INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

Reproducibility of visual brain mapping using fMRI

Pedro Miguel Guerreiro Brito da Cruz

Dissertação para a obtenção de grau de mestre em

Engenharia Biomédica

Júri

Profª Drª Patrícia Figueiredo
Prof. Dr. Fernando Lopes da Silva
Prof. Dr. Mário Secca
Dr. João Xavier

Setembro 2008

Acknowledgements

Ao longo do meu trabalho, no âmbito da tese de mestrado, reuni importantes contribuições de diversas pessoas e entidades, sem as quais esta investigação não teria sido possível.

Em primeiro lugar quero agradecer à minha coordenadora, a Prof. Patrícia Figueiredo, por todo o apoio, interesse e motivação, bem como pela constante partilha de "*know-how*", que muito contribuiram para este trabalho, em particular nos momentos mais críticos.

Deixo um agradecimento especial à Ginoeco, nomeadamente ao Dr. João Xavier, ao Dr. João Teixeira e à Cláudia Azevedo, pela colaboração neste projecto oferecendo as condições de trabalho para a aquisição das imagens de ressonância magnética funcional.

Gostaria também de agradecer à Fundação BIAL pelo apoio financeiro à realização das experiências através da Bolsa No.16/04.

Deixo ainda uma nota de agradecimento aos voluntários que gentilmente "cederam o seu cérebro" em favor deste trabalho de investigação.

Finalmente, quero agradecer à minha família e amigos pelo incentivo e "força" que me deram durante esta etapa.

Abstract

Functional MRI (fMRI) techniques are extensively used today to assess brain activation upon the application of specific stimuli or tasks. The human visual cortex, in particular, exhibits a retinotopic organization such that neighbouring areas in the brain respond to the stimulation of proximal positions in the visual field. It is possible to use fMRI to perform retinotopic mapping in order to determine this correspondance. Retinotopic mapping methods 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

Resumo

As técnicas de Ressonância Magnética funcional (RMf) são, hoje em dia, usadas para avaliar a activação cerebral quando aplicado um estímulo específico ou perante uma tarefa. O córtex visual humano apresenta uma organização retinotópica, ou seja, regiões adjacentes no córtex respondem a estímulos visuais adjacentes no campo visual. É possível utilizar RMf na determinação desta correspondência. Os métodos de mapeamento retinotópico podem ser utilizados para avaliar situações de re-organização em casos de doenças com perda de campo visual ou ainda contribuir no planeamento pré-cirúrgico de remoção de cancro. Em consonância com estas responsabilidades clínicas o mapeamento retinotópico requer um protocolo de estimulação visual robusto e com variabilidade mínima. O objectivo deste trabalho foi desenvolver um protocolo que cumprisse tais requisitos e determinar a respectiva reprodutibilidade. Implementou-se um paradigma de estimulação baseado em quadrantes (do campo visual) e testou-se este paradigma num universo de 5 voluntários em dois dias distintos (sessões), com 4 repetições ("*runs*") por sessão.

Os resultados foram analisados utilizando séries de dados com 2 dimensões distintas ("*datasets"* compostos por 2 ou 4 "*runs*") e também diferentes combinações de parâmetros de pré- e pós processamento. Avaliou-se a reprodutibilidade intra- e inter-sessões, utilizando os 2 tipos de "*datasets*" através do cálculo de diferentes medidas de variabilidade dos diferentes parâmetros de activação.

Em geral, verificou-se que a reprodutibilidade era maior para comparações intra-sessão comparativamente com comparações inter-sessão. Para além disso, verificou-se que "*datasets*" compostos por mais "*runs*" produziam resultados mais reprodutíveis. Em todo o caso, não foram registadas diferenças significativas, em termos de reprodutibilidade, no uso de parâmetros de processamento diferentes. Escolhendo a combinação de parâmetros mais favorável, a distância mediana entre picos de activação foi de ~3 mm tanto para comparações intra-sessão de "*datasets*" de dimensão 2 como para comparações inter-sessão de "*datasets*" de dimensão 4. Para comparações inter-sessão de "*datasets*" de dimensão 2, este valor aumentou para ~5 mm. Em termos de volume e magnitude de activação, os coeficientes de variação da mediana variaram entre ~ 11 e 33 % para o volume de activação e entre ~ 2 e 4 % para a variação média de sinal, para comparações inter-sessão de "*datasets*" de dimensão 4 e 2, respectivamente. Como medida global de reprodutibilidade, a mediana da percentagem de sobreposição de voxels variou entre ~ 44 e 57 % para comparações inter-sessão de "*datasets*" de dimensão 4 e 2, respectivamente.

Em suma, desenvolveu-se um protocolo de mapeamento visual robusto baseado em RMf que detecta e localiza a activação retinotópica do cortex visual associada à estimulação de quadrantes do campo visual.

Palavras-chave: RMf, Retinotopia, Mapeamento Visual, Actividade Cerebral, Reprodutibilidade

Contents

List of Abbreviations

BET	Brain Extraction Tool
BOLD	Blood-Oxygenation-Level-Dependent
CBF	Cerebral Blood Flow
CBV	Cerebral Blood Volume
COPE	Contrast Of Parameter Estimate
CV	Coefficient of Variation
EPI	Echo-Planar Imaging
FA	Flip Angle
FEAT	FMRIB's Expert Analysis Tool
FILM	FMRIB's Improved Linear Model
FLAME	FMRIB's Local Analysis of Mixed Effects
FLIRT	FMRIB's Linear Image Registration Tool
fMRI	functional Magnetic Resonance Imaging
FMRIB	Functional Magnetic Resonance Imaging of the Brain
FSL	FMRIB's Software Library
FWHM	Full Width Half Maximum
HRF	Hemodynamic Response Function
MRI	Magnetic Resonance Imaging
ROI	Region of Interest
SNR	Signal to Noise Ration
SPGR	Spoiled Gradient Recalled
TR	Time of Repetition
TE	Time of Echo
VFS	Visual Field Sign

List of Figures

List of Tables

Chapter 1

Introduction

Since the late years of the 19th century, with the first steps of radiography, medical imaging has suffered a great evolution, increasing the effectiveness and accuracy of medical diagnosis in such a way that it is now a central area in modern medicine. Among the various imaging techniques, Magnetic Resonance Imaging (MRI), being the youngest modality, is also commonly regarded as one of the most dynamic. Its capacity to produce high contrast images between soft tissues, particularly in the brain, together with its versatility in producing a variety of types of contrastmakes it a very good candidate to assess the human brain in a "functional" perspective. In particular, it is possible to evaluate which parts of the brain respond to a certain stimulus or are responsible for a specific task. Furthermore, along with its technical advantages, MRI also excels for its safety, as it produces no harm to the patients, as opposed to to ionizing radiation based imaging techniques. Over the last years a great research effort has been dedicated to this modality of MRI, commonly known as functional MRI (fMRI), with hundreds of studies being published, increasingly, every year. The fMRI technique has already contributed enormously to basic Neuroscience research and is now at the stage of becoming a useful tool in the clinical setting as well. In order to fulfill the great potential of this translational scientific effort, a strongly multi-disciplinary work involving biomedical engineers and clinicians is necessary.

1.1 Principles of functional MRI

1.1.1 Basic principles of MRI

Magnetic Resonance Imaging (MRI) is based on different magnetic properties of tissues, which ultimately yield different image intensities, enabling a distinguishable contrast between tissues. The MRI signal relies on the abundance of hidrogen atoms in the human body. Each hydrogen nucleus has a spin property such that it induces a magnetic moment. Because the spin number of the hydrogen nucleus is ½, most of the phenomenology associated with MRI can be described using a classical approach. In natural

circumstances, each magnetic moment points in a random direction, generating a global net magnetization that is infinitesimally small. If one applies a specific static magnetic field B_0, this will make the nuclei to precess (gyroscopic motion) around an axis along the direction of B_0. This structured orientation of the nuclei will generate a non-zero net magnetization M_0 in the direction of the static magnetic field B_0. If one then applies an electromagnetic excitation to the nuclei at their resonance frequency, dictated by the Larmor equation ($\omega_L=\gamma B_0$), their magnetic moment will change. As the transient electromagnetic excitation ends, the nuclei tend to return to their initial equilibrium condition in a process called relaxation, which has specific time constants associated with it. The fundamental MRI contrast mechanisms are based on the fact that different tissue types, besides having different proton densities, also have different relaxation times and thus can produce different signal intensities when assessed at a specific time point during the relaxation process.

1.1.2 Brain activity and the BOLD effect

In oposition to structural MRI, functional MRI (fMRI) demands an insight of the processes involved in neuronal activity detection. The question is: "how does one assess neuronal activity based on the magnetic properties of the brain?". In the late 1990s, Ogawa and colleagues showed that different blood oxygenation levels induced different image intensities (S. Ogawa, December 1990). Previous studies had shown that paramagnetic deoxyhemoglobin locally produces smaller image intensities, as it causes local magnetic field inhomogeneities (distortions), while diamagnetic oxyhemoglobin had negligible effects on the magnetic field. Ogawa et al. manipulated blood oxygenation on an animal model to measure the so called blood oxygenation level-dependent (BOLD) contrast and suggested its potential application in brain activity detection. In 1992, the groups headed by Ogawa and Kwong simultaneously published the first reports of the measurement of BOLD signal changes related to brain activation in response to sensory stimulation (Kwong, 1992) (Ogawa, 1992). Despite the fact that BOLD fMRI assesses neuronal brain activity in an indirect way, it is today the most widespread fMRI technique.

As we could expect, increased brain (neuronal) activity results in increased metabolic demands, which in turn elicit changes in the local hemodynamics. Altough the process is not fully understood, there are some consensual features that explain, to some extent, the relation between the fMRI (BOLD) signal and hemodynamic parameters, such as the cerebral blood flow (CBF), volume (CBV) and oxygen saturation (Nikos K. Logothetis, 2001). The time response of BOLD fMRI to a short duration, spatially localized neuronal activity pulse is known as the hemodynamic impulse response function (HRF). Because BOLD fMRI produces an indirect measure of brain activity, it is of central importance to understand the underlying signaling events in order to understand the limitations of the technique and its interpretation in assessing brain activity. Biophysical models have been developed to describe the HRF as a function of neuronal activity and brain hemodynamics. Moreover, experiments have been carried out in order to measure directly the relationship between the BOLD signal and the underlying electrical activity of the neurons.

1.1.2 The hemodynamic response function

The HRF consists of three stages describing the BOLD signal behavior in response to a transient increase in neuronal activity, as illustrated in Figure 1.1 (Heeger, 2002). Following an increased neuronal activity the metabolic demands will also increase. This results in an increased oxygen consumption that will locally induce variations in concentration of both oxy- and deoxy-hemoglobin, which will respectively decrease and increase. As seen before, the higher the concentration of deoxyhemoglobin is the lower the image contrast produced, i.e., a below baseline BOLD signal decrease (Vanzetta, 1999). This initial response stage is known as initial dip, and it is somewhat controversial (Heeger, 2002). The second stage of the HRF has a 2 second delay, approximately, relative to the initial neuronal stimulus. After the initial dip, a large increase in CBF is verified, which accounts for for the glucose consumption increase. Along with an augmented CBF, and due to vessel elasticity, CBV is also increased, as described by Buxton's ballon model (Buxton, 1998). According to this model, the increased CBF causes a higher inflow of blood in the venous system rather than blood outflow, yielding an augmented CBV. Despite the higher oxygen consumption, CBF and CBV increases overcome the absolute increase in deoxyhemoglobin, whose concentration will decrease. Such events will trigger an increase in BOLD signal, which will reach a peak. The third and final stage of the HRF comprehends a decrease in both CBF and CBV, though at different rates. The CBF decreases much faster than the CBV which results in a raised concentration of deoxyhemoglobin, consequently induces BOLD signal to go below baseline for a certain period, while CBV returns to a normal value. This stage is known as overshoot. The graphical representation of the HRF is shown in Figure 1.1 - A.

Altough Buxton's work has contributed to a better understanding of the physiological processes involved in BOLD fMRI, it has also raised some issues on the spatial resolution of the technique. As the BOLD signal relies mostly on vein drainage processes, it might be questioned if one is detecting brain activity on its real site or on the neighbour veins. Another issue is the fact that HRF varies among individuals and brain regions which can be misleading in BOLD fMRI signal interpretation. Besides these complications, several assumptions are made in order to obtain brain activity measures from the analysis of BOLD signal changes. Nevertheless, empirical findings have consistently shown that the HRF can be approximated by a gamma function in most situations. This is the model used in the large majority of fMRI studies.

The analysis of fMRI data is based on the linear transform hypothesis, which states that BOLD responses are proportional to local average neural activity averaged over a period of time (Figure 1.1- B). The first study to corroborate this hypothesis (Boynton, 1996) used three different tests to evaluate this assumption, based on human primary visual cortex (V1) stimulation, which were, in the end, compatible with the hypothisized model. In their data, Boynton et al. (Boynton, 1996) verified that fMRI responses on V1 depend separately on stimulus timing and stimulus contrast. The second test was to check whether responses to long duration stimulus could be predicted based on shorter duration stimulus fMRI responses, and finally the independence of fMRI data noise and stimulus contrast and temporal period. Furthermore this study presented a model of the "fMRI impulse response function", commonly know as

hemodynamic response function, which is the (BOLD) fMRI time response to a brief, spatially localized pulse of neural activity, which was described by a gamma function. More recently a study recorded simultaneously fMRI responses and neural signals, measured intracortically, proving undoubtably that BOLD signal increase reflects increase neural activity. On the other hand the obtained "double" data was consistent with previous description of a lineariy relation between BOLD and neural activity.

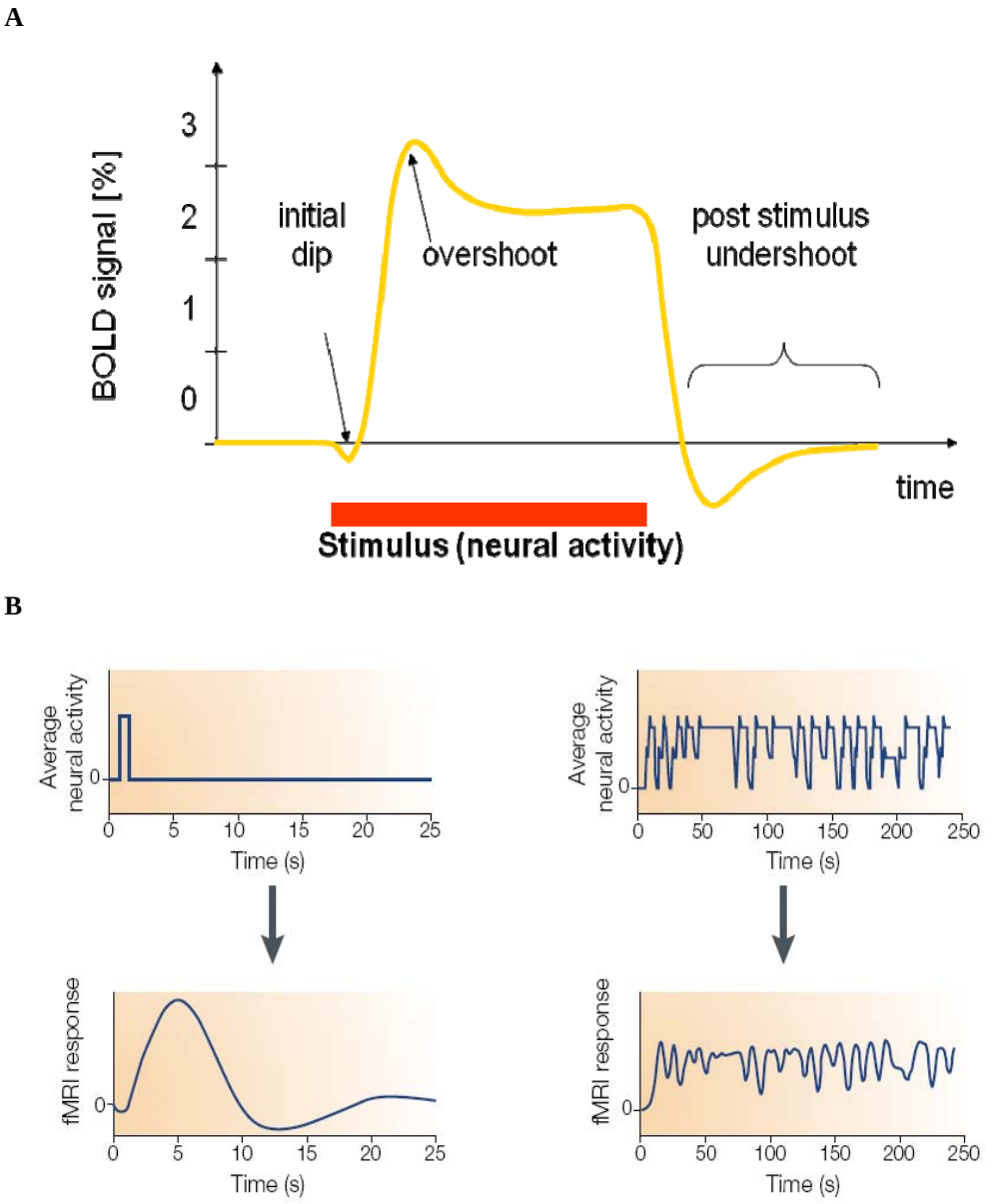

Figure 1.1 – Illustration of the Hemodynamic Response Function: **A**) BOLD signal change measured in response to a stimulus of a finite duration (courtesy of Patricia Figueiredo, IST). **B**) BOLD response to an impulse of neural activity, corresponding to the HRF (Left) and BOLD response to a a series of impulses of neual activity, resulting from the linear summation of the HRF (Right) (adapted from Boynton et al., 1996).

1.1.4 Image analysis

In order to detect brain activation it is necessary to process each dataset with a number of tools that try to improve the quality of the image and thus reduce to a minimum the level of undesired noise. The image analysis process comprehends 4 main steps.

- First of all it is necessary to pre process the data to increase its signal to noise ratio (SNR) and prepare it for the statistical analysis. Spatial smoothing, with a gaussian FWHM above 5 mm, is typically performed to further validate the Gaussian Random Field assumption (which assumes that the activation map follows a Gaussian curve) and also in terms of computation, reducing the number of effective statistical tests to perform. High-pass filtering is extensively used as well, to remove undesired noise-related low frequencies. Motion correction is normally done by correcting all the volumes to a specific one (typically the middle one), i.e., applying transformations that minimize deviation. Slice time correction is, according to user preferences, used in order to correct for the time deviation that each slice has.
- The next step is to perform the statistical analysis. Based on our block-type time-series (or task) it is necessary to convolve it with the HRF in orther to obtain our explanatory variable (regressor) that will be used, upon determining the best weighting (coefficients), to fit optimally to the experimental (now pre processed) data.
- In post processing, the obtained weighting coefficients map from the statistical analysis is thresholded for significance, typically with a Z or T map with the respective p value. Applying thereafter a threshold we obtain the "activated areas" (so to speak).

1.2 Clinical Applications and Reproducibility

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 sistematically manipulating the input in order to study the outputs makes it a very "trackable" brain function (Tootell, 1998).

1.2.1 Retinotopic organization of the visual cortex

The human visual cortex is divided into several functional areas that have different local neuronal properties (Warnking, 2002). It is widely accepted that there is a hierarchy in these functional areas. Some of there 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 homeomorphy, 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). Altough being an 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.

1.2.2 Mapping the Human Virtual Cortex

The concept of retinotopic mapping only gained form when visual field sign mapping (VFS) was developed (Sereno, 1995). 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 (centered on the fixation point) which permitted retinotopic mapping in a phase-locked way, in oposition to the rotating wedge, that allowed mapping from the foveal visual area to the retinal periphery (Engel, 1994).

Eccentricity mapping and polar angle mapping (which is conceptually similar to rotating wedge mapping), although mapping the same visual field, extract distinct retinotopic properties. Eccentricity mapping as a radial varying stimulus will exhibit the magnification property of retinotopy. As seen on Fig.1.3 (left) colors located in foveal regions will have a corresponding larger cortical region in relation to peripheral colors (although the blue color does not appear as it is located in a sulcus region of the cortex). Polar angle mapping allows direct correspondence from angular colors in the visual field into the corresponding regions of the visual cortex, as seen in Fig.1.3 (right). For this type of mapping the symmetry property of retinotopy can be seen with lower visual field colors being mapped on higher regions of the visual cortex and similarly lateralized colors mapped on contralateral regions. For both cases color continuity is verified.

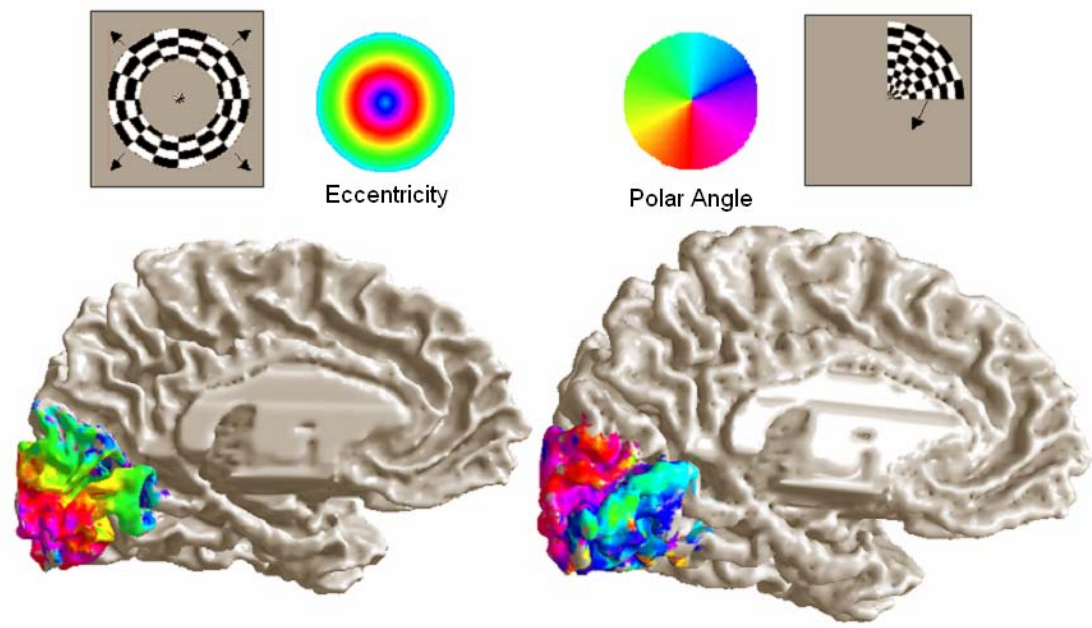

Figure 1.2 – Retinotopic mapping of the visual field: eccentricity mapping using an expanding ring (left) and polar mapping using a rotating wedge (right).

1.2.3 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 (Sunaert, 2006): 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 tumor margin and essential structure for effective nearly risk-free surgical resection is of 10 to 15 mm (Sunaert, 2006). 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. (Sunaert, 2006)

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.

1.2.4 Variability in fMRI measurements

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

- The equipment itself introduces important noise components on the data. Constant field inhomogeneities or gradient differences, even minimal, contribute sistematically 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 a 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
- Behavioral – 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 undoubtably yield brain activation in certain cortical reagions. 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 (Wagner, 2005). 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 (Miki, 2000), the motor cortex (Yetkin, 1996) and the frontal language areas (Brannen, 2001).

1.2.5 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 achive this, we assessed the variability of a number of parameters closely related with brain activation, in terms of its localization, extent, intensity and stability. 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.

Chapter 2

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 togther to yield dataset A1/B1 and Runs 3 and 4 were paired togther 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 toghether, 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.

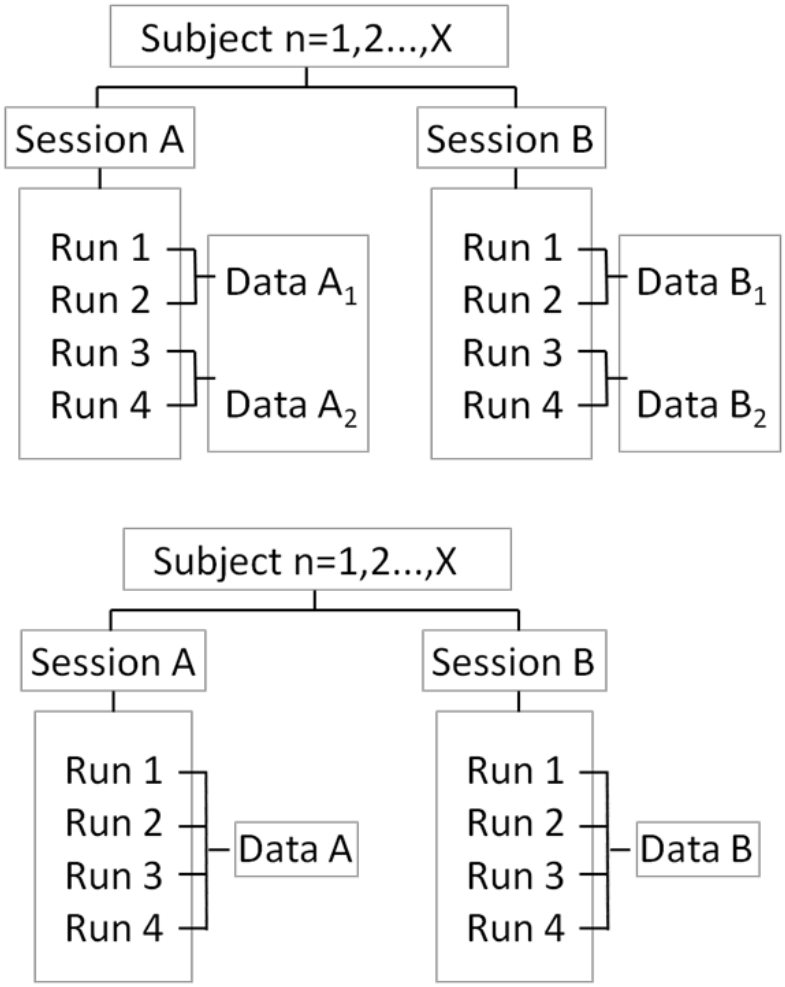

Figure 2.1 – Schematic representation of the study design. Top: Grouping of runs in order to produce datasets of size 2 (~11min duration – *datasets 11*): A1, A2, B1, B2; Bottom: Grouping of runs in order to produce datasets of size 4 (~22 min duration – *datasets 22*): A and B.

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 (Warnking et al., 2002). 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).

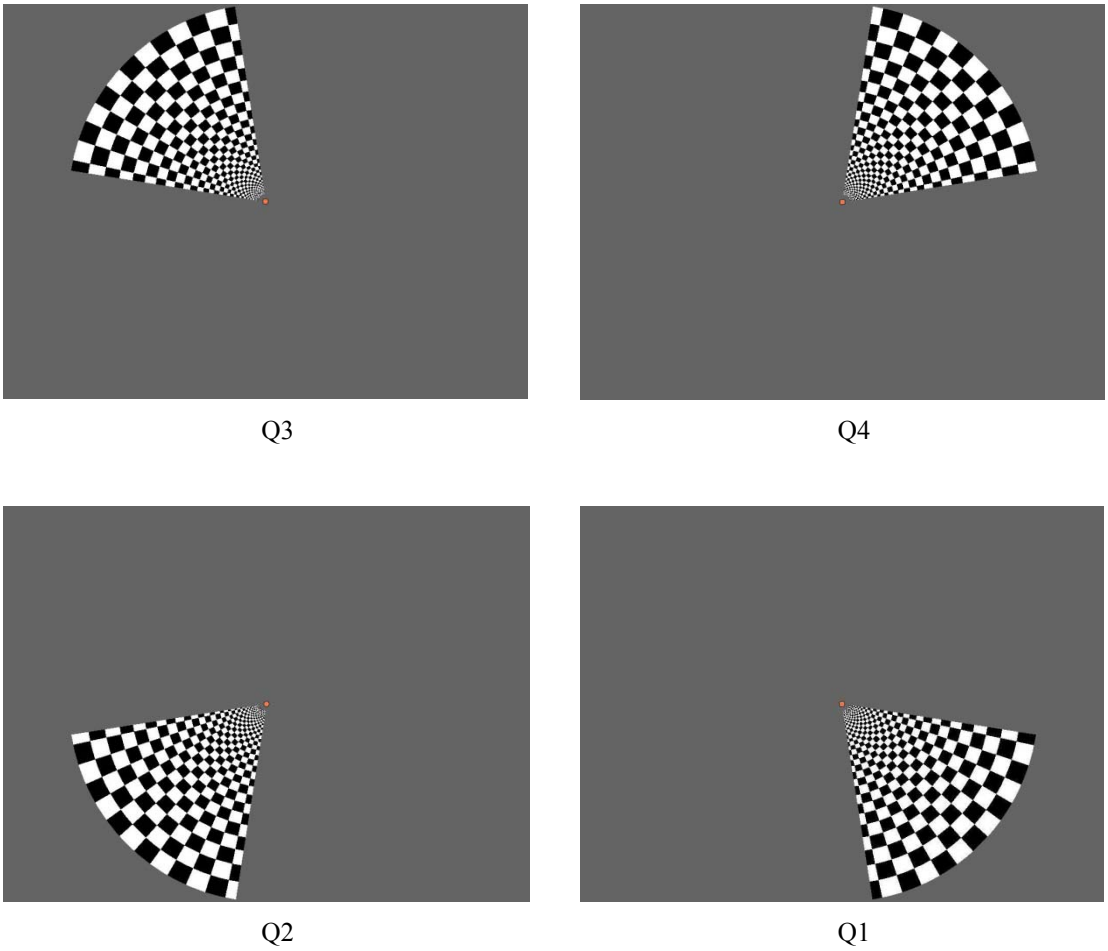

Figure 2.2 – Representation of the checkerboard wedges used to stimulate each of the 4 visual field quadrants. Each wedge was named after the quadrant of the visual field that it mapped (Q1 – Q4).

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 simetric in relation to block 10, i.e. the numeric sequence of the quadrant index is a palindromic number (*capicua*). Althogh 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.

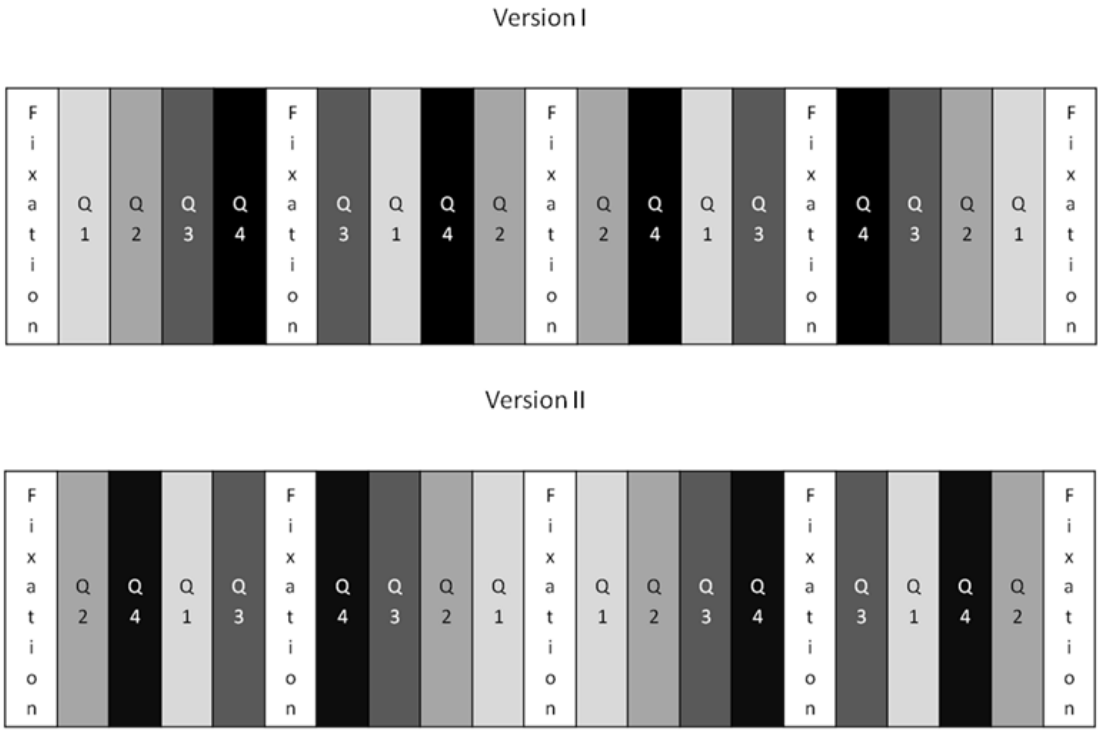

Figure 2.3 – Block structure order of the stimulation paradigm versions I and II.

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.

As we are performing visual mapping, it is important to determine the visual angle that the stimulus produces in the subject's visual field. As illustrated in the scheme of the visual stimulation setup shown in Figure 2.4, we can compute the visual angle (*V*) that we are using in our visual mapping exam, based on the visual stimulus width (*S*) and the distance between the subject's eyes and the projection screen (*D*), according to the equation:

$$\alpha = 2 . arctan\ (S/2.D) \qquad (1)$$

The stimulus has a radial configuration and produces a backprojected image of approximately 100 cm in diameter (S = 100 cm). The distance between the subject's eyes and the center of the stimulus was of approximately 300 cm (D = 300 cm). Therefore, the used visual angle was approximately V ~ 19º.

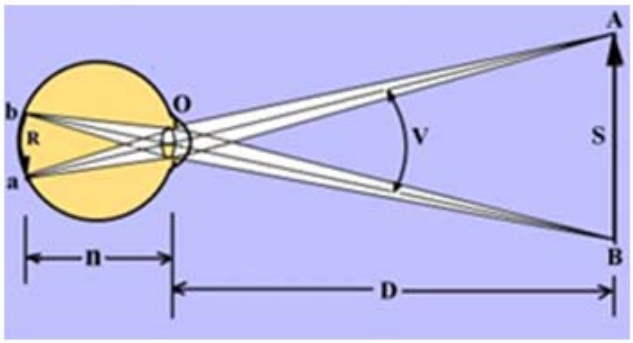

Figure 2.4 – Visual angle calculation (in http://en.wikipedia.org/wiki/Image:EyeOpticsV400y.jpg).

2.3 Image Acquisition

Imaging data were collected on a 1.5 T Philips Gyroscan Intera whole-body system, using a head coil (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. Total duration of the four functional scans was ~22 min (5 min 36 sec per run) A spoiled gradient recalled echo (SPGR) pulse sequence was used to collect high-resolution T_1-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 ~1 mm^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 [Smith, 2002]); 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 [Jenkinson et al., 2002].
- *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 [Friston et al., 1994; Woolrich et al., 2001]. Task periods (Q1-Q4) were modeled by convolving a square function with stimulus duration width (16 sec) with the canonical Gamma-variate haemodynamic response function (HRF) [Boynton et al., 1996]. 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 haemodynamic response across the brain. Motion parameters, three translational and three rotational, totaling six, were further included in the GLM as covariates of no interest, in order to account for any signal variability due to the detected motion. Linear contrasts of parameter estimates (COPE) 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.

On the first level, each run was analyzed separately. The resulting COPE images were then entered into a higher level analysis, joining together groups of runs according to the study design shown in Fig. 2.1. In order to conduct higher level statistical analyses, all datasets were first 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 [Collins et al., 1994], 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 [Jenkinson and Smith, 2001; Jenkinson et al., 2002]).

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 (Beckmann et al., 2003;Woolrich et al., 2004).

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 all the pre-processing and post-processing options, which were combined in all possible ways, yielding a total of 16 combinations.

Preprocessing	Postprocessing
Spatial Smoothing FWHM = {5,8}	Cluster Thresholded Statistical Map (Z > 3,0)
Slice Time Correction = {1,0}	
Motion Parameters Inclusion = {1,0}	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 2.2. 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 2.3. 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:

$$D = 2 . \sqrt{(V1_x - V2_x)^2 + (V1_y - V2_y)^2 + (V1_z - V2_z)^2} \qquad (2)$$

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 factor of 2.

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 were then calculated as a measure of variability, according to the equation:

$$CV\ (\%) = \frac{\sigma(D_1, D_2)}{\mu(D_1, D_2)} \cdot 100\% \qquad (3)$$

where σ is the standard deviation and μ 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's 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.

$$\underset{Relative\ Change\ (\%)}{Mean\ (Z/COPE)} = \frac{|\mu(D_1) - \mu(D_2)|}{\mu(D_2)} \cdot 100\% \qquad (4)$$

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:

$$Detection\ Efficiency\ (\%) = \frac{ROI_{D_1}}{ROI_{D_1} \cup ROI_{D_2}} \cdot 100\% \qquad (5)$$

where $ROI_{D\,i}$ is the numer 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:

$$Cluster\ Overlap\ (\%) = \frac{ROI_{D_1} \cap ROI_{D_2}}{ROI_{D_1} \cup ROI_{D_2}} \cdot 100\% \qquad (6)$$

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	

Chapter 3

Results

In this chapter, the results obtained in our study will be presented. Firstly, the imaging results will be described, in terms of the cortical mapping of the visual field quadrants. Secondly, the multiple reproducibility measures determined for both within- and between-session comparisons, using two different dataset sizes, as well as a number of processing combinations, will be described. Finally, a summary of the optimal and average reproducibility measures found will be presented.

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 overlayed 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. An example of the representation of the right lower quadrant (Q1) of the visual field obtained using our quadrant mapping protocol is shown in Figure 3.1.

This example shows datasets A1 and B1, comprising 2 runs each (*datasets 11*), from one subject. We can clearly see an activated region in the left upper portion of the occipital region of the brain which corresponds to quadrant 1 (recall quadrant indexation) located in a symetrical opposite region of the visual field, according to its retinotopic properties. The figure represents the result of the same quadrant mapping, under the same acquisition and processing conditions but on different days. The first reproducibility assessment is visual and in this case we can say that apparently for both conditions the result is sufficiently reproducible. In Fig. 3.2, we can see an 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.

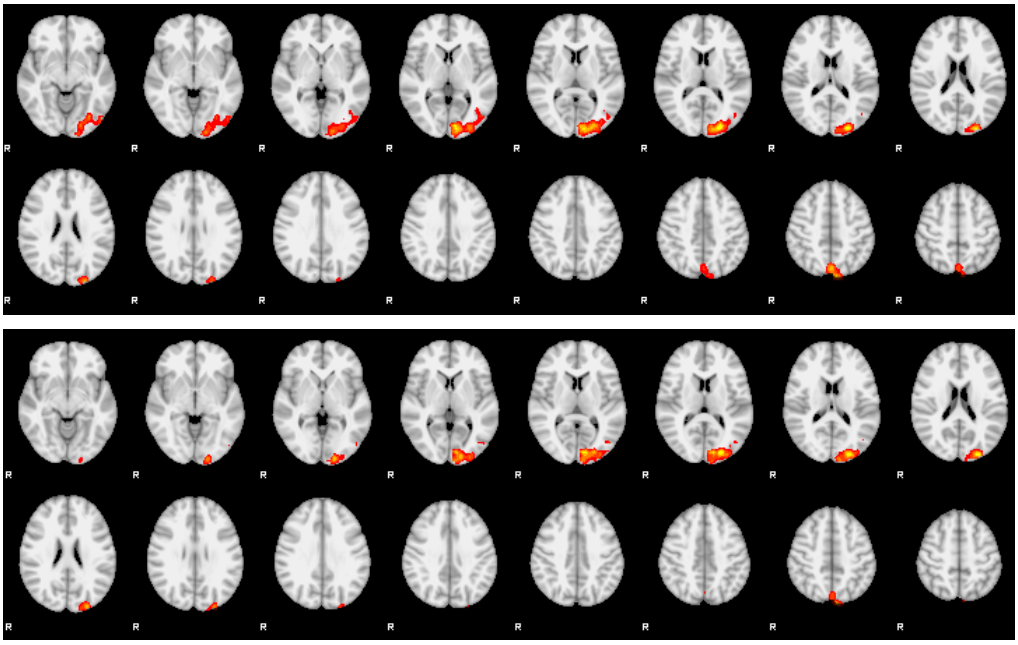

Figure 3.1 – Representative example of quadrant mapping: the Z statistic map representing the activation obtained upon stimulation of the right lower quadrant (Q1), overlaid on the template MNI anatomical image, for one subject, over sessions A and B (Upper and Lower panels, respectively). A stack of axials slices is shown in each case, with color representing Z score.

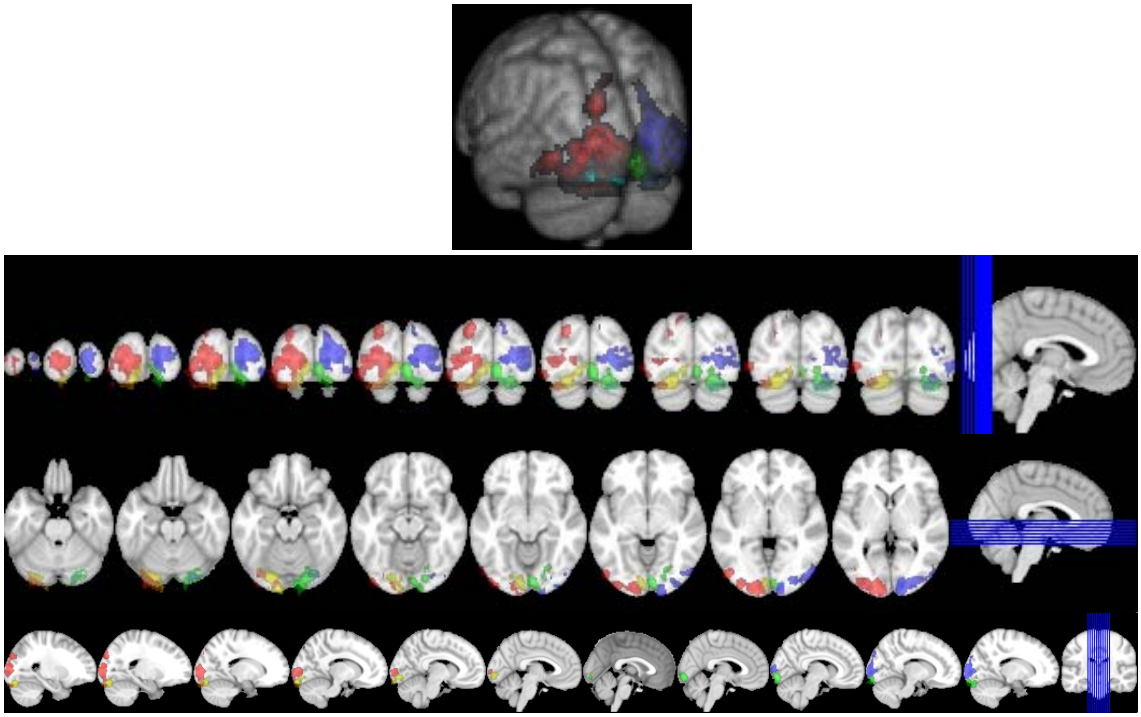

Figure 3.2 – Representative example of quadrant mapping: Representative example of the Z statistic map representing the activation obtained upon stimulation of the four quadrants (Q1 – Q4), overlaid on the template MNI anatomical image, for one subject. In *red* we have quadrant 1, *blue* represents 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 saggital slice images

In order to visually assess the reproducibility of quadrant mapping results, we compare them with homologous results (within- and between-session). This was done by overlaying the activated regions in each of the datasets. Figure 3.3 illustrates an example of an optimal intersection of the results of 3 datasets obtained and processed under the same conditions but acquired at distinct times.

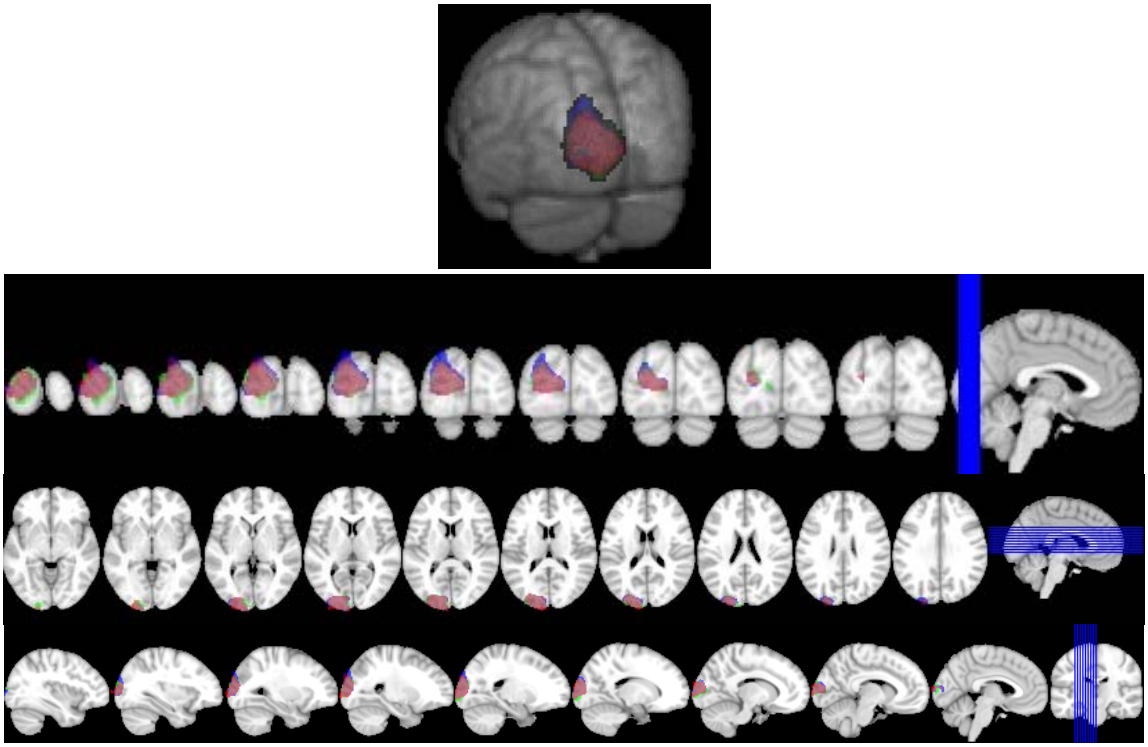

Figure 3.3 – Example of quadrant cluster mapping under the same conditions on different times. In *red* we have the actual dataset, in *blue* the homologous within-session dataset and in *green* the between-session dataset. On the top the 3 dimensional rendering of the image and on the bottom the corresponding coronal, axial and saggital slice images.

3.2. Reproducibility measurements

The results of the measures of activation in terms of localization, extent and intensity, as well as detection efficiency, generated a very extensive table of results. For *datasets 11* (size 2) this table had 2034 entries (the product of 4 Quadrants, 2 Sessions, 16 Parameter combinations. 2 Datasets per session, 4 subjects and 2 variables, Z score and COPE, which yields 2048, plus the same factors for 1 subject that only performed 1 session, producing 256 entries). For *datasets 22* (size 4) the numbers of entries reduces to a forth (~600 entries) has there is only 1 dataset per session and cluster thresholding was only performed using $Z > 5,0$, because, during the results analysis, this processing option seemed to give better results.

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. As we are representing the results as a function of the processing parameters and because these are very extensive in length we represented them on the axis as a code which is made of 3 digits. The 1st digit can be either 5 or 8 in accordance with the FWHM of the spatial smoothing gaussinan filter. Accordingly, the 2nd and 3rd digits can be either 0 or 1 depending on the inclusion of slice time correction (2nd digit) and on the inclusion of the motion parameters on the GLM (3rd digit).

The initial idea was to represent each of the series as an average and the respective standard deviation. However, we realized that the normal distribution assumption was not fulfilled in all cases. Through inspection of the series histograms, as the one shown in Figure 3.4, we corroborated this hypothesis: the distribution of the measures obtained typically had a behaviour where most of the values were located near zero and some outliers were present having high values. After putting aside the normal distribution, 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.

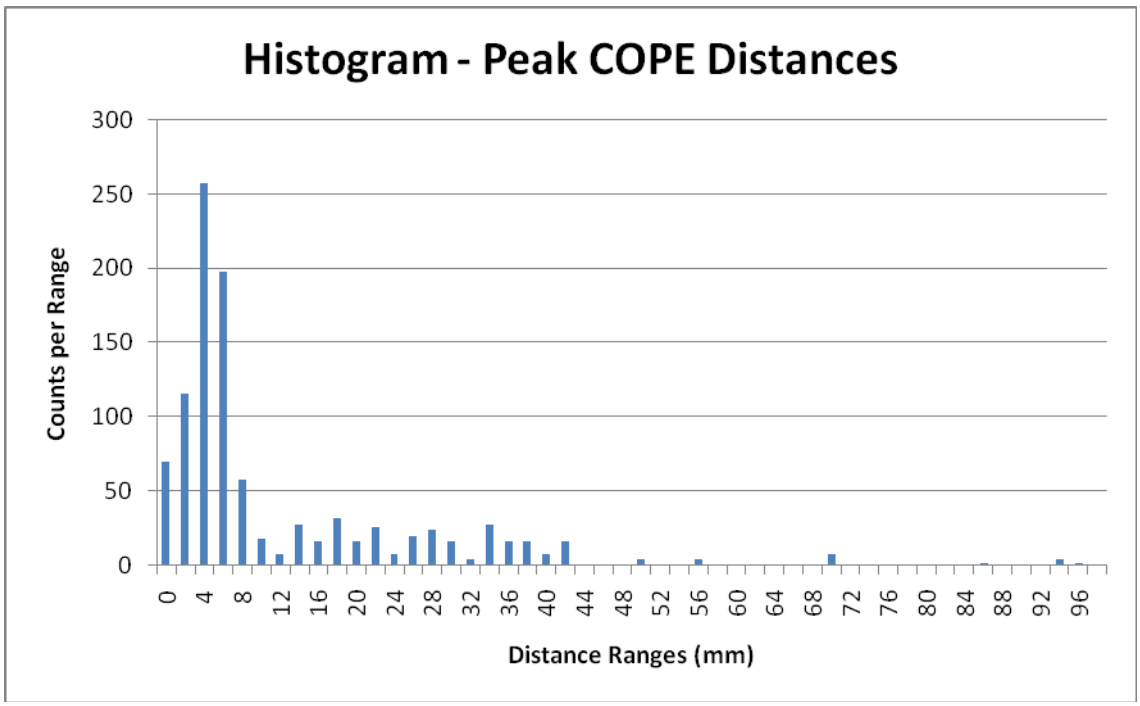

Figure 3.4 – Histogram showing the distribution of peak COPE distances obtained for dataset comparison BS11 (between-session datasets of size 2) across all subjects and quadrants.

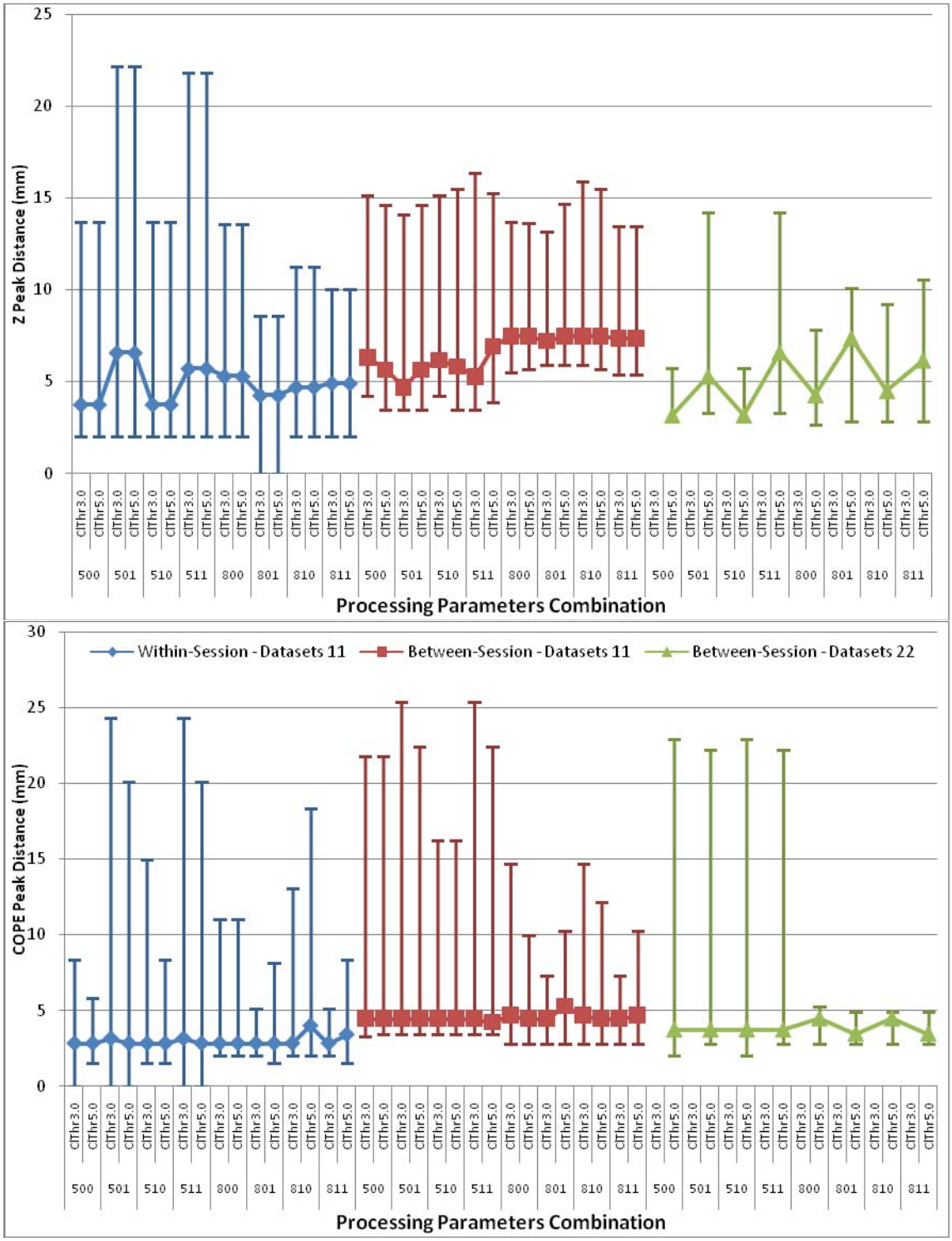

Figure 3.5 – Z Peak Distance (mm) (Top) and COPE Peak Distance (mm) (Bottom) - the median D values obtained for all subjects and quadrants is shown, for each type of comparison considered. Error bars represent the confidence intervals between the 25th and 75th percentiles.

From Figure 3.5, we can see that the median distance between activation peaks is around 5 mm, for both COPE and Z score maps, altough slightly higher for the statistical Z score maps. Despite this difference, the variability introduced by the percentile "error bars" is much higher for the COPE peak distance than

for the corresponding statistical map, as we reach a 25 mm distance on the first case (COPE) and get around 15 mm for most of the top limit values of the Z score maps. Comparing the 3 series, we could not find significant differences in variability between within- and between-session comparisons, nor between different dataset sizes. Furthermore, no significant differences between parameter combinations could be observed. However, there was a clear trend for the between-session distances of datasets of size 2 to be larger than the other two. This observation is consistent with better reproducibility of the results when using larger datasets, or when making within- relative to between-session comparisons.

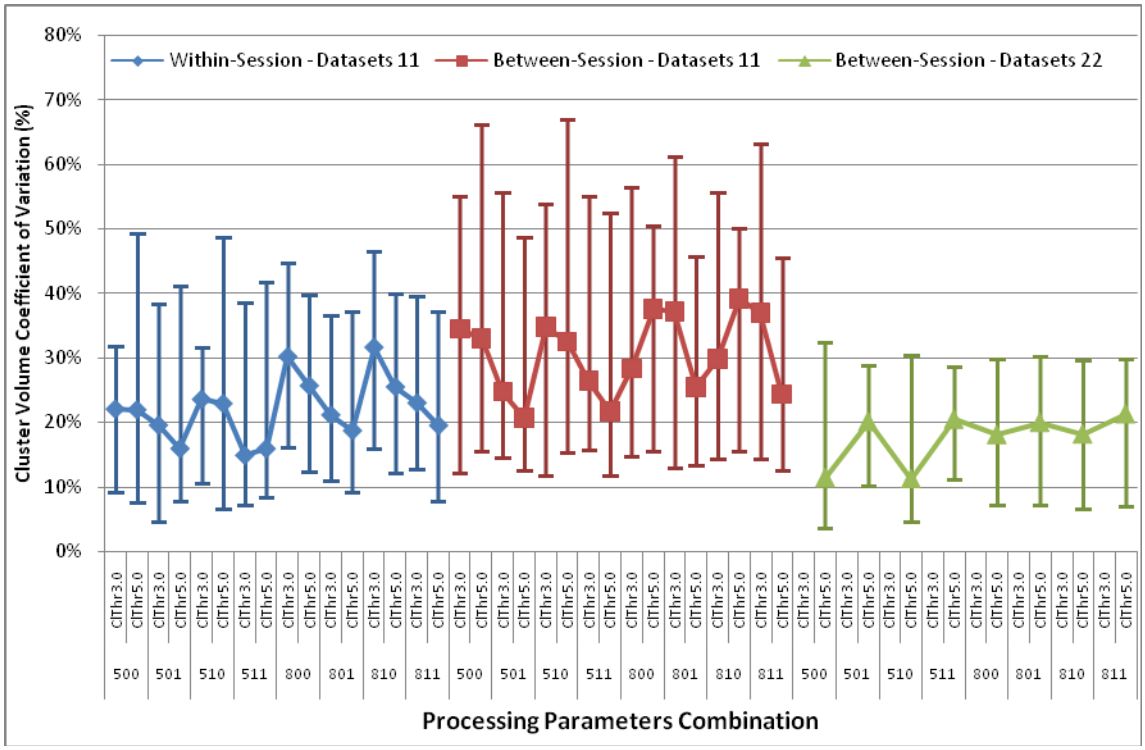

Figure 3.6 – Cluster Volume Coefficient of Variation (%) - the median overlap values obtained for all subjects and quadrants is shown, for each type of comparison considered. Error bars represent the confidence intervals between the 25th and 75th percentiles.

As shown in Figure 3.6, cluster volume CV clearly suffers from high values (reaching ~40%) and large variations as well, yielding no significant differences among the different combinations of processing parameters. Between-session comparisons of *datasets 22* (size 4) seem to have slightly better results in relation with the other 2 types of comparisons.

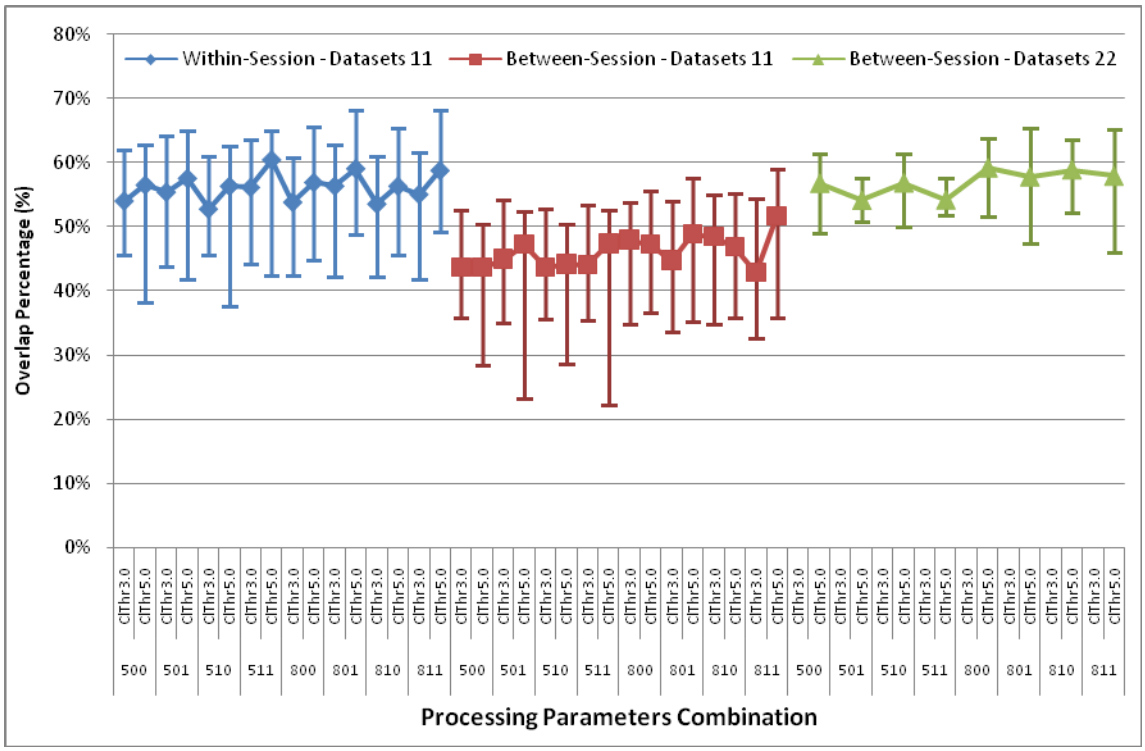

Figure 3.7 – Cluster Overlap Percentage (%) - the median overlap values obtained for all subjects and quadrants is shown, for each type of comparison considered. Error bars represent the confidence intervals between the 25th and 75th percentiles.

Assessing cluster overlap as shown in Figure 3.7, we can see that the series between-session for *datasets 11* stand out by having lower values and higher variability. In comparison to the between-session for *datasets 22* series, which has much less variable percentile "error bars", we can say with some certainty that both series represent statistically different results. The within-session series also presents high overlap values, but with some more variability. Again in this case, different processing parameter combinations do not yield significant overlap differences.

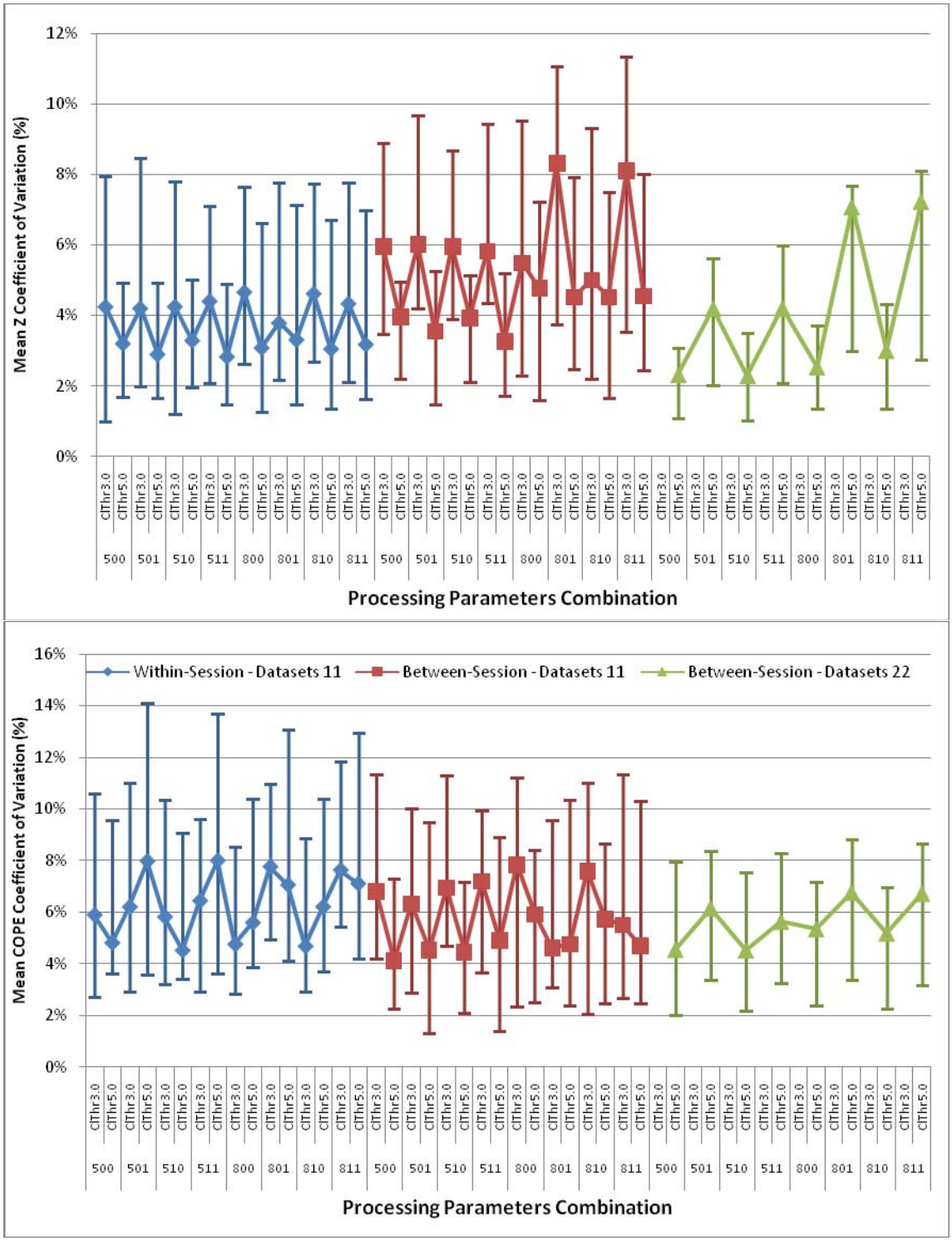

Figure 3.8 – Mean Z Coefficient of Variation (Top) and Mean COPE Coefficient of Variation (Bottom) - the median overlap values obtained for all subjects and quadrants is shown, for each type of comparison considered. Error bars represent the confidence intervals between the 25th and 75th percentiles.

Once again, Figure 3.8 shows that mean COPE CV yields slightly more variability than the corresponding statistical Z-score map, particularly for the within-session series. As expected, for the Z score map, the mean value CV is higher and more variable for the lower cluster thresholding option (for the *datasets 11*

series) but such behaviour is not seen on the mean COPE CV, in particular for the within-session series. Furthermore for the mean Z CV between-session series, the values appear to have a slightly higher variability average (CV), while the opposite is seen on mean COPE CV average (higher variability for the within-session series). No significant differences exist between processing parameters combinations.

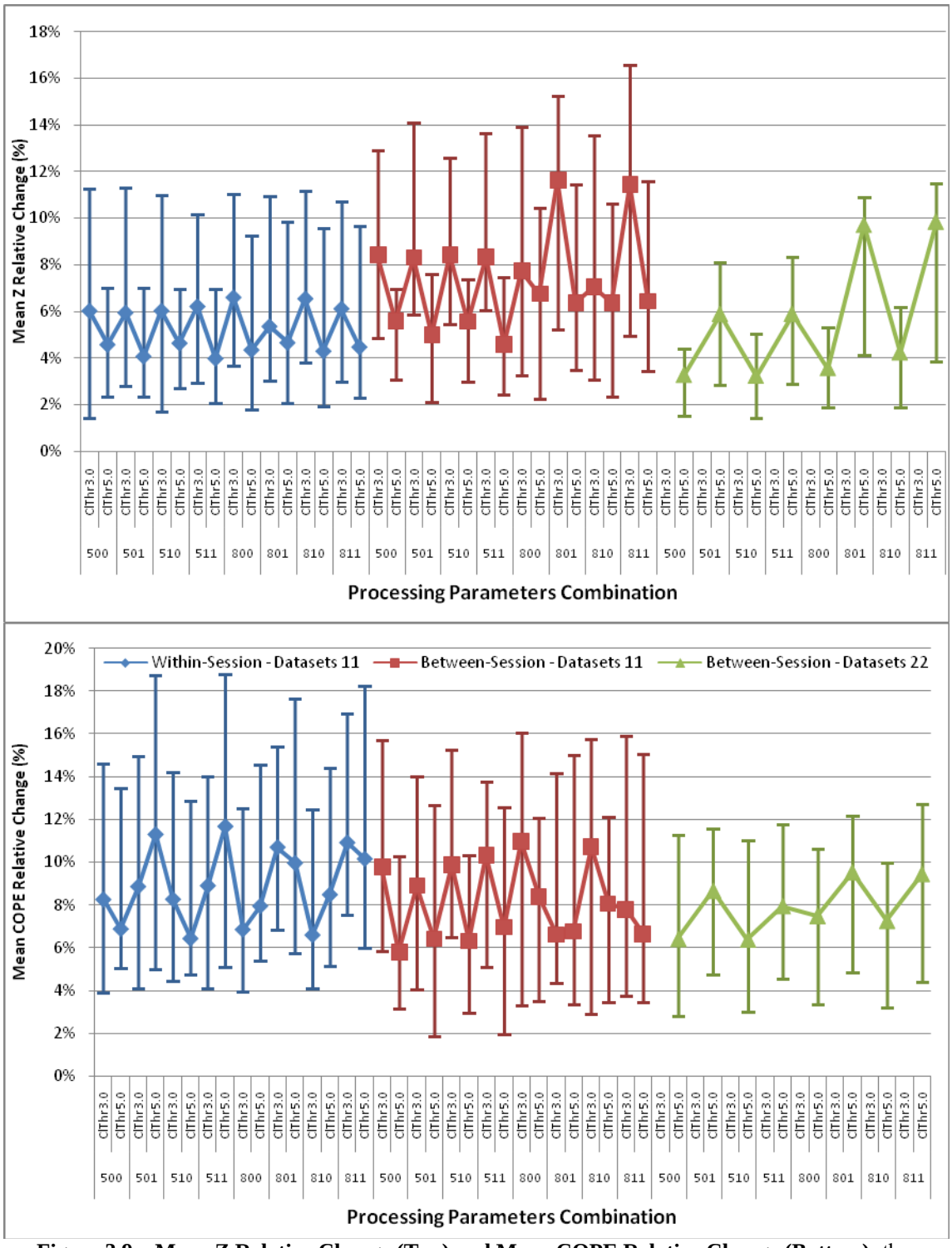

Figure 3.9 – Mean Z Relative Change (Top) and Mean COPE Relative Change (Bottom)- the median overlap values obtained for all subjects and quadrants is shown, for each type of comparison considered. Error bars represent the confidence intervals between the 25th and 75th percentiles.

If we compare the charts plotting mean Z/COPE CV and mean Z/COPE relative change, in Figure 3.9, we can see that the results are essentially the same although on a different scale.This means that although being different these 2 measures are directly interdependent. Again, mean COPE relative change is affected by more variability, in comparison to mean Z relative change and no significant differences exist between processing parameters combinations.

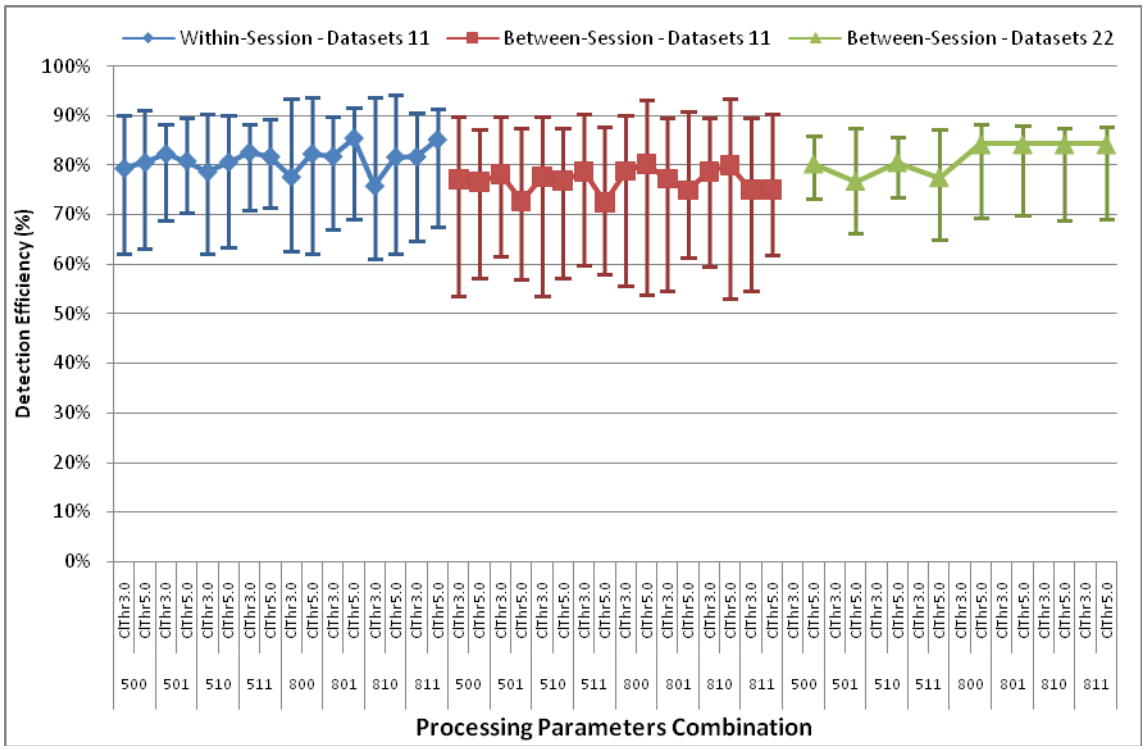

Figure 3.10 – Detection Efficiency (%) – The median overlap values obtained for all subjects and quadrants is shown, for each type of comparison considered. Error bars represent the confidence intervals between the 25th and 75th percentiles.

As far as detection efficiency is concerned, Figure 3.10 shows that among the 3 series no significant differences seem to exist, although between-session comparisons for *datasets 11* (size 2) denote a slightly lower detection efficiency as well as affected by more variability. Within-session comparisons appear to have better results and can be somewhat distinguishable from the between-session comparisons (size 2). *Datasets 22* (size 4) apparently introduce lower variability as well as a slight higher result when comparing with between-session (size 2) series. Once again no conclusive differences are found between the processing parameters combinations.

3.3. Summary of reproducibility measurements

After assesing in detail the results of each reproducibility measure, it is now important to summarize the information in a simpler way. We considered 2 approaches to sum up the results.

First, since no significant differences were found among the different combinations of processing parameters, we chose to display the best choice of parameters combination in terms of the respective reproducibility results. For each measure and series, the median value of each parameter combination was considered and given a rank. The best choice of parameters was selected as the one with the lowest rank and corresponded to: pre-processing using 5 mm FWHM spatial smoothing, no slice time correction and no inclusion of motion parameters in the GLM (500); and post-processing using cluster thresholding with $Z>5.0$. We represent, for each measure, the median values obtained (for this combination) for within-session and between-session comparisons (*datasets 11* and *datasets 22*). The reproducibility results for the optimal choice of parameter combinations are shown in Figures 3.11 to Figure 3.17. The second approach consisted on computing the average and standard deviation of all the median values of each processing parameters combination, for each measure. These results are shown in Figures 3.18 to Figure 3.22.

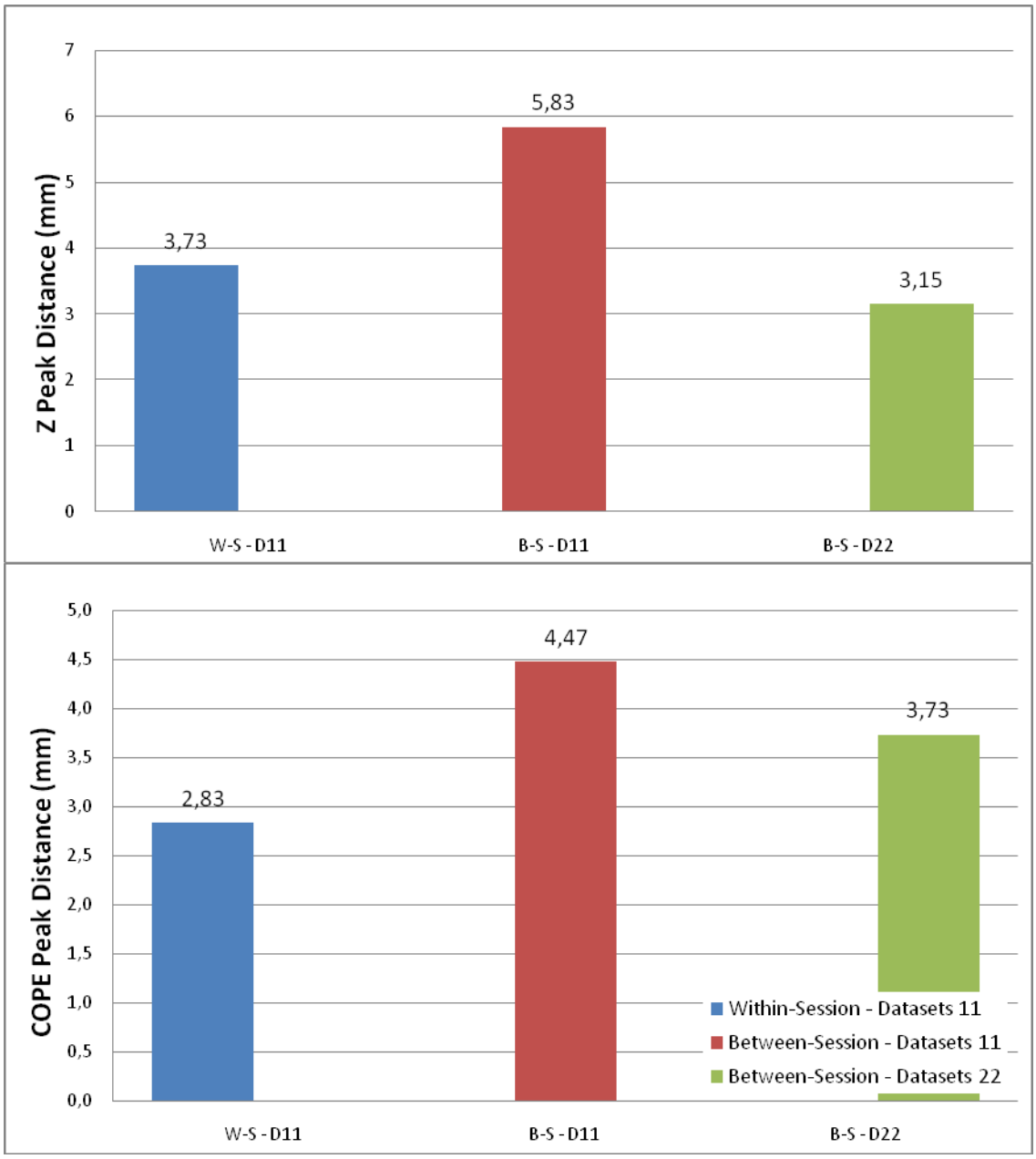

Figure 3.11 –Reproducibility measurements of localization obtained with the optimal combination of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

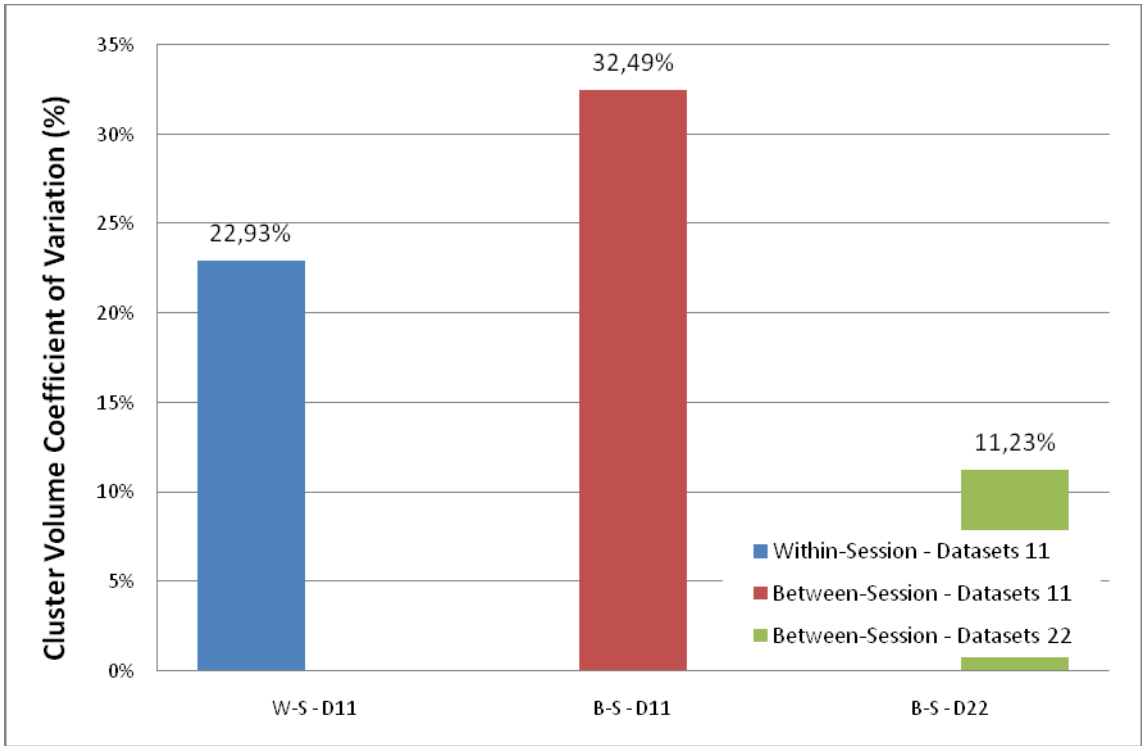

Figure 3.12 –Reproducibility measurements of activation extent obtained with the optimal combination of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

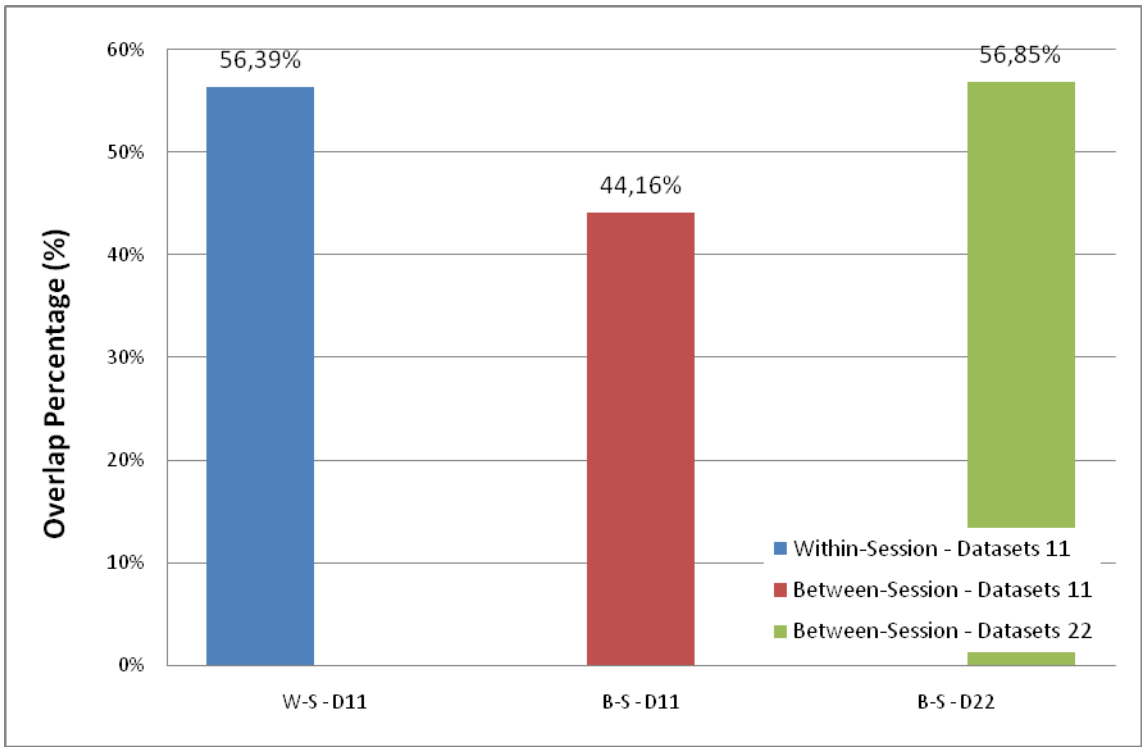

Figure 3.13 –Reproducibility measurements of activation overlap obtained with the optimal combination of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

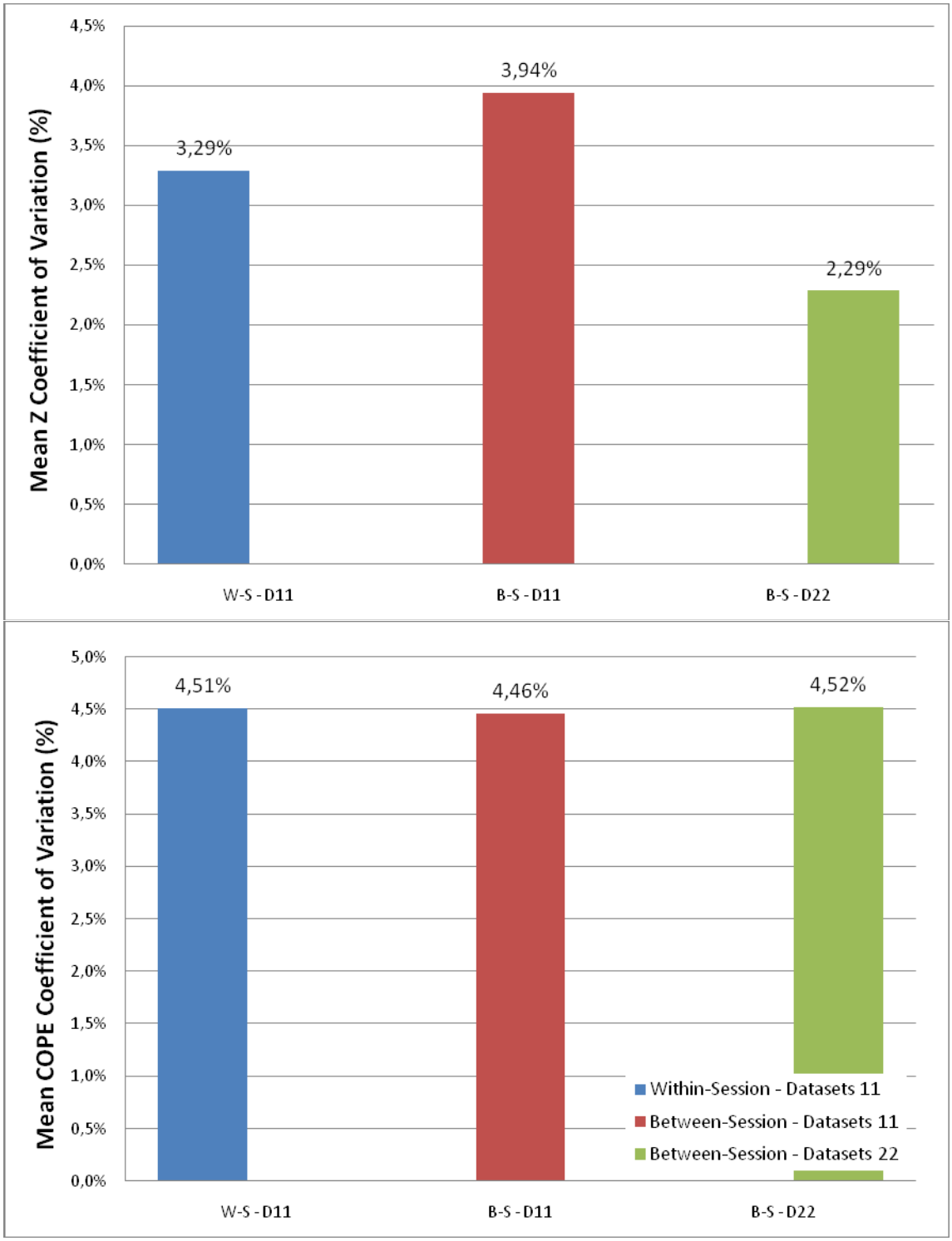

Figure 3.14 –Reproducibility measurements of activation intensity obtained with the optimal combination of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

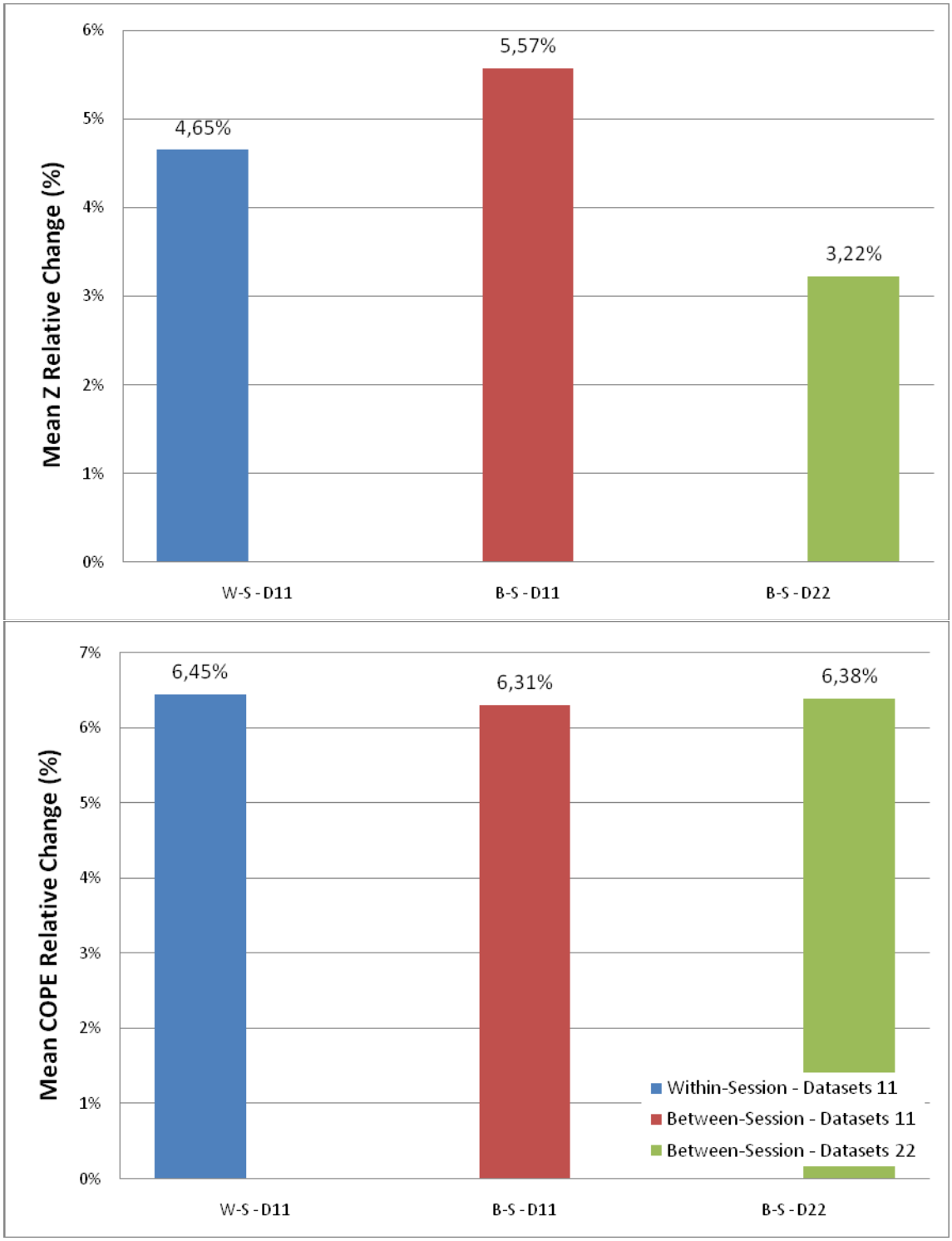

Figure 3.15 –Reproducibility measurements of activation intensity obtained with the optimal combination of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

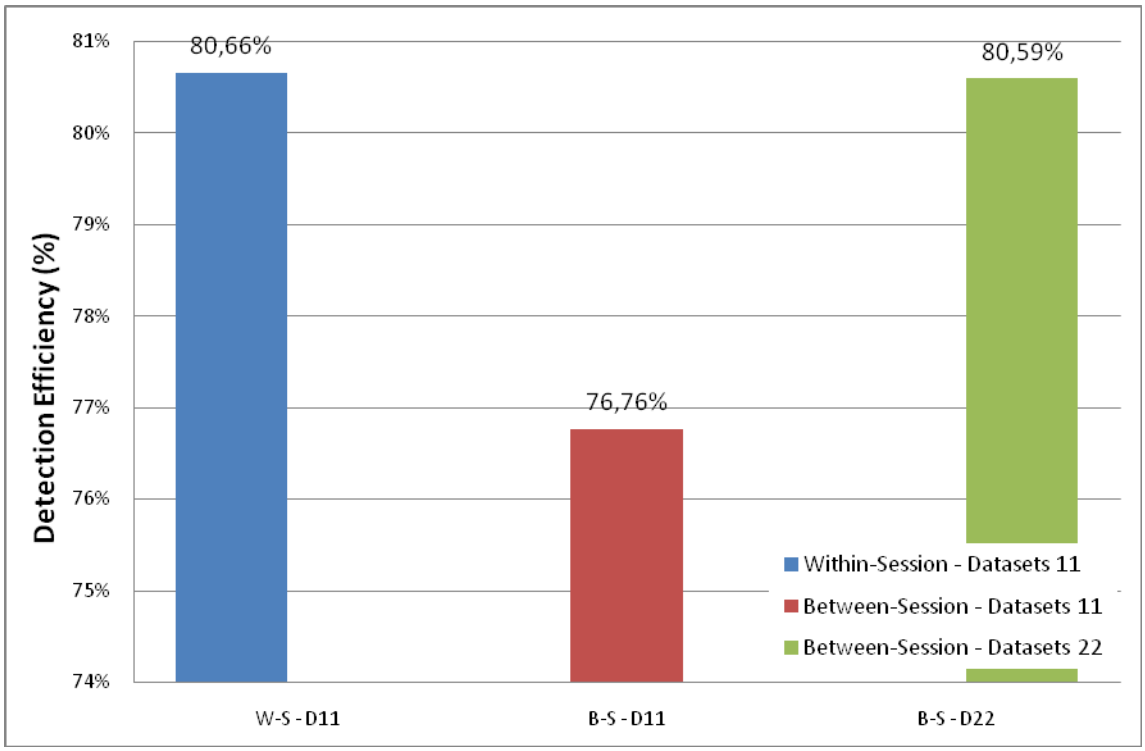

Figure 3.16 –Reproducibility measurements of detection efficiency obtained with the optimal combination of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

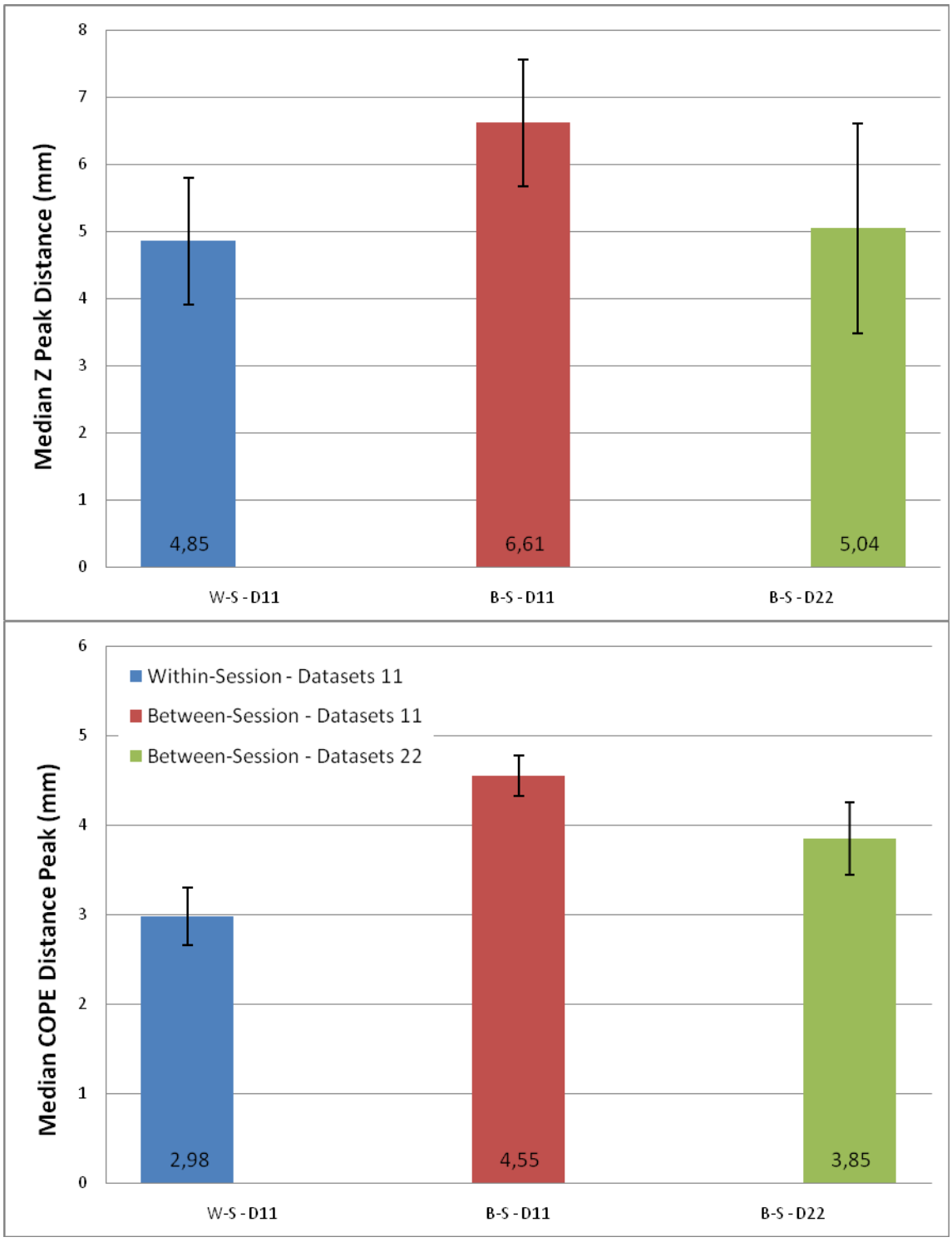

Figure 3.17 –Average reproducibility measurements of localization obtained with all combinations of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

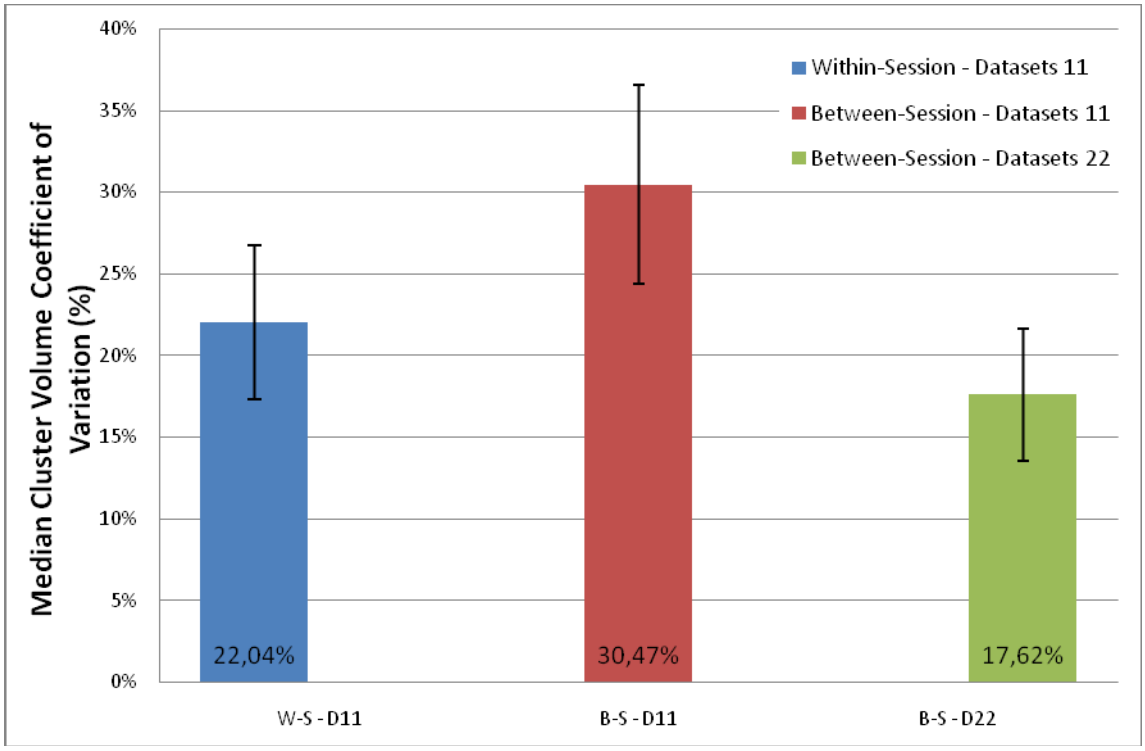

Figure 3.18 –Average reproducibility measurements of activation extent obtained with all combinations of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

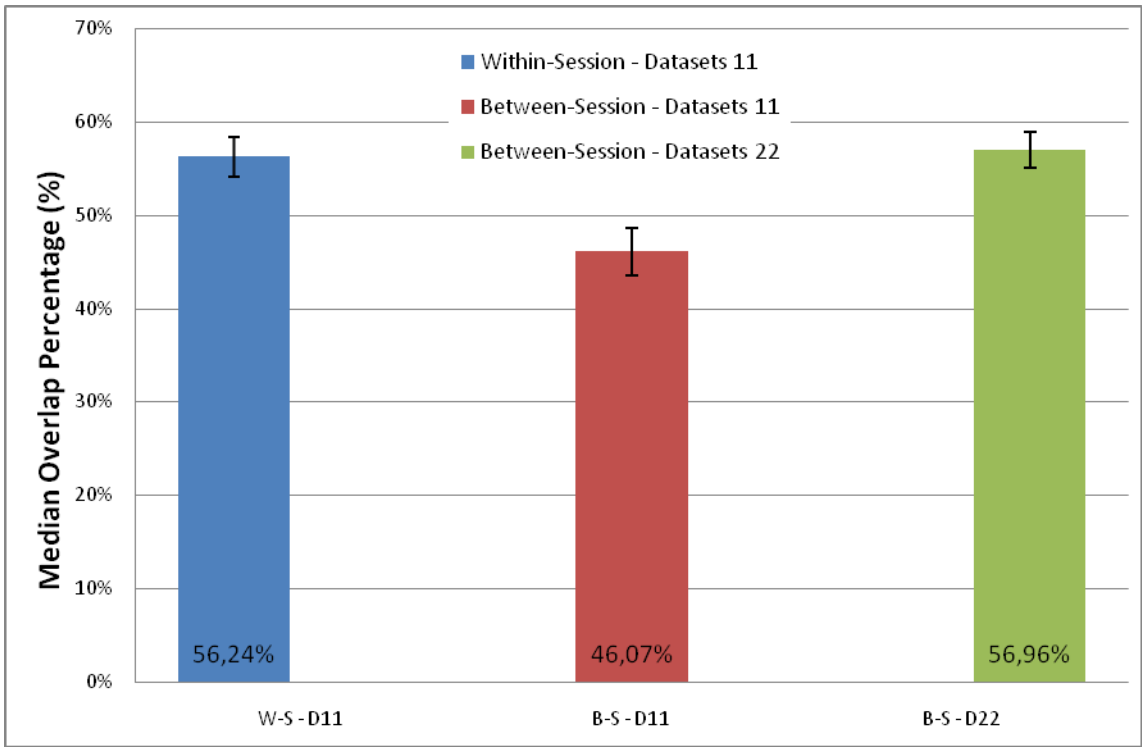

Figure 3.19 –Average reproducibility measurements of activation overlap obtained with all combinations of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

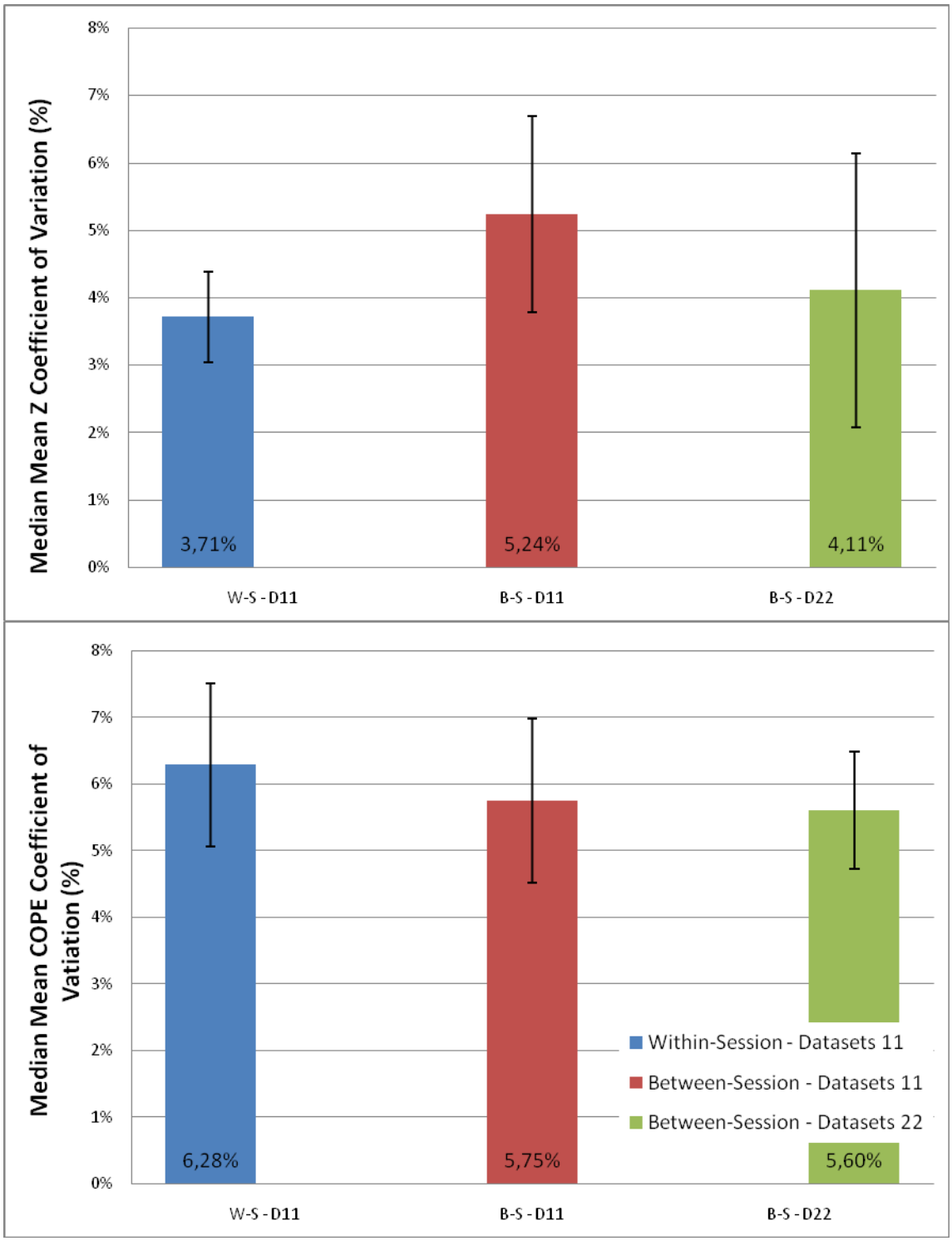

Figure 3.20 –Average reproducibility measurements of activation intensity obtained with all combinations of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

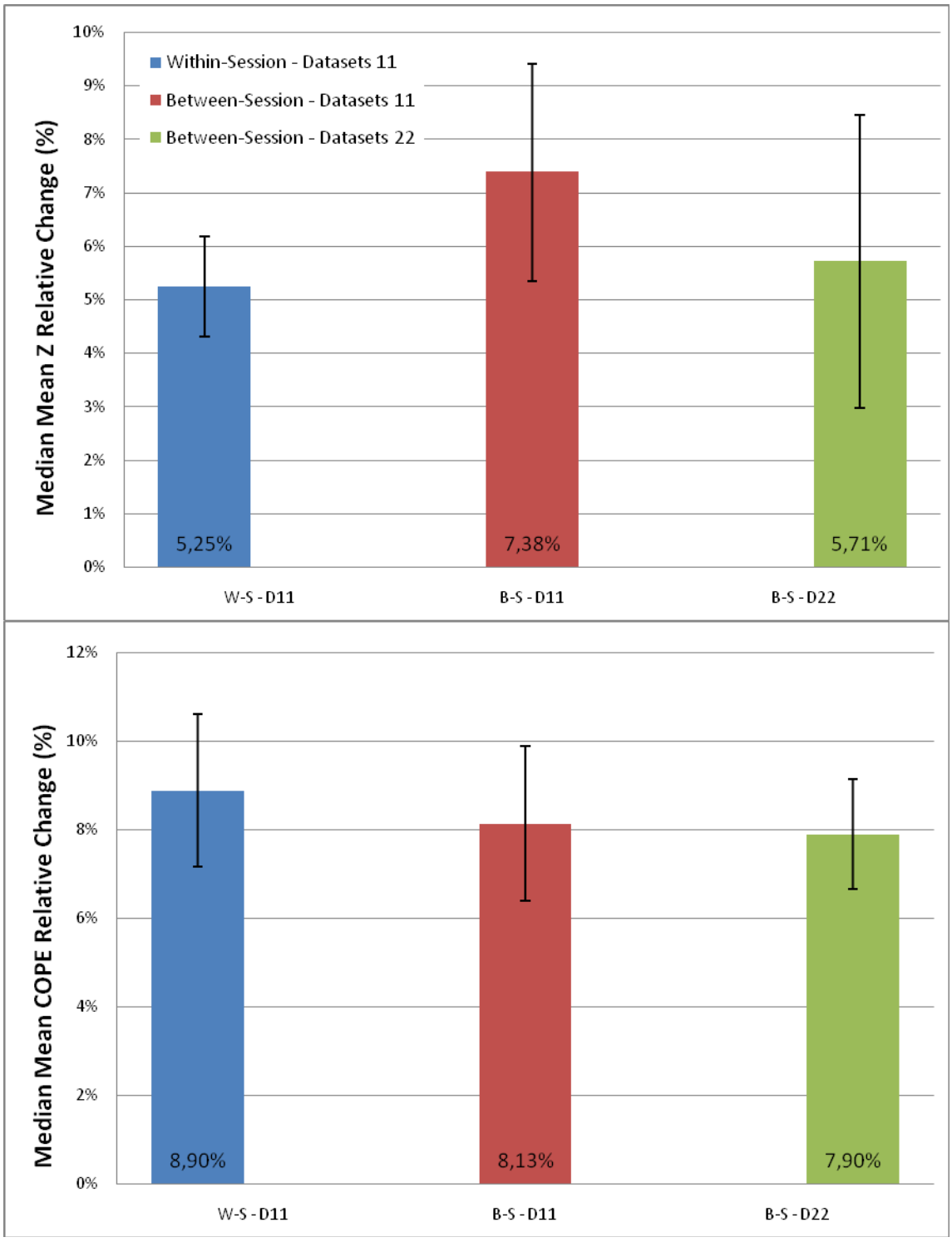

Figure 3.21 –Average reproducibility measurements of activation intensity obtained with all combinations of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

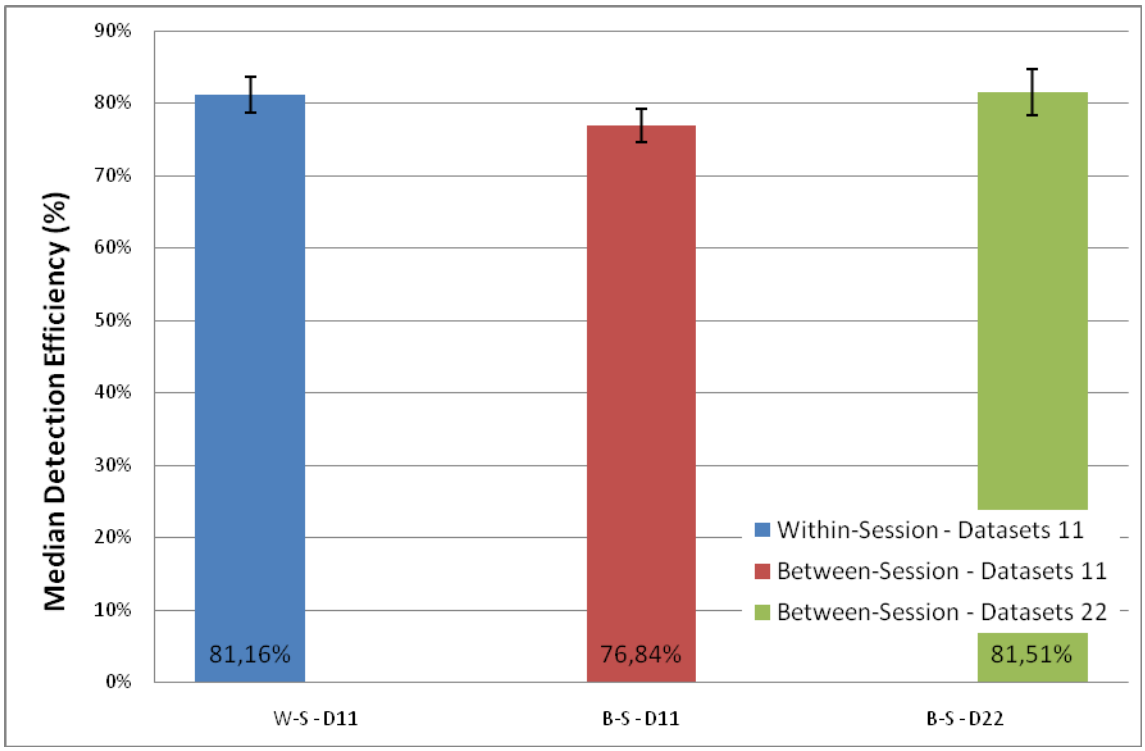

Figure 3.22 –Average reproducibility measurements of activation efficiency obtained with all combinations of pre- and post-processing parameters, for each of the three types of comparisons considered: within-session (datasets 11), between-session (datasets 11) and between-session (datasets 22).

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, and partially seen on the initial charts, 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 whithin-session comparisons (*datasets 11*) yield worst results than the other 2 comparison types.

Chapter 4

Conclusions and Future Work

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) (Peelen & Downing, 2005). 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 (Wagner et al., 2005). 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 (Tjandra, 2005).

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 (Laurienti, 2002). 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.

Bibliography

Bartsch, A. J., Homola, G., Biller, A., Solymosi, L., & Bendszuz, M. (2006). Diagnostic Functional MRI: Illustrated Clinical Applications and Decision-Making. *Journal of Magnetic Resonance Imaging* , 921-932.

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

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

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

Buxton, R. B. (1998). Dynamics of blood flow and oxygenation changes during brain activation. *Magn. Reson. Med. 39* , 855-864.

Engel, S. et al. (1994). fMRI of human visual cortex. *Nature 369* .

Friston, K. J. (1994). Analysis of functional MRI time-series. *Human Brain Mapping, 1* , 153-171.

Geoffrey M. Boynton, S. A. (July 1996). Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. *The Journal of Neuroscience, Vol. 16, No. 13* , 4207-4221.

Heeger, J. R. (2002). What does fMRI tell us about neuronal activity? *Nature, Vol.3* , 142-151.

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

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

Kwong, K. K. (1992). Dynamic magnetic resonance of human brain activity during primary sensory stimulation. *Proc. Natl Acad. Sci USA 87* , 5675-5679.

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

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

Nikos K. Logothetis, J. P. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature 412* , 150-157.

Ogawa, S. et al. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl Acad. Sci USA, 89* , 5951-5955.

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

S. Ogawa, T. M. (December 1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA, Vol. 87* , 9868-9872.

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

SM, S. (2002). Fast robust automated brain extraction. *Human Brain Mapping 17* , 143-155.

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

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.

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

Vanzetta, I. & Grinvald, A. (1999). Increased cortical oxidative metabolism due to sensory stimulation: implications for functional brain imaging. *Science 286* (286), 1555-1558.

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

Warnking, J. et al. (2002). fMRI Retinotopic Mapping - Step by step. *NeuroImage 17* , 1665-1683.

Woolrich MW, et al. (2001). Temporal autocorrelation in univariate linear modelling for fMRI data. *Neuroimage 14* , 1370-1386.

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