data and perform a registration with the intraoperative structural data. This way any brain shifts arising from tumour resection and surgical brain manipulation in general can be overcome with this technique.

Variability in fMRI measurements

There are a number of factors that contribute, with more or less weight, to variability in fMRI measurements.

- The device itself introduces important noise components on the data. Constant field inhomogeneities or gradient differences, even minimal, contribute systematically to variability in fMRI.
- Patient motion is also a significative component of variability. It is impossible to keep any subject completely motionless and thus the slightest motion will produce a "kind of" blurred effect as data, for specific locations, will be acquired from neighbour areas.
- Stimulation noise – particularly on a visual stimulus with a fixation point, it is naturally impossible for any subject to keep focused on a point and thus whenever deviations from this point occur, undesired cortical visual areas will be stimulated, which introduces noise into the dataset
- Behavioural – ideally we would like to introduce only our stimulus of interest and cancel all the others. One knows that such is impossible. Each individual during each exam will have thoughts or be stimulated externally in general that will undoubtedly yield brain activation in certain cortical regions. Such activation can therefore amplify noise and variability in fMRI measurements.

Reproducibility is a direct measure of variability. The more variable a measure is, the less reproducible it will be. Previous works have assessed reproducibility both qualitatively, by analyzing the consistency of suprathresholded activated regions, and quantitatively. A number of reproducibility measures has been used: number of activated voxels; overlap ratio; correlation of activation values or lateralizations; intraclass correlation coefficient; intersect maps; conjunction analysis [6]. Past research has shown that the choice of processing parameters, data preprocessing and the threshold level, does influence the results.

Furthermore, studies have shown that certain brain areas can be satisfactorily reproducible, such has the visual cortex [7], the motor cortex [8] and the frontal language areas [9].

Objectives of the current work

The first objective of the present work was to design a robust retinotopic mapping paradigm for the human visual cortex, with particular focus on the primary visual cortex (V1). This paradigm would have to be short enough for acceptable clinical examination durations, but also long enough in order to allow sufficient signal averaging to produce a good SNR and hence good sensitivity. Therefore, the trade-off between exam duration and sensitivity has to be considered. In order to accommodate both requirements, we opted for a simple quadrant stimulation paradigm.

The second objective of the work was to assess the within- and between-session reproducibility of the visual mapping protocol. In order to achieve this, we assessed the variability of a number of parameters closely related with brain activation, in terms of its localization, extent and intensity. As different pre- and post- processing options can influence the results, we also aimed to assess the independence of the result from the combination of processing parameters used.

2. MATERIALS AND METHODS

2.1 Study Design

The study was carried using 5 healthy volunteers (2 male, 3 female) with normal or corrected to normal vision and an age range of 23-35 years old. Informed consent was obtained from each subject according to guidelines approved by our institutional ethical committee. Each subject was scanned on two different days (*Session A* and *Session B*), except for one case were only one session was accomplished. In each session, the subject underwent four fMRI runs (*Run1 – Run 4*). For the experiment, two versions of the stimulation paradigm were created (*Version I* and *Version II*, as explained in Section 2.2): Version I was used for Runs 1 and 3 and Version II was used for runs 2 and 4. The stimulation paradigm was the same for both sessions and no technical or software updates were done on the imaging system during the study.

The four fMRI runs could then be grouped together in order to produce datasets of different sizes. We tested two different dataset sizes, by grouping together either 2 or 4 runs as shown in Fig. 2.1. This resulted in datasets A1, A2, B1 and B2, with ~11 min equivalent acquisition duration, and datasets A and B, with or ~22 min equivalent acquisition duration. We shall recall datasets A1, A2, B1 and B2 as *datasets 11* and similarly, datasets A and B as *datasets 22*.

Firstly Runs 1 and 2 of each session were paired together to yield dataset A1/B1 and Runs 3 and 4 were paired together to yield dataset A2/B2, as schematically represented in Fig. 2.1 (top). On a second analysis, all 4 runs of each session were grouped together, corresponding to doubling the exam duration, as represented in Fig- 2.1 (bottom). Larger datasets are usually acquired in the perspective of increasing activation detection power. By using these two dataset sizes, we can then investigate whether longer acquisitions provide significantly better reproducibility results. This is important because the minimization of exam duration is also a central objective in this work, and so that the trade-off between dataset size and reproducibility should be considered.

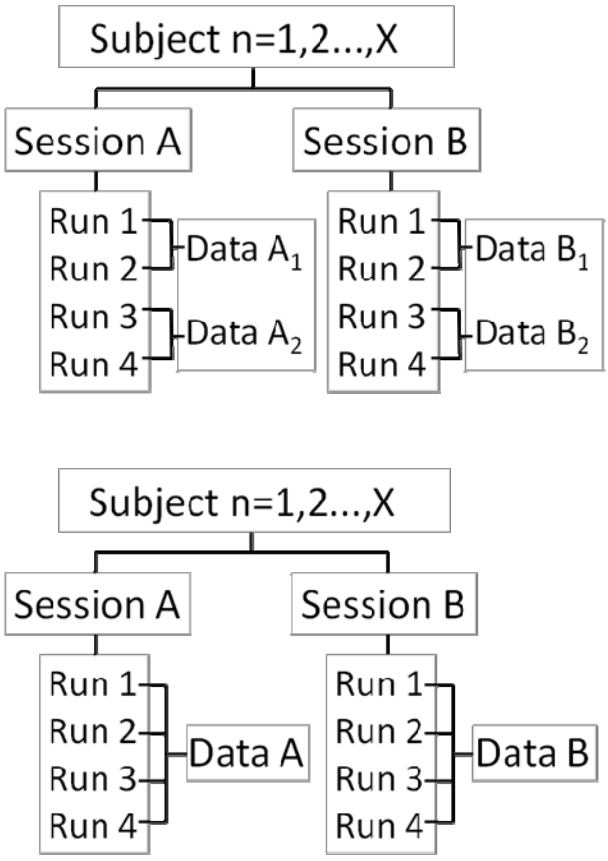

Figure 2.1 – Schematic representation of the study design

2.2 Paradigm (Stimuli and task)

We used a quadrant mapping paradigm, consisting of the presentation of a black and white contrast reversing checkerboard at each of the four quadrants of the visual field in a block design. Each quadrant was stimulated by a wedge approximately 78º wide (we used 78º instead of 90º, in order

to enhance separation of the cortical activation fields for each quadrant). Each square of the checkerboard pattern had an angular width of 6º, yielding a total number of 13 squares per wedge. Figure 2.2 illustrates the four quadrant stimuli. Contrast reversal occurred at a temporal frequency of 8 Hz, which has been shown to produce optimal visual activation [2]. Each subject was asked to focus on a central fixation point during the exam, in order to keep the visual field locked to the stimulation screen. The fixation point used was a yellow circle of 6 pixel radius and separated by approximately 8 pixels from the most central part of the wedge. The paradigm was implemented in C++ using code specifically developed for this work, based on the programme provided by John Serences (http://sites.google.com/site/pcclab/software-downloads).

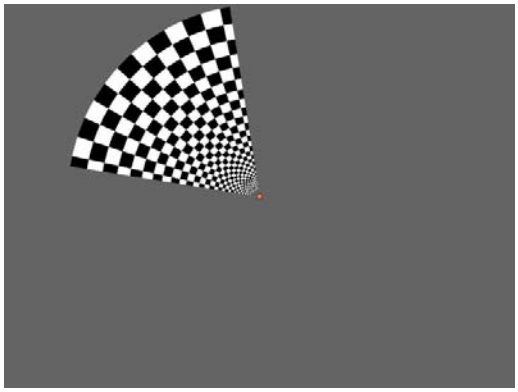

Figure 2.2 – Representation of one of the checkerboard wedges used to stimulate the respective visual field quadrant. Each wedge was named after the quadrant of the visual field that it mapped (Q1 – Q4)

Each fMRI run was built with 21 blocks of 16 seconds duration, yielding a total duration of 5 minutes and 36 seconds. Blocks 1,5,10,16 and 21 were fixation baseline conditions and the remaining were stimulation blocks of one of the four quadrants. All four quadrants appeared in the stimulus the same number of times, in the order shown in Fig. 2.3. The order of the 21 stimulation blocks is symmetric in relation to block 10, i.e. the numeric sequence of the quadrant index is a palindromic number (*capicua*). Although being different, versions I and II of the stimulation paradigm are related: the first part of version I is the second part of version II and vice-versa, as shown in Fig. 2.3.

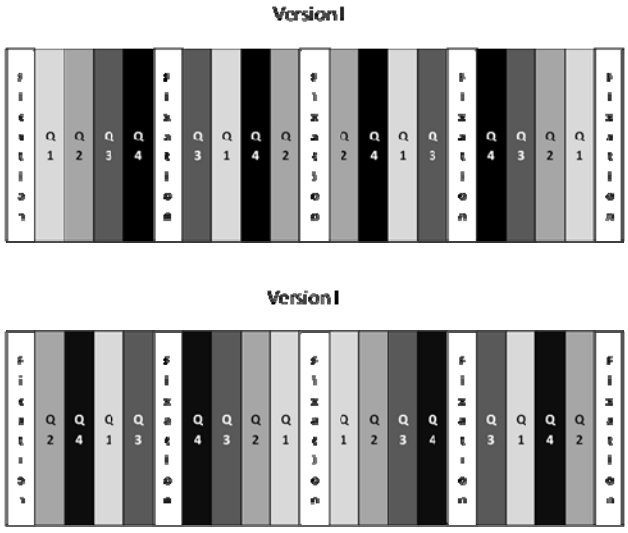

Figure 2.3 – Block structure order of the stimulation paradigm versions I and II

The visual stimuli were backprojected on a translucent screen placed at the feet of the scanner bed, which could be seen by the subject to be examined by means of an angled mirror attached to the head coil. The approximate distance between the subject's eyes and the centre of the screen was 300 cm, translating into a visual angle of

2.3 Image Acquisition

Imaging data were collected using a 1.5 T Philips Gyroscan Intera whole-body system (Philips Medical Systems, Best, The Netherlands).
An Echo-Planar Imaging (EPI) pulse sequence was used to acquire BOLD functional data and a spoiled gradient recalled (SPGR) pulse sequence was used to collect a high-resolution T1-weighted structural image. For the functional images, the scanning parameters were: repetition time TR = 2000 ms, echo time TE = 50 ms, flip angle FA = 90º. The whole brain was covered with a total of 24 oblique slices and an acquisition matrix of 64 x 64. Slice thickness was 5 mm (no gap) and the field-of-view was 240 × 240 mm^2, yielding an in-plane resolution of 3.75 x 3.75 mm^2. A spoiled gradient recalled echo (SPGR) pulse sequence was used to collect high-resolution T1-weighted structural images in the same session, with 1-mm-thick axial slices of 240 x 240 mm^2 field of view and a 256 x 256 acquisition matrix, yielding a reconstructed voxel size of ~1mm^3.

2.4 Image analysis

Data analysis was carried out using a number of tools from FMRIB's Software Library (FSL; Oxford; UK, online at http://www.fmrib.ox.ac.uk/fsl). In particular, FEAT (FMRI Expert Analysis Tool), version 5.92, was used to detect brain activation based on significant changes in BOLD signal. A series of pre-processing steps were applied to each time series of BOLD images, in order to minimize experimental variability and thus improve the validity of further statistical analysis:

- *Non-brain removal*: in all cases, non-brain structures were detected and removed from each image using FSL's Brain Extraction Tool (BET [10]); In this way, all further processing and analysis is restricted to the brain voxels only, for which a binary mask is created;
- *Spatial smoothing*: all images underwent spatial smoothing, using a Gaussian kernel of 5 mm or 8 mm FWHM. Spatial smoothing not only improves the SNR of the data, but most importantly it is used so that the requirements of the Gaussian random field (GRF) theory are fulfilled for subsequent statistical analysis;
- *High-pass temporal filtering*: a Gaussian weighted least squares straight line fitting with 100 s cutoff was performed in order to eliminate undesirable, slow drifts present in the signal;
- *Motion correction*: also used by MCFLIRT motion correction tool [11].
- *Slice timing correction*: a correction was performed to each voxel's time-series for the fact that later processing assumes that all slices were acquired exactly half-way through the relevant volume's acquisition time (TR), whereas in fact each slice is taken at slightly different times. Slice timing correction works by using (Hanning-windowed) sinc interpolation to shift each time-series by an appropriate fraction of a TR relative to the middle of the TR period.

Statistical analysis of the pre-processed time-series was carried out through a general linear model (GLM) approach, using FMRIB's Improved Linear Model (FILM) with local autocorrelation correction. The model is setup in order to test for stimuli-related brain activity changes [12,13]. Task periods (Q1-Q4) were modelled by convolving a square function with stimulus duration width (16 sec) with the canonical Gamma-variate hemodynamic response function (HRF) [14]. The first time derivative of the canonical HRF was also included in the model as a regressor, in order to account for any potential variability in the delay and dispersion of the hemodynamic response across the brain. Motion parameters, three translational and three rotational, totalling six, were further included in the GLM as covariates of on interest, in order to account for any signal variability due to the detected motion. Linear contrasts between the visual stimulation conditions and the controlling baseline were calculated for each explanatory variable (Q1-Q4) and each subject, producing statistical maps of increased brain activity for each visual quadrant.

In order to conduct higher level statistical analyses, all datasets were normalized into a standard space. Low-resolution functional images were first registered to the corresponding high-resolution T1-weighted anatomical image of the subject. The latter was in turn registered to the Montreal Neurological Institute (MNI) template for a standard brain [15], with a spatial resolution of 2 mm, in each of the three directions, so each voxel represents a 2x2x2 mm cube. Finally, all functional images were normalized through co-registration with the MNI space. In each case, registration was accomplished using FMRIB's Linear Image Registration Tool (FLIRT [11,16]).

For producing *datasets 11* (size 2), each normalized COPE image was paired with the complementary normalized COPE image (i.e., Run 1 with Run 2, yielding Dataset A1 / B1, and Run 3 with Run 4, yielding Dataset A2 / B2). For *datasets 22* (size 4), A and B, all normalized COPE images were analyzed together A fixed-effects approach was employed, using Bayesian estimation techniques implemented in FLAME [17,18].

Several combinations of pre-processing and time-series post-statistics options were employed in order to produce diverse and potentially conclusive comparisons, as shown in Table 2.1: In the preprocessing steps, three parameter combinations were used:

- spatial smoothing (Gaussian kernel) with FWHM of 5 mm or 8 mm;
- slice time correction was performed or not;
- motion parameters were included in the GLM or not.

At the post-processing stage, the following approach was used:

- The first one includes thresholding the Z-test significance map ($P < 0,05$) with a clustering tool, inherent to FEAT software, with 2 threshold value analysis, 3,0 and 5,0 ($Z > 3,0$ and $Z > 5,0$).

Table 2.1 - Summary of the combinations of pre-processing and post-processing parameters

Preprocessing	Postprocessing
Spatial Smoothing = {5,8} Slice Time Correction = {0,1} Motion Parameters = {0,1} (8 Combinations)	Cluster Thresholded Statistical Map ($Z > 3,0$) Cluster Thresholded Statistical Map ($Z > 5,0$)

2.5 Reproducibility measurements

In order to evaluate the within-session and between-session reproducibility of the quadrant mapping procedure, we considered a number of comparisons across all possible datasets, as shown in Table 2. Recalling on the study design shown Fig. 2.1., for datasets 11 (size 2), we have 4 combined datasets for each subject: A1 A2 from Session A and B1, B2 from Session B. We can therefore establish 4 comparisons among them, 2 within-session and 2 between-session. We decided not to make the between-session comparisons A1 with B2 and A2 with B1, in order to avoid the confound of possible acquisition order effects. For datasets 22 (size 4), only a between-session comparison was performed: A vs. B.

Table 2.2 - Comparisons established between datasets for reproducibility assessment

Datasets of size 2 (~11 min duration)	
Within-Session Comparison	**Between-Session Comparison**
A1-A2	A1-B1
B1-B2	A2-B2

Datasets of size 4 (~22 min duration)
Between-Session Comparison
A-B

Several reproducibility measurements were performed and are summarized in Table.3. Both localization and extent of activation were assessed, as well as the intensity of activation, using information from Z-statistics maps and the corresponding COPE images. In terms of localization, the peaks of both the Z-statistic and COPE maps were considered. In this case, variability was measured as the distance (D) between the peaks of the two comparison datasets in each case:

$$D = 2 \cdot \sqrt{(V1_x - V2_x)^2 + (V1_y - V2_y)^2 + (V1_z - V2_z)^2} \quad (1)$$

where $Vi_{x,y,z}$ represents the x, y, z coordinate of peak voxel i. Has we are working on the MNI standard space with cubic voxels of 2 mm^3 we convert the digital distance into a metric distance by multiplying it by a 2 factor.

In terms of the extent and intensity of the activation, the relevant statistical cluster in each case was used to define a region of interest (ROI). The volume of the activation cluster was measured, and the mean signal change (COPE) and mean statistical significance (Z statistic) were determined for the ROI. The respective coefficients of variation (CV) between each pair of datasets was then calculated as a measure of variability, according to the equation:

$$CV\ (\%) = \frac{\sigma(D_1, D_2)}{\mu(D_1, D_2)} \cdot 100\% \quad (2)$$

where (is the standard deviation and (is the average of each measurement across the two comparison datasets in each case. As the clustering post-processing is applied to the Z statistical maps, and typically more than one cluster is found, it was necessary to perform a "correct" cluster selection. We assumed that the cluster with the highest statistical score was the one of interest, and thus a simple algorithm was designed to perform this selection automatically. One-by-one inspection of the results showed that a small number of clusters did not verify this assumption. In these cases, the cluster of interest was manually selected.

Besides these standard variability measures, we aimed to further assess the functional reproducibility of the results by estimating the change in mean COPE and mean Z score that would be obtained if the ROI produced by one dataset was used for these calculations in the other dataset. In other words, for each dataset the homologous within and between-session datasets' ROIs were considered and applied on the signal change (COPE) and Z score maps, generating different mean values, a technique used in a similar work (Peelen & Downing, 2005). This functional reproducibility measure assesses the relative change in the COPE (signal change) and Z score maps.

$$Mean\ (Z/COPE)\ Relative\ Change\ (\%) = \frac{|\mu(D_2) - \mu(D_1)|}{\mu(D_1)} \cdot 100\% \quad (3)$$

where $\mu(D_i)$ is the mean Z/COPE of dataset D_i.

Finally, we aimed to assess the efficiency, as well as the reproducibility, of our visual mapping protocol in detecting the activated voxels in each portion of the visual cortex. Our main concern was to minimise the presence of false negatives, because this situation is the most problematic when considering pre-surgical planning applications. In fact, in this case it is preferable to obtain a false positive leading to a more conservative decision in terms of surgical removal of brain tissue. With this concern in mind, we designed one measure of efficiency as follows: the activation clusters detected with each pair of datasets were joined together to produce a cluster reunion; the fraction of voxels of this reunion belonging to each of the individual clusters was then calculated:

$$Detection\ Efficiency\ (\%) = \frac{ROI_{D_1}}{ROI_{D_1} \cup ROI_{D_2}} \cdot 100\% \quad (4)$$

where $ROI_{D\,i}$ is the number of voxels (volume) of dataset D_i and U is the reunion operator.

Another measure of reproducibility was also obtained by determining the fraction of voxels of the reunion belonging

to both individual clusters, i.e., the percentage of common voxels in relation to the reunion voxels, which we shall call cluster overlap percentage:

$$Cluster\ Overlap\ (\%) = \frac{ROI_{S_1} \cap ROI_{S_2}}{ROI_{S_1} \cup ROI_{S_2}} \cdot 100\% \qquad (5)$$

where ∩ is the intersection operator.

Table 2.3 - Reproducibility measures estimated for the quadrant mapping results, for each of the comparisons considered in Table 2.2

Localization	Activation extent	Activation intensity	Detection efficiency
Z peak distance	Cluster Volume CV	Cluster mean Z CV and relative change	Detection efficiency
COPE peak distance	Cluster Overlap Percentage	Cluster mean COPE CV and relative change	

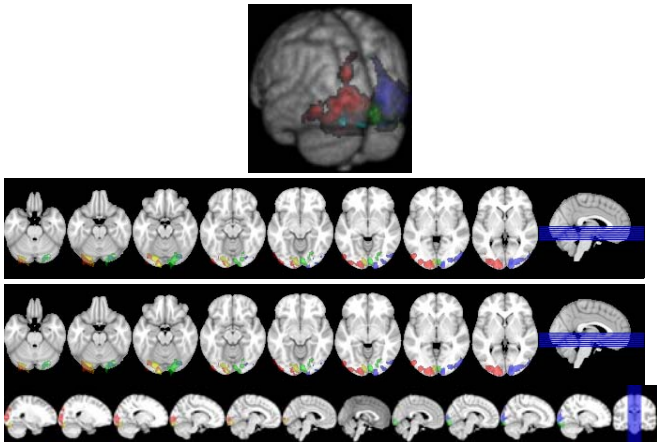

Figure 3.1 – Representative example of quadrant mapping: Representative example of the Z statistic map representing the activation obtained upon stimulation of the four quadrants (Q1 – Q4), overlaid on the template MNI anatomical image, for one subject. In *red* we have quadrant 1, *blue* represents quadrant 2, *green* maps quadrant 3 and *yellow* corresponds to quadrant 4. On the top the 3 dimensional rendering of the image and on the bottom the corresponding coronal, axial and sagittal slice images

3. RESULTS

3.1 Visual quadrant brain mapping

The first output of each dataset will be, among others, images/maps representing the contrasts of parameter estimates (COPE) of each of the explanatory variables and their corresponding Z-score statistical maps. These images, when overlaied on an anatomical template upon registration from one space to another, will show the brain regions that were activated upon stimulation. In our present case, as we are mapping each of the 4 quadrants of a fraction of the visual field, we expect to obtain the cortical representation of each quadrant of the visual field. Fig 3.1 gives a representative example of the final result of the quadrant brain mapping for all four visual field quadrants. Overlaying together the Z statistic maps representing the activation upon stimulation of each of the 4 quadrants, we get the retinotopic map of the full visual field we are stimulating.

3.2 Reproducibility measurements

For each of the 9 measures performed, we directly compared the factors within-session and between-session, as well as dataset dimension (2 and 4, i.e., 11 and 22), resulting in 3 series of comparison data: *within-session - datasets 11* (WS11), *between-session - datasets 11* (BS11) and *between-session - datasets 22* (BS22).

The results were separated by combination of processing parameters, yielding 16 series of data for comparisons WS11 and BS11 and 8 series of data for comparisons BS22 (for these datasets cluster thresholding was only performed using Z > 5,0, because, during the results analysis, this processing option seemed to give better results).

During the statistical analysis process of these series, for each measure, it was clear that the data distributions did not follow normal distributions and therefore we opted to characterize each series based on their median values and considering an interval of confidence between percentile 25 and percentile 75, which accounts for 50% of the data of the series.

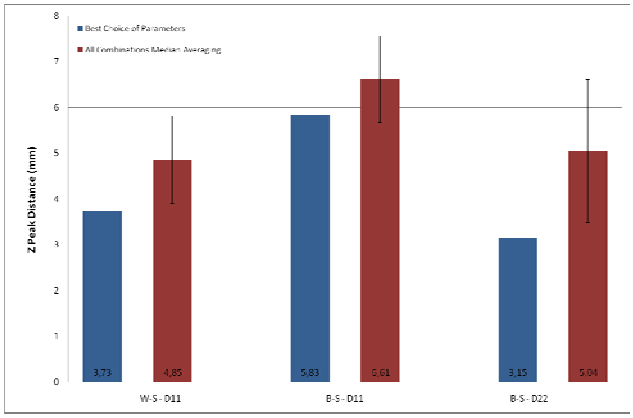

Figure 3.2 – Reproducibility measurements of localization (Z peak) for optimal choice of processing parameters and averaging of median values of all combinations.

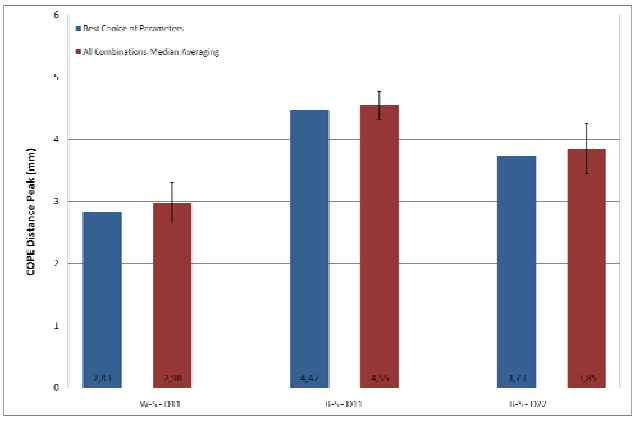

Figure 3.3 – Reproducibility measurements of localization (COPE peak) for optimal choice of processing parameters and averaging of median values of all combinations.

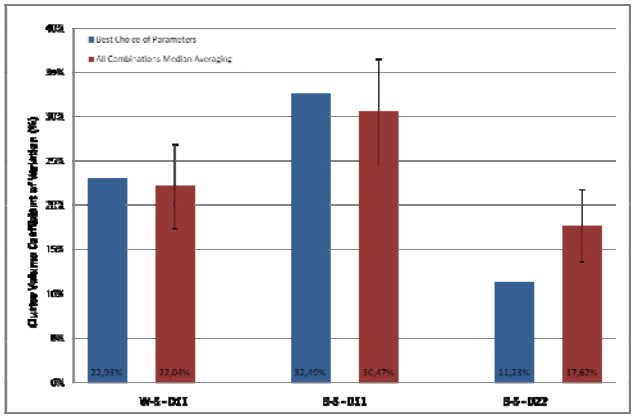

Figure 3.4 – Reproducibility measurements of activation extent (cluster volume CV) for optimal choice of processing parameters and averaging of median values of all combinations.

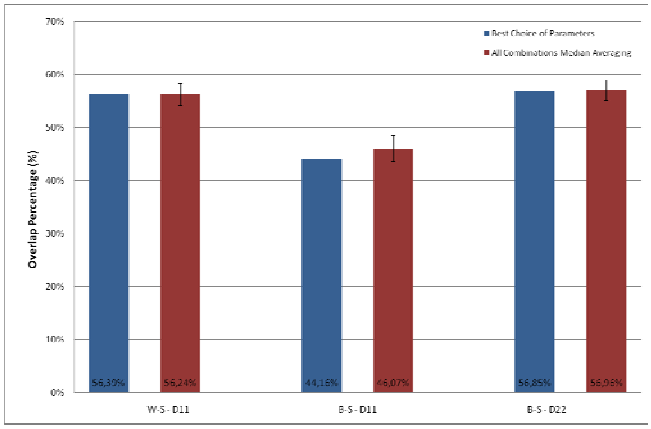

Figure 3.5 – Reproducibility measurements of activation extent (overlap percentage) for optimal choice of processing parameters and averaging of median values of all combinations.

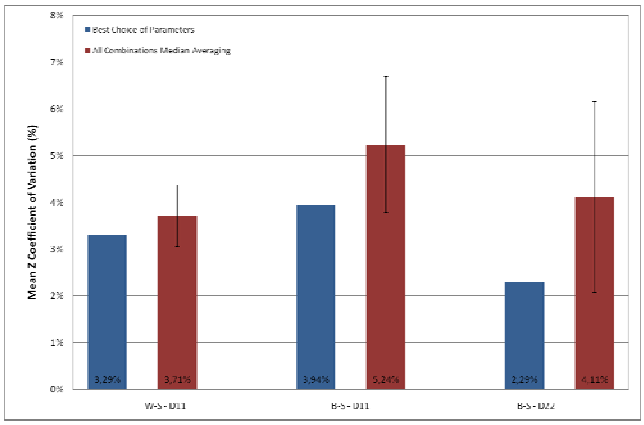

Figure 3.6 – Reproducibility measurements of intensity of activation (Mean Z CV) for optimal choice of processing parameters and averaging of median values of all combinations.

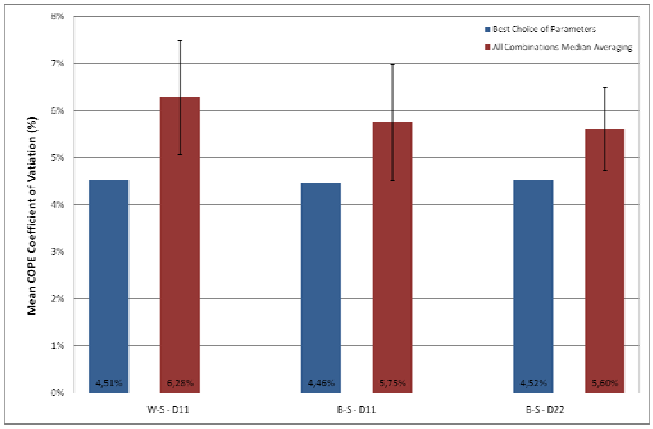

Figure 3.7 – Reproducibility measurements of intensity of activation (Mean COPE CV) for optimal choice of processing parameters and averaging of median values of all combinations.

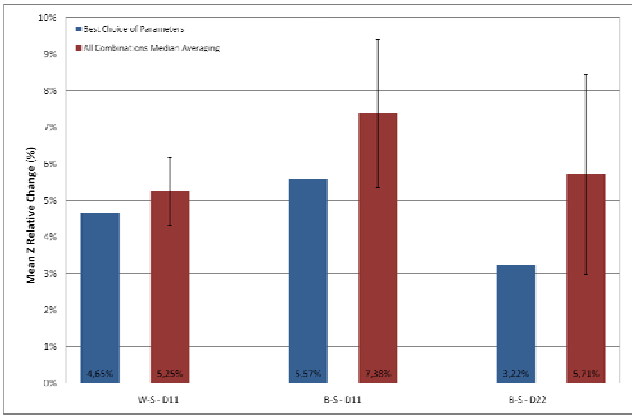

Figure 3.8 – Reproducibility measurements of intensity of activation (Mean Z Relative Change) for optimal choice of processing parameters and averaging of median values of all combinations

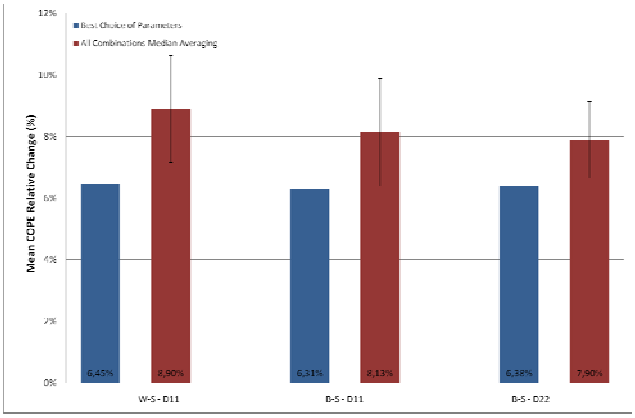

Figure 3.9 – Reproducibility measurements of intensity of activation (Mean COPE Relative Change) for optimal choice of processing parameters and averaging of median values of all combinations

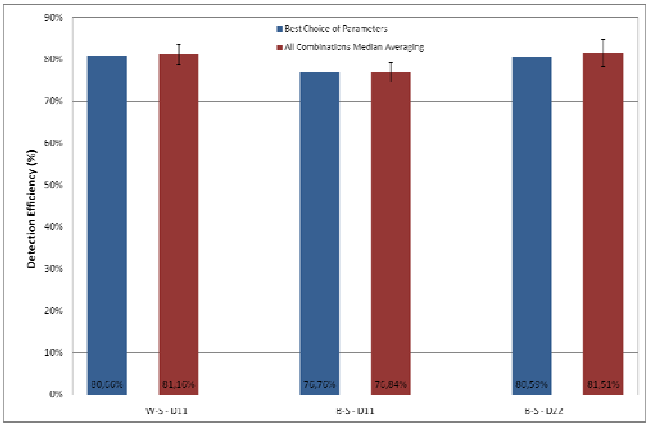

Figure 3.10 – Reproducibility measurements of detection efficiency for optimal choice of processing parameters and averaging of median values of all combinations

Comparing both approaches we essentially verify that the results seem to be similar, apart from some exceptions, for both cases in terms of the relative differences among the 3 types of comparisons assessed.

As expected, between-session comparisons for *datasets 11* (size 2) typically have worst results in relation to both within-session comparisons (*datasets 11*) and between-session comparisons for *datasets 22*. For some measures within-session comparisons yield the best results, such as for Z and COPE peak distances, except for the Z peak distance for the best choice of parameters analysis. In other cases, between-session comparisons using *datasets 22* produce the best results, as for cluster volume CV were the results seem to be much more stable in relation to both comparison types in *datasets 11*. For the mean COPE CV and mean COPE relative change measures very few difference is produced among all three comparisons, and if we consider the average of all medians (of the different processing combinations) we can see that within-session comparisons (*datasets 11*) yield worst results than the other 2 comparison types.

4. DICUSSION AND CONCLUSION

We have developed a complete fMRI protocol for mapping the retinotopic areas of the visual cortex and have subsequently assessed its reproducibility in a group of healthy volunteers. The protocol was designed so as to accommodate the requirements of the clinical setting, where non-cooperative patients may hinder the success of very long and demanding experiments. In order to find a compromise for the trade-off between the duration of the exam and its sensitivity, a simple and short, quadrant stimulation paradigm was implemented.

A robust fMRI protocol is one that yields reproducible results, i.e., results with small variability over a number of experiments. Our results suggest that reproducibility tends to be higher in within-session relative to between-session comparisons, although the difference is not significant for all measures of variability. In terms of the size of the dataset used for analysis, a similar trend for a higher reproducibility was observed when comparing the larger datasets composed of 4 runs with the smaller datasets composed of 2 runs. In fact, within-session variability measures for the smaller datasets were very similar to between-session variability measures of the larger datasets, and both were lower than the between-session variability measures for the smaller datasets. This means that the sources of between-session variability can at least partially be compensated by using larger datasets, with the disadvantage of longer acquisition times.

In terms of the combinations of processing options, no significant systematic differences in the reproducibility of the results could be detected. This means that the result of the protocol is roughly independent of the exact parameters used in the analysis of the data, which is consistent with a robust and reproducible protocol.

Let us analyze each of the reproducibility measures separately. In terms of localization of brain activation, average median Z peak distance (among all processing parameters combinations) is in the range of ~5 to 8 mm, while for percent signal change peaks the distances range from ~3 to 6 mm. In both cases, within-session comparisons represent the smallest distances and between-session comparisons (smaller datasets) are responsible for the largest. Recalling on the "golden rule" for tumour resection, the safety distance from functional areas and tumour boundaries obtained for previous studies was of about 10 mm. As our result represents a distance between peaks and not boundaries, it would be necessary to assess to what extent it is possible to extrapolate peak distances to ROIs' boundaries distances in order to compare our results with the safety distance "golden rule". Nevertheless, our results suggest that this safety distance in the case of occipital lobe lesions could possibly be considerably reduced, allowing the optimization of tumor resection planning. A similar study assessing reproducibility evaluated visual areas (FFA - Fusiform Face Area, OFA - Occipital Face Area, PPA - Parahippocampal Place Area PPA and EBA - Extrastriate Body Area) on datasets of size equivalent to our size 4. They obtained peak distances between 1.5 mm (within-session) and 2.9 mm (between-session) [19]. These results are in the same range of the ones obtained in this study, corroborating the validity of our reproducibility assessment.

Relatively to the extent of activation, we can observe that the cluster volume CV measure is somewhat variable. By analyzing the average median cluster volume CV we get values between ~20% and 30 %. We therefore conclude that this measure might not be adequate to assess the extent of brain activation in a reproducible way. The cluster overlap measure, however, revealed a relatively low variability ranging from 46% to 57% in the average median assessment, with a low standard deviation. Again the best results were obtained for within-session comparisons and higher size datasets. A reproducibility study of memory related activation of the medial temporal lobe used a cluster overlap measure of reproducibility and obtained values around ~18% cluster overlap [6] Although the brain areas studied are very different, compared to ours (~45% to ~55%) these results bring confidence to the reproducibility assessment of our paradigm.

With respect to the intensity of activation, we have an average median mean Z CV ranging from 4 to 7 % and for percent signal change (COPE) a range of 6 to 8 %. For the related measure mean Z/COPE relative change we get an average median between 6 and 10 % for Z and 9 to 11 % for COPE. In both cases for Z score the best (lowest) values were from within-session comparisons and the worst (highest) for between-session comparisons (datasets of size 2). In opposition, for the COPE signal all 3 comparisons showed similar results with the highest (worst) yield by within-session and the best (lowest) by between-session comparisons of datasets with size 4. In terms of intensity, a study evaluating cerebral motor areas got a range of coefficients of variation between ~15% to 30% which, again, is consistent with the validity and relevance of the present study [20].

Finally, average median detection efficiency had a range between ~77 and 82 %, again with between-session comparisons of datasets of size 2 assuming the low values of the range and the other 2 comparisons equally sharing the high value of the range.

This study has several limitations that we note here. A number of external factors, such as eating habits and medication, were not taken into account in the present study. However, it is well known that coffee (caffeine) can produce quite variable results in the BOLD signal of a subject due to its vasodilation properties [21]. Furthermore, it would have been interesting to assess the inter-subject variability of the visual mapping results since this is probably very high and makes an important contribution to the observed variability of the our results. Also, the anatomical variability between different subjects could have hindered the alignment of the structural images with one common template.

Similarly to every "new product", the first assessment is normally given to make improvements for the subsequent version. Still, as a first assessment the results are globally acceptable which translates into a robust-with-potential-to-improve paradigm that is able to compete with the state-of-art of the "golden rule" of fMRI guided tumour resection surgery.

Future work should focus on improvements on the paradigm and on the choice of the most appropriate reproducibility measurements. Furthermore, studies of specific clinical cases are necessary in order to assess the robustness of the paradigm in the conditions it was created for.

5. ACKNOWLEDGEMENTS

I thank Dr. João Xavier, Dr. João Teixeira and Cláudia Azevedo at Ginoeco for providing the conditions and assistance for data acquisition. I acknowledge the financial support by Projecto Fundação BIAL No.16/04.

6. REFERENCES

[1] Tootell, R. et al. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in Cognitive Sciences - Vol.2, No.5* , 174-183.

[2] Warnking, J. e. (2002). fMRI Retinotopic Mapping - Step by step. *NeuroImage 17* , 1665-1683.

[3] Sereno, M. e. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance. *Science 268* , 889-893.

[4] Engel, S. e. (1994). fMRI of human visual cortex. *Nature 369* .

[5] Sunaert, S. (2006). Presurgical Planning for Tumour Resectioning, 23. *Journal of Magnetic Resonance Imaging*, 887-905.

[6] Wagner, K. e. (2005). The reliability of fMRI activations in the medial temporal lobes in a verbal episodic memory task. *NeuroImage 28* , 122-131.

[7] Miki, A. e. (2000). Reproducibility of visual activation in functional MR imaging and effects of postprocessing. *AJNR Am J. Neuroradiol. 21* , 910-915

[8] Yetkin, F. e. (1996). Test-retest precision of functional MR in sensory and motor and task activation. *AJMR Am J. Neuroradiol. 17* , 95-98

[9] Brannen, J. e. (2001). Reliability of functional MR imaging with word-generating tasks for mapping Broca's area. *AJNR Am. J. Neuroradiol. 22* , 1711-1718

[10] Smith, SM. (2002). Fast robust automated brain extraction. *Human Brain Mapping 17* , 143-155.

[11] Jenkinson, M. e. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brains images. *Neuroimage 17* , 825-841.

[12] Friston K.J., e. a. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping* , 214-220.

[13] Woolrich MW, e. a. (2001). Temporal autocorrelation in univariate linear modelling for fMRI data. *Neuroimage 14* , 1370-1386.

[14] Boynton, G. E. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J.Neurosci., 16* , 4207-4221.

[15] Collins DL et al. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *Proc IEEE-Nuclear Science Symposium and Medical Imaging Conference* , 1813-1817.

[16] Jenkinson M, Smith S. (2001). A global optimization method for robust affine registration of brain images. *Med Image Anal 5* , 143-156.

[17] Beckmann CF et al. (2003). General multilevel linear modeling for group analysis in fMRI. *Neuroimage 20* , 1052-1063.

[18] Woolrich MW et al. (2004). Multi-level linear modelling for fMRI group analysis using Bayesian inference. *Neuroimage 21* , 1732-1747.

[19] Peelen, V. M., & Downing, E. P. (2005). Within-subject reproducibility of category-specific visual activation with functional MRI. *Human Brain Mapping* , 402-408.

[20] Tjandra, T. et al. (2005). Quantitative assessment of the reproducibility of cuntional activation measured with BOLD and MR perfusion imaging: Implications for clinical trial design. *NeuroImage 27* , 393-401.

[20] Laurienti, P. et al. (2002). Dietary Caffeine Consumption Modulates fMRI Measures. *NeuroImage 17* , 751-757.