Doctoral Dissertation
Doctoral Program in Electronic and Telecommunications (28th Cycle)

Quality data assessment and improvement in pre-processing pipeline to minimize impact of spurious signals in functional magnetic imaging (fMRI)

By

Anna Nigri

******

Supervisors:
Ing. Filippo Molinari, Supervisor
Dr. Maria Grazia Bruzzone, Co-Supervisor

Politecnico di Torino
2017

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Anna Nigri

2017

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Acknowledges

Grazie

alla mia famiglia,

ai miei colleghi e amici,

a tutte le persone che mi hanno

supportato e "sopportato" in questi anni.

Abstract

In the recent years, the field of quality data assessment and signal denoising in functional magnetic resonance imaging (fMRI) is rapidly evolving and the identification and reduction of spurious signal with pre-processing pipeline is one of the most discussed topic. In particular, subject motion or physiological signals, such as respiratory or/and cardiac pulsatility, were showed to introduce false-positive activations in subsequent statistical analyses.

Different measures for the evaluation of the impact of motion related artefacts, such as frame-wise displacement and root mean square of movement parameters, and the reduction of these artefacts with different approaches, such as linear regression of nuisance signals and scrubbing or censoring procedure, were introduced. However, we identify two main drawbacks: i) the different measures used for the evaluation of motion artefacts were based on user-dependent thresholds, and ii) each study described and applied their own pre-processing pipeline. Few studies analysed the effect of these different pipelines on subsequent analyses methods in task-based fMRI.

The first aim of the study is to obtain a tool for motion fMRI data assessment, based on auto-calibrated procedures, to detect outlier subjects and outliers volumes, targeted on each investigated sample to ensure homogeneity of data for motion.

The second aim is to compare the impact of different pre-processing pipelines on task-based fMRI using GLM based on recent advances in resting state fMRI preprocessing pipelines. Different output measures based on signal variability and task strength were used for the assessment.

Keywords: fMRI, motion, preprocessing, pipeline

Contents

Introduction

1.1 Original Contributions

Most of the work presented in this thesis is an original contribution to knowledge and has been published, submitted or part of future works:

- the auto-calibrated method for the detection of outlier subjects and volumes in fMRI data was currently submitted to an international Journal (Chapter 4).
- the effect of pre-processing pipeline on GLM in task-based fMRI (Chapter 5) was presented as poster at 2015 OHBM national conference.
- the olfactory and auditory fMRI studies on disorder of consciousness (Chapter 6) were presented as poster at 2013 and 2015 OHBM national conference, respectively and were published in 2015 and 2016 as Original Research on International Journals (European Journal of Neurology and Brain Imaging and Behaviour, respectively).

The studies were conducted at Neuroradiology Department, IRCCS, Foundation Neurological Institute "C. Besta".

1.2 Thesis Organization

The thesis is organized in the following sections.

Chapter 1 starts with an introduction to MRI and fMRI principles. We detailed how MRI images were obtained combining a static magnetic field, a radiofrequency and three localization gradients. Subsequently, the mechanism of Blood Oxygen level dependent contrast at the basis of fMRI signal was explained. At the end of the chapter we focused on possible noise source in fMRI data.

Chapter 2 introduce the unsupervised learning methods, which were used in the Chapter 4 to detect motion outlier subjects and volumes in fMRI studies with the auto-calibrated fMRI data assessment for motion.

A description of the fMRI pre-processing approaches was presented in Chapter 3. In particular, the standard and advanced strategies to reduce the sources of noise in fMRI data previously described in Chapter 1, were outlined. Moreover, a brief description of statistical analysis applied to fMRI data was detailed: theory assumptions and implementation of general linear model (GLM).

Chapter 4 explains the algorithm implementations of the auto-calibrated methods for the detection of outlier subjects and outlier volumes in three task-based fMRI datasets of healthy participants,

characterized by high and low movements. At the end of the chapter, the results were presented and discussed in the framework of previous studies.

In Chapter 5, the impact of different pre-processing pipelines was assessed on task-based fMRI using GLM. For these purpose, a sample of healthy participants characterized by low motion was used. The order of step and the different signals included as sources of noise in pre-processing pipelines were identified on the basis of the literature reported in Chapter 3. Different output measures based on signal variability and task strength were used for the assessment.

In Chapter 6, two task-based fMRI studies on patients with disorder of consciousness were presented as other applications of pre-processing pipeline and optimization of subsequent analyses for patients characterized by high motion.

Finally, a conclusion with discussion of some limitations and possible future extensions of this work.

List of abbreviations

Amplitude of Low Frequency Fluctuations (ALFF)

Anterior Commissure - Posterior Commissure (AC-PC)

Associatively related trials (WWr)

Auditory Cortex (AC)

Blood Oxygenation Level-Dependent (BOLD)

Brainstem Auditory Evoked Potentials (BAEPs)

Cerebral Blood Flow (CBF)

Cerebral Blood Volume (CBV)

Cerebral Metabolic Rate of oxigen (CMRO2)

CerebroSpinal Fluid (CSF)

CoPhenetic Correlation Coefficient (CPCC)

Davies-Bouldin index (DB)

Default Mode Network (DMN)

deoxyHemoglobin (dHb)

Derivative of rms VARiance over voxelS (DVARS)

Disorder Of Consciousness (DOC)

Echo Planar Imaging (EPI)

Echo Time (TE)

Emerged from the Minimally Conscious State (EMCS)

Event-Related Potential (ERP)

False Discovery Rate (FDR)

Family-Wise Error (FWE)

Field Of View (FOV)

First derivative of realignment parameters (RP d/dt)

Fractional ALFF (f/ALFF)

Framewise Displacement (FD)

Free Induction Decay (FID)

Frequency encoding Gradient (GF)

functional Magnetic Resonance Imaging (fMRI)

Fusiform Face Area (FFA)

General Linear Model (GLM)

generalized Extreme Studentized Deviate (ESD) test

Hamming Distance (HD)

Hemodynamic Response Function (HRF)

Hierarchical Clustering (HCl)

Independent Component Analysis (ICA)

Independent Component maps (ICs)

K-Means clustering (KM)

Late Positive Complex (LPC)

lateral-medial Orbitofrontal Cortex (lOFC)

Local Field Potentials (LFPs)

Locked-In Syndrome (LIS)

Low-Level Auditory contrast (LLA)

Magnetic Resonance Imaging (MRI)

Mahalanobis Distance (MD)

Minimally Conscious State (MCS)

Multi-Unit Activity (MUA)

Nuclear Magnetic Resonance (NMR)

Number of Phase encoding steps (NP)

Object Face Area (OFA)

Orbitofrontal Cortex (OFC)

oxygenated Hemoglobine (Hb)

Parahyppocampal Place Area (PPA)

Phase encoding Gradient (GP)

Physiological Parameter regression (PhyP)

Piriform Cortex and AMYgdala cortex (PC/AMY)

Pseudowords trials (PP)

RadioFrequency (RF)

Realignment Parameters and Physiological Parameters regression (RPPhyP)

Realignment Parameter (RP)

Region Of Interest (ROI)

Repetition Time (TR)

Root Mean Square (RMS)

Signal to Noise Ratio (SNR)

Silhouette index (Sh)

Slice selection Gradient (GS)

Slice-Time Correction (STC)

Standard Deviation (SD)

Standardized Uptake Value (SUV)

Statistical Parametric Mapping (SPM8)

Sum of Squared Error (SSE)

Superior Temporal Gyrus (STG)

temporal SNR (tSNR)

Traumatic (TBI), hemorrhagic (HBI) and anoxic (ABI) Brain Injury

Unrelated trials (WWur)

Vegetative State/Unresponsive Wakefulness Syndrome (VS/ UWS)

Verbal Fluency (VF)

White Matter (WM)

Word-pseudoword trials (WP)

Chapter 1

Theoretical background: functional magnetic resonance imaging (fMRI)

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique based on a physical phenomenon, defined Nuclear Magnetic Resonance (NMR) principle, in which magnetic nuclei within a magnetic field absorb and re-emit electromagnetic radiation. This principle is used in MRI scanner to provide images with different contrast among tissues based on their different chemical and physical properties.

1.1 MRI scanner

A MRI scanner is composed from three key components:

- a set of main magnet coils distributed in a bore to generate the static magnetic field,
- three gradient coils for signal localization,
- a radiofrequency transmitter and/or receiver coil to perturb the system and recorder the generated signal. Coils are application-specific.

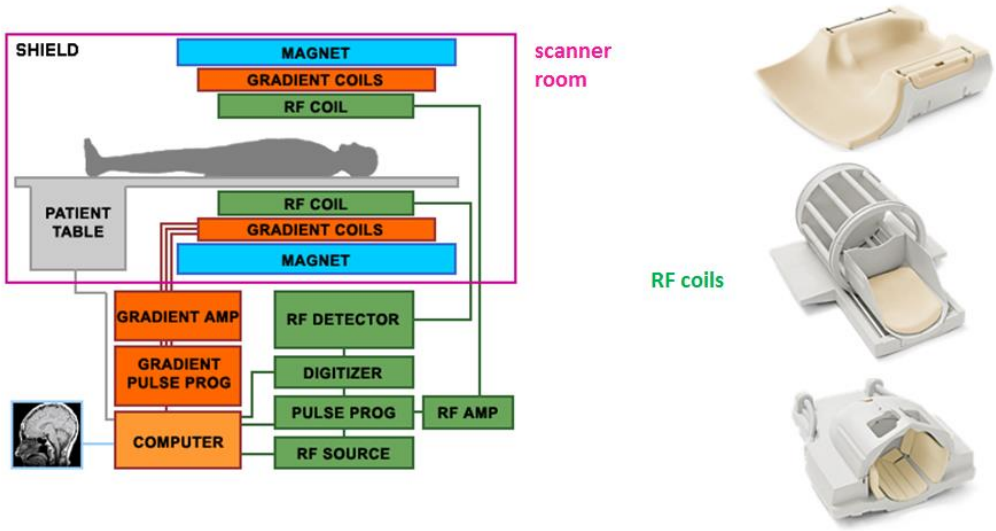


Figure 1.1 MRI scanner structure and coils (spinal cord, head 8-channel, head 32-channel; Philips).

1.2 Physical principles of MRI

The principle of NMR is that certain atomic nuclei, atomic nuclei with a non-zero spin (nuclei with an unpaired protons or neutrons, have intrinsic magnetic properties and these properties can be manipulated when placed in a magnetic field.

The attention is restricted to hydrogen nuclei ^{1}H with only one unpaired proton, because it constitutes approximately 63% of atoms in human body: living tissue consists mostly of water and fat, which both are rich in hydrogen.

Atomic nuclei with a non-zero spin possess spin angular momentum, as defined by Pauli in 1924, and a magnetic moment μ, characterizing the magnetic field around the nucleus (Fig. 1.2).

Without an external magnetic field, the magnetic moment of the nuclei are free to point in any directions. Since all directions are equally likely, the sum of all magnetization vectors of all nuclei magnetic moments is very close to zero (Fig. 1.2).

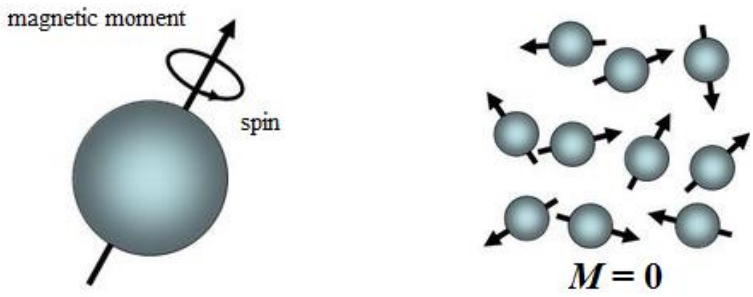


Figure 1.2 On the left: magnetic moment of a nucleus; on the right: orientation of magnetic moments without external magnetic field.

1.2.1 Effects of external static magnetic field

When an external static magnetic field B_0 is applied along a z-axis, it produces two different effects on magnetic moment of nuclei: the random magnetic moments of nuclei tend to align along the magnetic field direction and to follow a precession movement. The first effect is explained by quantum mechanism, which assigns a quantum number I to the spin (I = ½ for proton of hydrogen nucleus.

The external static magnetic field creates two different states ($\mu \pm$ ½) for these protons: the magnetic moment has a component parallel to B_0 in the state $\mu =$ ½, and anti-parallel to B_0 in the state $\mu = -$½. The two states for the nucleus are equally probable, because their energy difference

are small, but population of parallel state is usually higher that population of anti-parallel, due to low energy state. However, there is a dynamic exchange between these two populations of protons.

The second effect, is explained by the torque force produced by the magnetic field on each magnetic moment. As results, magnetic moments precesse around the axis of the magnetic field at a special frequency referred as Larmor frequency ω_L(in MHz):

$$\omega_L = \gamma B_0$$

where γ is the gyromagnetic ratio, intrinsic characteristic of the nuclear isotope, and B_0 is the strength of static magnetic field. For ^{1}H, γ = 42.58 MHz/T. The frequency of the precession is proportional to the strength of the magnetic field B_0. For example at 3 Tesla field, ω_L= 127.7 MHz.

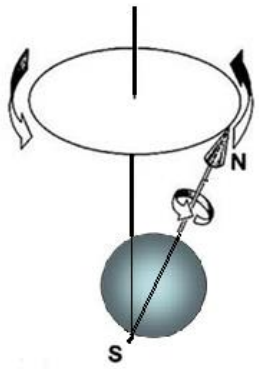


Figure 1.3 Precession of the magnetic moment of the proton around the external magnetic field axis.

A parallel and antiparallel hydrogen nuclei have equal, but opposite magnetic moments and cancel each other out. However, there are always slightly more hydrogen nuclei parallel to B_0, that give origin to net magnetization *M*. The individual magnetic moments create a surface of a double cone, and their joint alignment gives origin to the net magnetization *M*. The phase of an ensemble of magnetic moments are random. The net magnetization is the vector sum of all the individual magnetic moments, as follow:

$$M = \sum p_i \ \mu_i$$

where μ_i is the magnetic moment of the i-th state and p_i is its population, which follows Boltzman statistics. Between the two states exist a dynamic balance and the nuclei are in thermal equilibrium, so the resulting magnetization is called the equilibrium magnetization M_0, which can be divided into two components, *Mz* and *Mxy*.

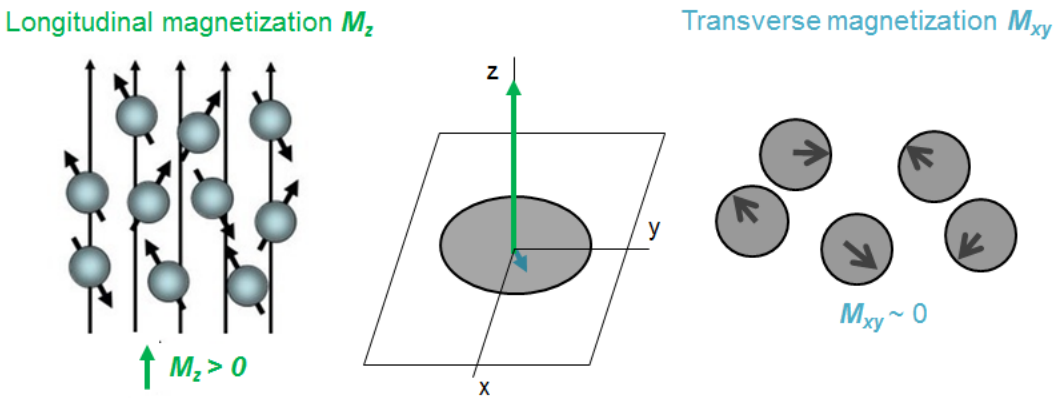


Figure 1.4 Orientation of magnetic moments with external magnetic field.

1.2.2 Radiofrequency

MR signal can be detected if time dependent transverse magnetization (perpendicular to B_0) is applied, because it induces a current in a receiver coil according to Faraday's law of induction. Indeed, without an application of a radiofrequency (RF), all the proton process at the same frequency, but with a random phase and a net magnetization constant and parallel to B_0.

When an external RF of amplitude B_1, with the same frequency of precession of spins, and perpendicular to B_0 is applied:

- the net magnetization M_0 from longitudinal axis is converted in transversal axis (90°) and
- all the spinning dipoles are exactly in phase.

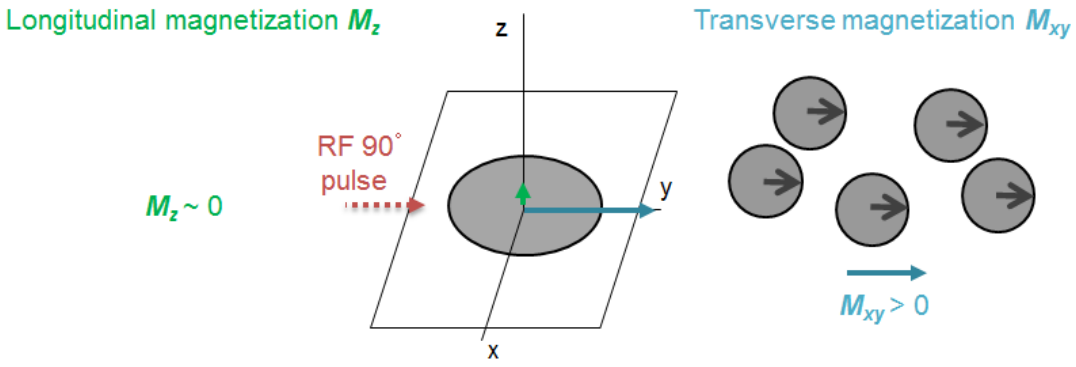


Figure 1.5 Orientation of magnetic moments with the application of radiofrequency.

The angle that the net magnetization *M* forms with the z-axis after the RF pulse is defined as flip angle. When the RF pulse is removed, the magnetization vector *M*, processing around the static magnetic field at Larmor frequency, returns to equilibrium. The MRI signal was generated by the rotating *Mxy* component and the receiver coil detects this signal as an oscillating current at Larmor frequency that gradually decays, known as a Free Induction Decay (FID).

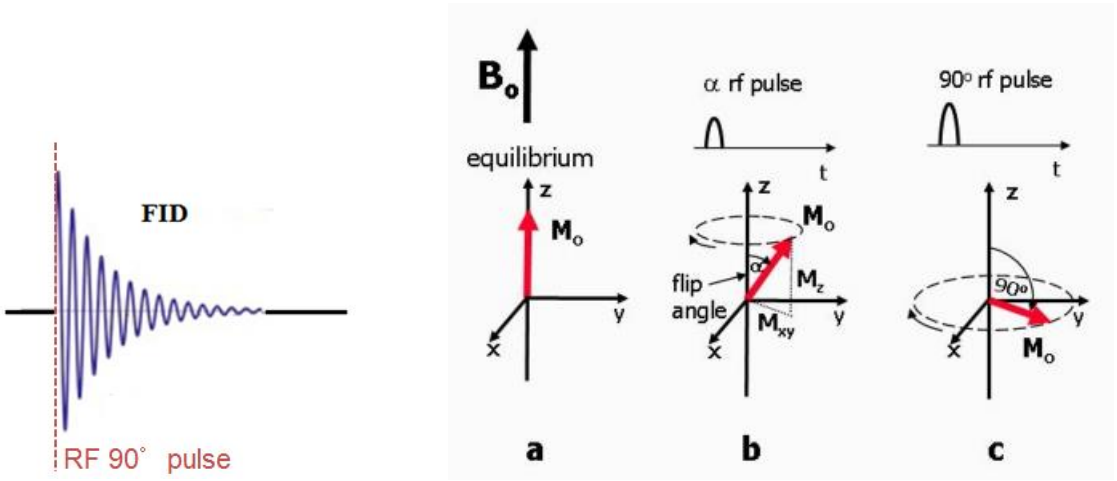


Figure 1.6 On the left: Free Induction Decay (FID). On the right: Effect of RF pulse: a) at equilibrium, the net magnetisation M_0 is aligned along the z-axis; b) when an RF pulse is applied, an angle, defined as flip angle, is created between the z-axis and *M*; c) when an RF pulse, defining a flip angle of 90° (90° RF pulse or saturation pulse), is applied, *M* lies entirely in the x-y plane and gives the maximum *Mxy* component (detectable signal amplitude) (*adapted from Ridgway, 2010*).

1.2.3 Relaxation times

The return to equilibrium of the net magnetization vector *M* is driven by three different relaxation phenomena, which differ according to the chemical and physical surrounding of each nucleus and to the tissue of belonging. The first phenomenon is related to the amplitude of net magnetization *M* in return to equilibrium, whereas the second and the third phenomena are related to loss of coherence or dephasing. These phenomena are:

- ***T1 relaxation time (spin-lattice relaxation time)***
  is a measure of how quickly the net magnetisation vector *M* returns to equilibrium (***amplitude***) in the direction of $B_0(M_z)$. In the transition between from the high energy state to the low energy, excited nuclei are associated with a loss of energy to the nearest nuclei. T1 relaxation is an exponential process. The length of the longitudinal magnetisation vector is given by the following equation:

$$M_z(t) = M_z(0)\left(1 - e^{-\frac{t}{T1}}\right)$$

  where *t*, the time after the 90° pulse RF, M_z is the magnetization at time t along z-axis, and $M_z(0)$ is the maximum amplitude of magnetization vector at full recovery. For *t* = T1, the signal recovers is about 63% of its initial value for *t* = T1.

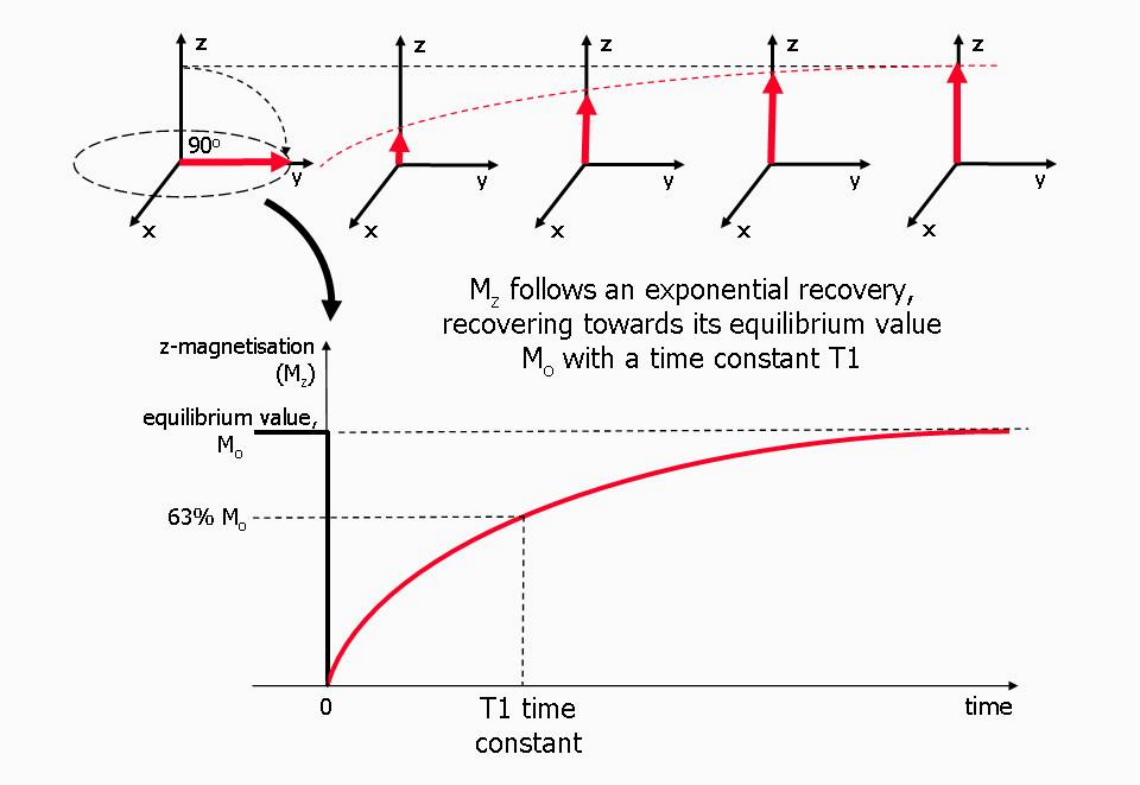


Figure 1.7 T1 relaxation time (*adapted from Ridgway, 2010*).

- ***T2 relaxation time (spin-spin relaxation)***

  is a measure of loss of coherence (***dephasing***) of spinning dipoles due to the rate of movement of protons in relation to tissue characteristics. When 90° pulse RF is applied, all the spinning dipoles are exactly in phase, and almost immediately RF is turn off, they lose coherence as some spin slightly faster than the others with a dephasing effect. The result is that the M_{xy} component of the magnetic vector decreases exponentially as a function of the T2 time constant.

$$M_{zy}(t) = M_{xy}(0)\left(e^{-\frac{t}{T2}}\right)$$

  When the magnetic moment are entirely out of phase, no net transversal magnetization remains.

- ***$T2^*$ relaxation time***

  is a measure of loss of coherence considering the ***inhomogeneity-induced dephasing*** of magnetic dipoles. These inhomogeneities may be due to intrinsic characteristics of the static field or to susceptibility-induced field distortions produced by the tissue or unwanted metallic materials. The previous T2 relaxation effect associated with the inhomogeneities of

magnetic field defined the T2* relaxation effect. It is exponential process with a time constant T2* and is less than or equal to T2 according this relation:

$$1/T2^* = 1/T2 + 1/T2_i$$

where $1/T2_i = \gamma \Delta B_i$ is the relaxation rate due to field inhomogeneities ΔB_i across a voxel. T2* is defined the "observed" or "effective" T2, whereas the T2 is considered the "natural" or "true" T2 of the tissue.

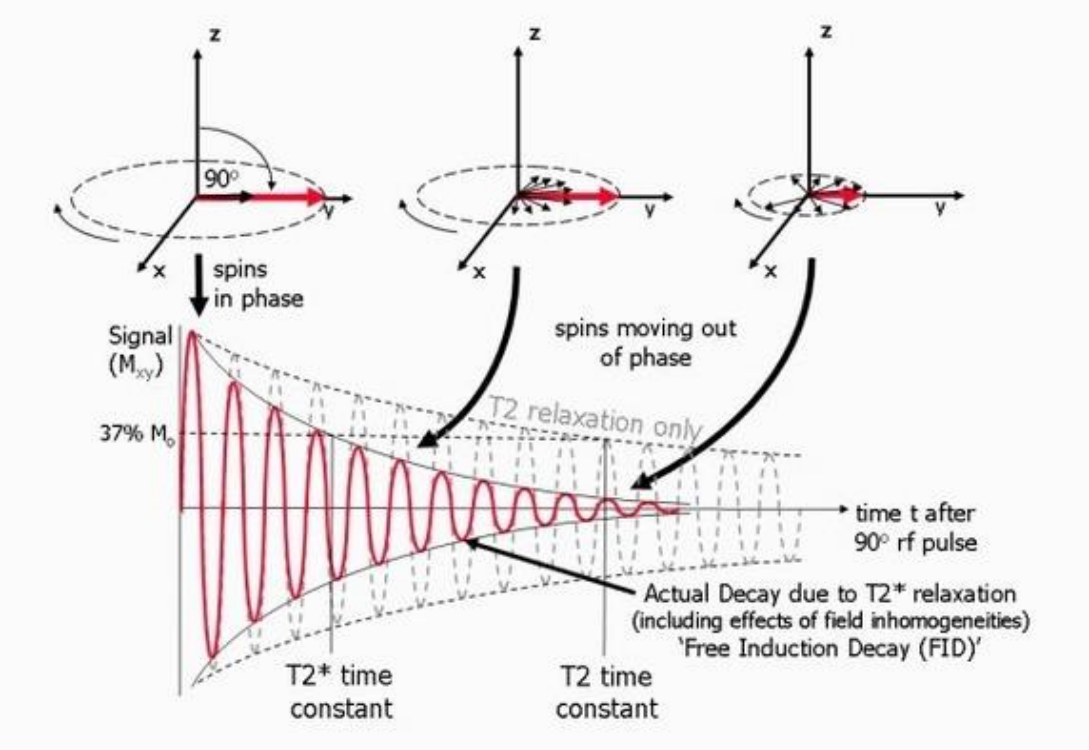


Figure 1.8 T2 and T2* relaxation time (*adapted from Ridgway, 2010*).

To increase the detectability of FID signal, is preferred to generate and measure the MR signal in the form of an ***echo***, because the subsequent localization gradients introduced additional unwanted de-phasing of the signal. Two different types of echo can be generated for MR imaging: gradient and spin echoes. Gradient echoes are generated with controlled application of magnetic field gradients, while spin echoes with the application of a 180° refocusing RF pulse after the 90° excitation pulse. Details are reported in the "*Image contrast*" section.

1.2.4 MRI signal localization: gradients

To localise and encode the MR echo signals is necessary to introduce magnetic field gradients. The application of a magnetic field gradient along a defined direction (x, y, or z) causes the strength of

the magnetic field only for that direction. Three different gradients are applied for the localization of the MRI signal source along the 3 main axis.

Step 1 - Selection of an image slice

To identify the resonance of protons of a single slice of tissue (along z-axis), a gradient magnetic field is applied simultaneously to the RF excitation pulse. When the RF pulse is trasmitted, only protons in a plane, that corresponds to the Larmor frequency, will respond, effectively defining a slice of tissue. Usually, the RF pulse comprises of a small range of frequencies, which create the transmit bandwidth of the RF pulse. This process is defined as "slice selection" and the gradient as the ***slice selection gradient***, ***GS***.

Step 2 - Phase encoding

Similarly, to identify the resonance of protons along y-axis, a ***phase encoding gradient, GP***, is applied along this direction after the RF excitation pulse, but before the echo signal detection. The protons change their relative phase according to their position along this gradient.

Step 3 - Frequency encoding

To identify the resonance of protons along x-axis, the ***frequency encoding gradient***, ***GF***, is applied along this direction during the echo signal detection.

The thickness of the slice is produced by the combination of the RF pulse bandwidth and the strength of the GS gradient, while the field of view (FOV) of the slice is related to the receiver bandwidths and the strength of GP and GF gradients.

1.2.5 MRI signal: pulse sequence diagram

The time of application of RF pulse, the 3 magnetic field gradients and the registration of echo signal are organized in a precise sequence that is represented by a pulse sequence diagram. This is an example of the application of the 3 previous steps. The time between the middle of the first RF pulse and the peak of the spin echo is called ***echo time (TE)***.

To ensure the maximum possible signal in the middle of the MR signal echo, additional gradient pulses are inserted both after the GS and before the GF. These additional gradient pulses compensate the de-phasing that is caused by these two gradients and are applied along the same direction as the imaging gradients, but with opposite slope. To compensate and reverse de-phasing effect for the GS, a re-phasing gradient that is only half the length of the slice selection gradient is applied after RF pulse. To compensate de-phasing effect for the GF and ensure a maximum

amplitude of echo in the middle of the sampling period, GF is normally preceded by a re-phasing gradient that is only half the length of it.

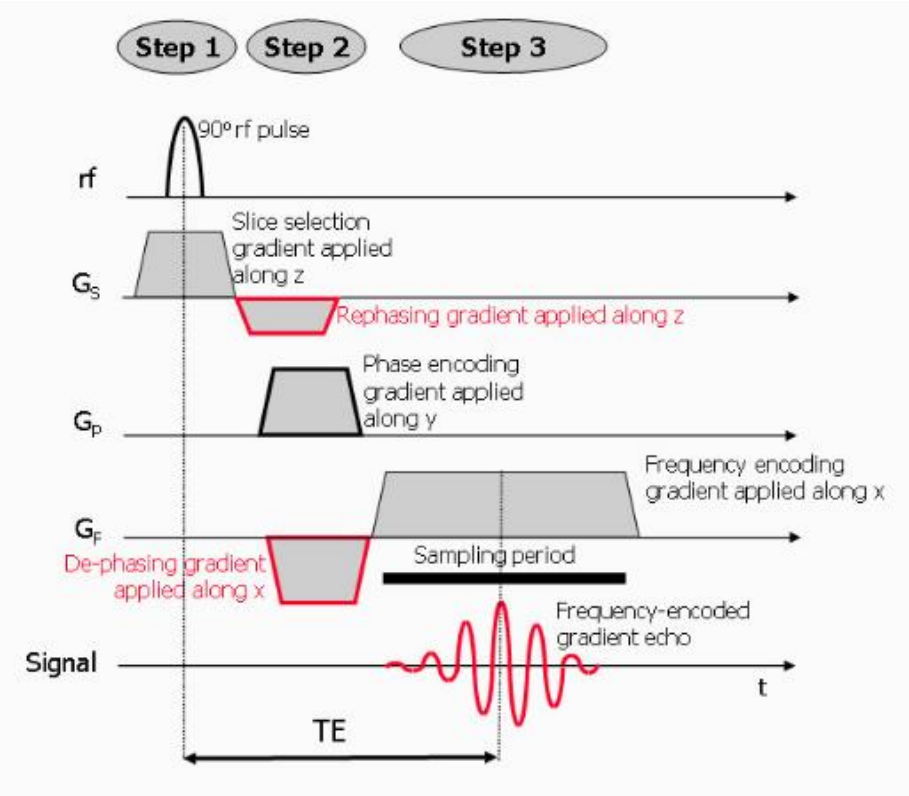


Figure 1.9 Scheme of a pulse sequence (*adapted from Ridgway, 2010*).

1.2.6 Image reconstruction: Fourier analysis and k-space

To transform the time-dependent MR signal into its different frequency components a Fourier transform is used: the amplitude of each frequency component can be mapped onto a location along the GF to determine the relative amount of signal at each location. To observe the phase changes induced by the GP, the above three-step process (slice selection, phase encoding and frequency encoding), each time applying the same slice selection and frequency encoding gradient, but changing phase encoding, is repeated to generate different signal echoes. This is done by increasing the strength of the GP for each repetition by equal increment. The time interval between each repetition of this process is known as the ***repetition time (TR)***. In each phase encoding step, the signal recorded are digitalized in a spatial frequency domain matrix, which is called ***K-space***. The dimensions of this matrix are the defined by the sample number in the frequency encoding direction and the sample number in the phase encoding direction. A two-dimensional (2D) Fourier transform is used to decode the information present in frequency and the phase.

The relation between the final image space and the acquired *K*-space is inverse: whereas the coordinates of the image represent spatial position (x and y), the coordinates of *K*-space represent 1/x and 1/y, referred to as spatial frequencies, *kx* and *ky*. Each value of *K*-space quantifies the amount of spatial frequency contained within the corresponding image.

In *K*-space:

- a low spatial frequency (data in centre of *K*-space) contributes mostly the signal content and contrast of the image;
- a high spatial frequency (data at the edge of *K*-space) contributes fine details or edges, effectively defining the spatial resolution of the image.

A complete representation of the scanned object is obtained only when the whole range of spatial frequencies is acquired to obtain

The filling of *K*-space can be done by activating phase encoding and frequency encoding gradients in different ways to collect image. Some of these different filling techniques are:

- linear: involves filling line by line until our entire space is filled. This is the standard way of filling our *K*-space.
- centric: involves filling high signal amplitudes in the center of our *K*-space and filling outward to the periphery. This is useful when performing contrast enhanced imaging.
- elliptical: involves filling our *K*-space in a spiral fashion starting in the center and working our way out to the periphery. Similarly, this is beneficial in contrast enhanced imaging.
- radial: by collecting data like cutting a pie into slices, it allow to reduce motion in slice plane.

It is possible to change the way we fill our *K*-space to save scan time. For example, "half Fourier" is a technique where only half of *K*-space in the phase encoding direction is filled. The other half of the *K*-space is then calculated from this data. "Partial echo" is the analogous parameter for the frequency direction.

1.3 Image parameters and sequences

Some important parameters determine the characteristics of images:

- ***TR***: determines the speed of acquisition of MR images and affects the image contrast.
- ***FOV***: the number of phase encoding steps (N_{PE}) determines the number of pixels in the phase encoding direction of the final image.

- ***Image acquisition time***: the minimum value is determine by TR x N_{PE}.
- ***Pixel size***: the signal is proportional to the voxel volume and the appropriate sequence-specific relaxation factor. It is important to decide the optimal trade-off for the voxel size to obtain a good compromise between a large value for an adequate SNR (signal is proportional to the voxel volume) and a small value to permit the visualization of small anatomical or pathological details.
- ***Thickness of slice***: is determined by both RF pulse bandwidth and steepness (or strength) of the *GS* gradient.

1.3.1 Sequences

Spin echo pulse sequence

In spin-echo sequence, a 180° refocusing RF pulse is applied after the 90° excitation pulse. This method is based on the hypothesis that the de-phasing of T2 relaxation is a random, irreversible process, while the de-phasing of field inhomogeneities due to static field is potentially reversible. Therefore, this implies that signal obtained with this procedure can be influenced by T2 relaxation only, and not by the effect of T2* relaxation. The 180° refocusing RF pulse flips the direction of the magnetization vector *M* in the x-y plane through 180° to reverse the de-phasing of magnetic field inhomogeneities. As response, the proton spins come back into phase causing an increase in amplitude of FID, which reach a maximum at the echo time, TE. This MRI signal obtained with the 180° RF refocusing pulse is known as a ***spin echo***. For a complete compensation of the dephasing due to field inhomogeneities at TE, the 180° pulse must be applied at time TE/2.

Gradient echo sequence

In a gradient sequence, additional controlled magnetic field gradients are applied. Usually, the application of a magnetic field gradient produced de-phasing of proton spins along the direction of the gradient, associated reduction of the amplitude of the FID signal. This principle for this method is that the application of a second magnetic field gradient along the same direction with an equal amplitude and duration, but in the opposite direction can compensate this de-phasing effect.

The maximum echo signal is obtained when this second magnetic field gradient is applied and compensate the de-phasing effect. The signal that is obtained through the switching of the gradient direction is known as a ***gradient echo***.

1.3.2 Image contrast

MR imaging is able to generate contrast between different soft tissue types, because each tissue is characterized by different T1 and T2 relaxation times. The choice of the pulse sequence parameters permits to create a MR signal that depends from particular tissue on its relaxation properties is controlled by: TE, TR and flip angle. Tissue's T1 and T2* relaxation times on the signal are determined by:

- the choice of TR and TE, while flip angle is fixed at 90° in spin echo pulse sequences;
- the TR, TE and flip angle in gradient echo pulse sequences.

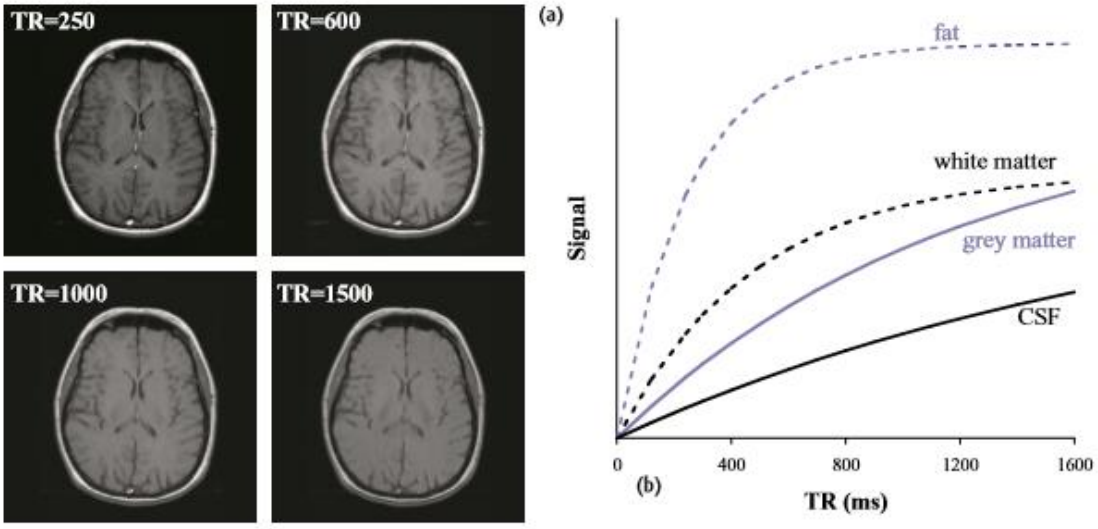
Spin-echo

TR	TE	
	Short (less than 40 ms)	Long (more than 75 ms)
Short (less than 750 ms)	T_1-weighted	Not useful
Long (more than 1500 ms)	PD-weighted	T_2-weighted

Gradient-echo

Flip angle α	TE	
	Short (less than 15 ms)	Long (more than 30 ms)
Small (less than 40°)	PD-weighted	T_2-weighted
Large (more than 50°)	T_1-weighted	Not useful

Table 1.1 Selection of TR, TE, flip angle in spin-echo and gradient-echo sequences (adapted from McRobbie, MRI from picture to proton).



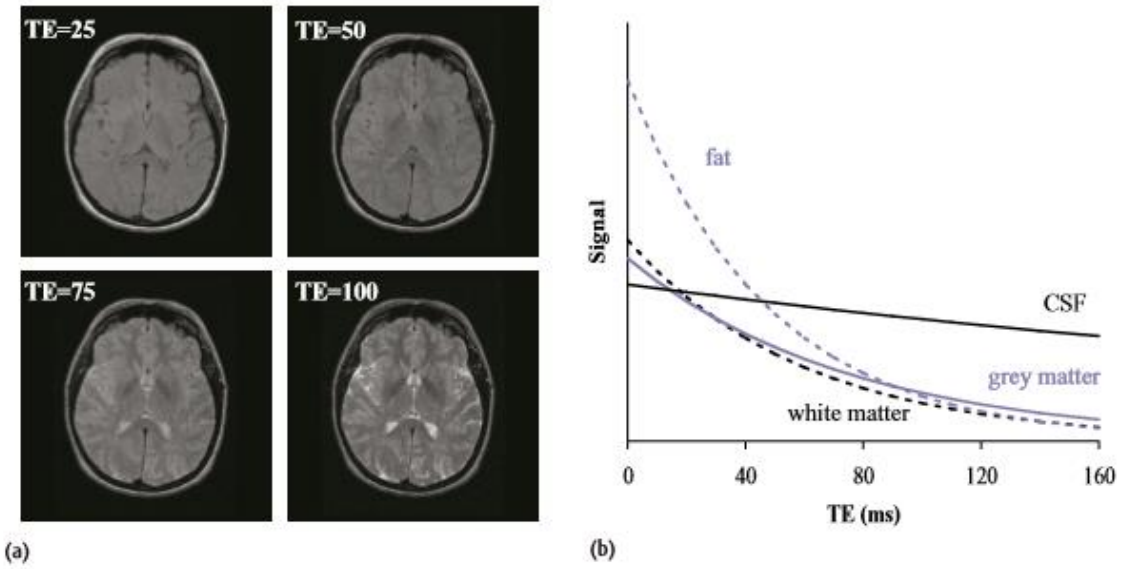


Figure 1.10 Image contrast in spin-echo sequence. On top: (a) T1 image contrast with TE = 10 ms and various TR. (b) Signal intensity in relation to TR. On bottom: T2 images contrast with TR = 1500 ms and various TE. (b) Signal intensity for brain tissues plotted against TE (adapted from McRobbie, MRI from picture to proton).

1.4 Functional MRI

1.4.1 Blood oxygen level dependent (BOLD) signal

Although comprising only 2% of the total body mass, the brain receives 12-15% of the cardiac output and consumes around 20% of the oxygen inhaled (Siesjo, 1978).

The BOLD contrast mechanism was first described by Ogawa (1990) and is based on the relation between physiological responses and brain activation: the coupling between the physiological responses and neural activation is tight and well localized, but the time delay of first response is relatively slow compared to the second activity.

BOLD imaging uses the different magnetic proprieties of endogenous MRI contrast agent of haemoglobin as the source of signal detectable in MRI: the oxygenated hemoglobine (Hb) is diamagnetic, because it has no magnetic moment due to unpaired electron and, while the deoxyhemoglobin (dHb) is paramagnetic, because it has magnetic moment due to unpaired electron. By accentuating the effects of this agent through the use of particular magnetic sequences, such as gradient-echo techniques, image of contrast reflecting the blood oxygen level can be obtained.

The hyper-perfusion of the local tissue is the basis of the BOLD contrast (McIntyre et al., 2003). During a rest phase, blood contains dHb in a magnetic field produced little field distortion induced

by the difference in magnetic susceptibility relative to the surrounding. When an area is involved in a specific task, changes in dHb content associated with hyper-perfusion of the local tissue can be detected with MRI signal.

1.4.2 BOLD as a vascular response

The oxygen consumption and blood flow in response to a neural stimuli determine the shape of the hemodynamic response function (HRF).

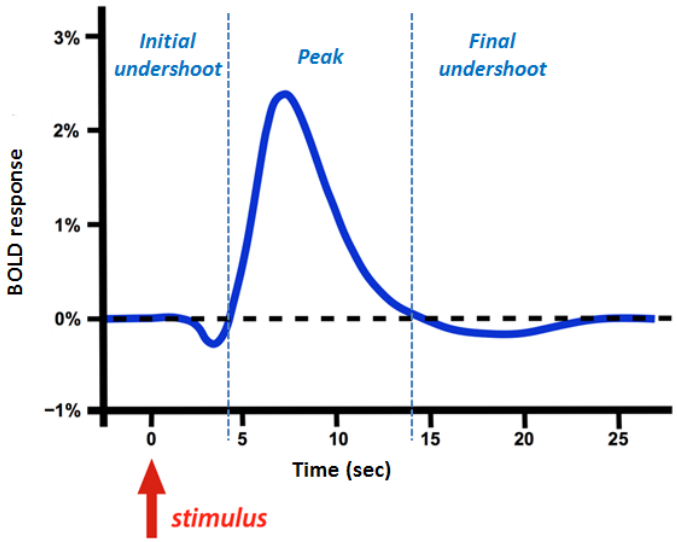


Figure 1.11 Vascular response of BOLD contrast: steps

The timeline of the HRF response is composed by an initial onset of the stimulus around 2 seconds, corresponding to the time that blood travels from arteries to capillaries and draining veins (Kwong et al., 1992). Subsequently, HRF response reaches a plateau in the interval of 6–12 seconds and returns to the baseline with a prolonged post-stimulus undershoot.

In detail, the process is structured in these three steps as follow (Fig. 1.11):

1. Initial undershoot (Hypoxic phase) = before any vascular response, when a neural activity start (onset), dendrites rapidly

- consume oxygen and
- increase of the cerebral metabolic rate of it ($CMRO_2$) (Kasischke et al., 2004).

The main consequences of this process lead to a reduction in oxygen pressure and Hb (Devor et al., 2005; Zhong et al., 1998), which corresponds to a decrease of the observed signal. The amplitude of

this undershoot is smaller than main vascular response, around 10% of the amplitude of peak of vascular response.

2. Peak of vascular response (Hyperoxic phase) = the vascular response to neural activity start with

- an increase of the cerebral blood flow (CBF).

The effect is a great concentration of Hb locally compared to dHb: this determines an increase of the signal. This phenomenon is principally related to an increase in blood cell velocity and/or capillaries dilatation and not to additional capillaries recruitment (Villringer et al., 1994).

3. Final undershoot = in the return to baseline, there is

- a decrease of CBF.

Two different phenomena that determine two different hypotheses were described for the final undershoot in HRF:

- first hypothesis: the differences between the active response of arteries (dilatation) and the passive venous response (Buxton et al., 1998) determine an increased cerebral blood volume (CBV). The concertation of dHB is more than Hb, so it is recorded as a signal decrease.

- second hypothesis: a residual the metabolic activity taking place after the stimulus neuronal activity related to uncoupling between CBF and $CMRO_2$. This option is introduced because Frahm et al. (2008) showed that in humans there was no change in CBV.

It is important to underline that HRF shows considerable variability between subjects (Aguirre, Zarahn & D'Esposito 1998, Handwerker et al., 2004) and between brain regions (Buckner et al., 1996), although a canonical estimation is performed for ideal response.

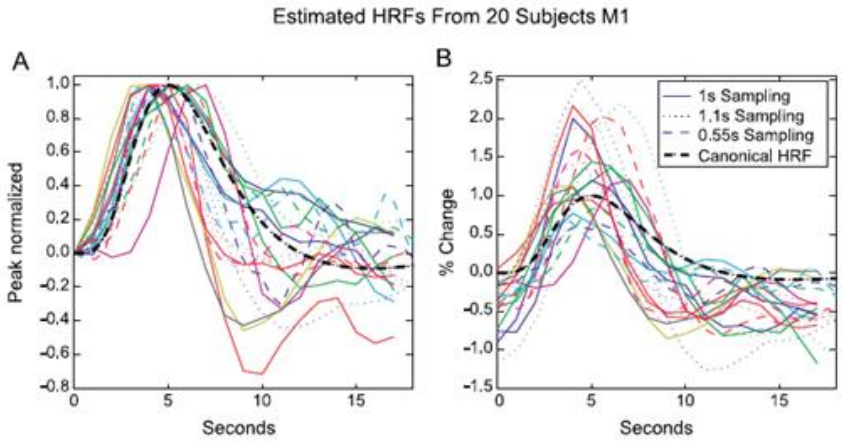


Figure 1.12 Estimation of HRF (Source: Handwerker et al., 2004)

Moreover, considering the setup of an experiment is important to consider that there are many factors that affect HRF, such as drugs (alcohol, caffeine), aging, disease (Handwerker et al., 2012).

1.4.3 BOLD as a neural response

Previous studies with simultaneous fMRI and electrophysiological recordings suggest that the BOLD contrast directly reflects the neural responses elicited by a stimulus (Logothetis, 2001). The inputs, which neurons receive through post-synaptic potentials, increase (excitatory post-synaptic potentials) or decrease (inhibitory post-synaptic potentials) the membrane voltage. If the sum of all post-synaptic potentials at the axon pushes the voltage above the threshold, the neuron will fire an action potential. In electrophysiology, two important measures are recorded: action potentials and local field potentials (LFPs). LFPs reflect post-synaptic potentials and multi-unit activity (MUA) reflect action potentials.

A initial theory hypothesized a linear relationship between BOLD and neural responses during short stimulus presentation. This appears particular evident between HRF and LFP, and less with MUA: this evidence suggests that BOLD is more related to post-synaptic activity than neural activity. However, variation of haemoglobin concentration did not respond linearly with the stimulus duration. This suggests that there are different factors that influence the BOLD response and contribute to the non-linearity behaviour between the neural response and the blood supply and during the oxygen consumption (Logothetis, 2003).

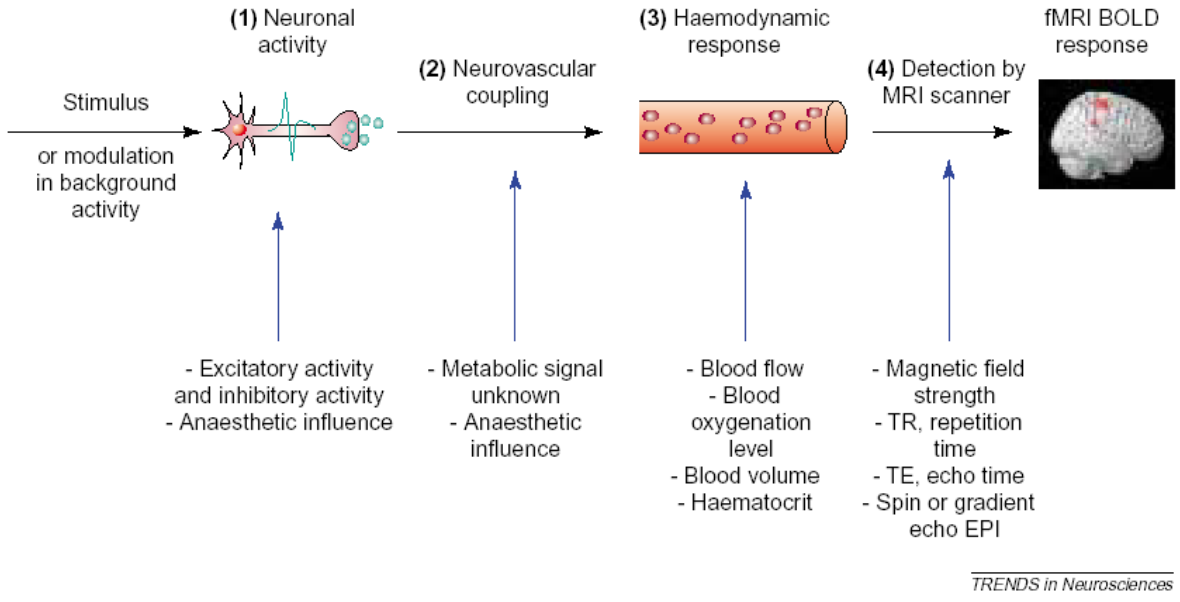


Figure 1.13 From stimulus to fMRI BOLD response (Source: Arthurs and Boniface, 2002)

1.4.4 Echo planar imaging

In echo-planar imaging (Mansfield, 1977; Ordidge, 1999), multiple lines of imaging data are acquired after a single RF excitation. A spin-echo-planar imaging sequence is created by two consecutive RF pulses in this order: a 90° and a 180° RF pulses. It differs from a conventional spin-echo sequences because, after the 180° RF pulse, the frequency-encoding gradient oscillates rapidly from a positive to a negative amplitude, forming a train of gradient echoes.

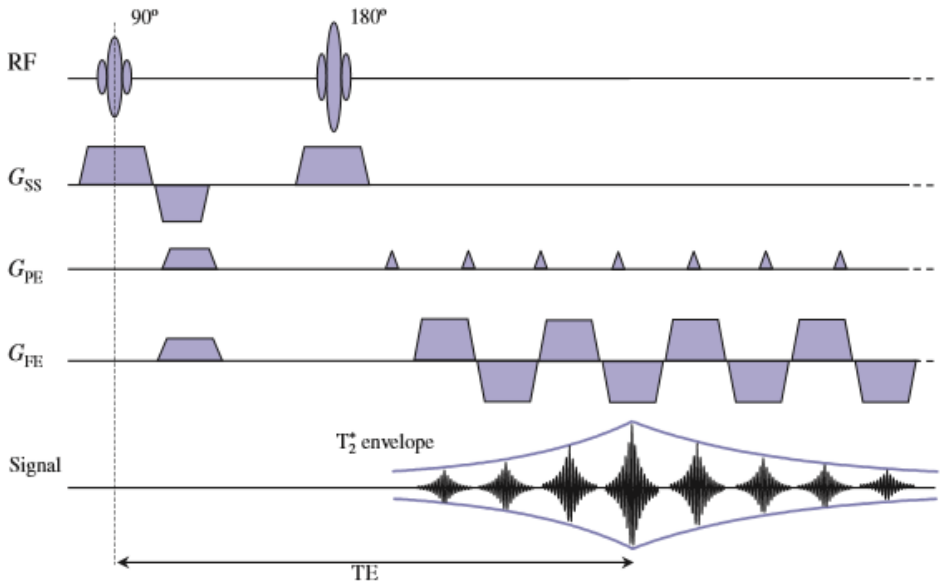


Figure 1.14 Pulse sequence scheme of a spin echo-planar imaging (adapted from McRobbie, MRI from picture to proton).

In a single-shot echo planar sequence, the entire range of phase encoding steps are acquired in one TR. In multi-shot echo-planar imaging, the range of phase steps is equally divided into several "shots" or TR periods.

The main advantages of this sequence are the reduced imaging time (high temporal resolution) and a decreased motion artefact. On the other hand, it is sensitive to susceptibility effects and long gradient echo train causes greater T2* weighting.

In the thesis, we will use the term:

- “voxel” single 3D unit of spatial resolution in the image;
- “volume” to refer an individual 3D acquisition performed in one TR of the brain (single time point).

- " run" or "scan" to refer a series of continuous volumes acquired during a single MRI sequence (multiple time points);
- "session" to refer a multiple runs with brief interruption among the single run.

For each subject that undergo an fMRI experiment, a single scan (or multiple scans) is acquired and the group of all fMRI data of the subjects constitutes a "dataset".

The basic strategy for the analysis of fMRI data aims at identifying the voxels in which the BOLD signal is significantly correlated with a reference condition of stimulation, and therefore can be re-framed as the problem of identifying the noise sources of signal variance.

1.5 Noise sources

Noise has many contributions, difficult to disentangle in order to elicit the real BOLD neurophysiological substrates of data. In fMRI data there are a variety of fluctuations induced by processes beyond the control of the experimenter, which make fMRI recorded signal inherently noisy. In fMRI data, noise may be roughly divided into two groups:

- ***MR physics-related artefacts*** (thermal noise, drift noise, signal dropout due to imperfect switching of the slice select gradients, EPI "ghosting"), and
- ***Subject-related noise*** (head motion effects, fluctuations induced by the cardiac and respiratory cycles, and spontaneous low frequency fluctuations of the baseline signal).

These sources of noise can be partially addressed during the periodic quality assurance protocol of the scanner and the acquisition stage (e.g. real-time shimming to reduce the inhomogeneity of some regions, padding to limit head motion). However, some of them persist and other post-processing steps are necessary to minimize their impact on signal of interest.

Factors can introduce unintended variations in fMRI measurement.

Category	Factor
1. Acquisition-related variations	Scanner make and model (Friedman and Glover, 2006b), sequence type (spiral vs. echo planar; single-echo vs. multi-echo) (Klarhofer et al., 2002), parallel vs. conventional acquisition (Feinberg et al., 2010; Lin et al., 2005), coil type (surface vs. volume, number of channels, orientation), repetition time, number of repetitions, flip angle, echo time, and acquisition volume (field of view, voxel size, slice thickness/gaps, slice prescription) (Friedman and Glover, 2006a)
2. Experimental-related variations	Participant instructions (Hartstra et al., 2011), eyes-open/eyes-closed (Yan et al., 2009; Yang et al., 2007), visual displays, experiment duration (Fang et al., 2007; Van Dijk et al., 2010)
3. Environment-related variations	Sound attenuation measures (Cho et al., 1998; Elliott et al., 1999), attempts to improve participant comfort during scans (e.g., music, videos) (Cullen et al., 2009), head-motion restraint techniques (e.g., vacuum pad, foam pad, bite-bar, plaster cast head holder) (Edward et al., 2000; Menon et al., 1997), room temperature and moisture (Vanhoutte et al., 2006).
4. Participant-related variations	Circadian cycle (Shannon et al., 2013), prandial (Haase et al., 2009), caffeine (Rack-Gomer et al., 2009), and nicotine status (Tanabe et al., 2011), sleepiness/arousal (Horovitz et al., 2008), sleep deprivation (Samann et al., 2010), scanner anxiety (de Bie et al., 2010), and menstrual cycle status (for women) (Protopopescu et al., 2005)

Table 1.2 Detailed description of factors that influence fMRI signal: MR physics-related artefacts (category 1) and subject-related noise (categories 2, 3, 4) (Source: Yan et al., 2013)

MR physics-related artefacts are intrinsic to the imaging acquisition process. For example, the random thermal noise arises primarily from the body that induce random signals in the receiver coil, and it is spread throughout the voxels of the image. The thermal noise is a uniform random Gaussian noise, that is depend on the voxel's signal amplitude and independent of the noise in the other voxels. These artefact related to MRI physic can be reduced if a regular quality assessment of the MRI imaging is performed on the scanner in order to control main field inhomogeneities, eddy non-linearity of imaging gradients and other technical problems.

Subject-related noise is related to displacement of head position of subject in the scanner due to the different coordinate systems of the two parts (scanner: global coordinate system; subject: local coordinate system) and vital function of subjects (breathing, heartbeat) (Lund et al., 2006). Motion movements can have a serious effect on the fMRI signal: small movement less than one voxel can cause signal dropout and can generate spurious activations or mask real activations. This is emphasized in long scanning session, where is very difficult to remain stationary for this long period. Subject movement is often measured with derived summary statistics based upon absolute head realignment parameters: the position of the head in space at each volume is estimate compared to a reference volumes (the first of the mean volume) using rigid body transforms. If these movements are not estimate and modelled correctly, these systematic spurious signal changes can obscure patterns of interest in BOLD signal.

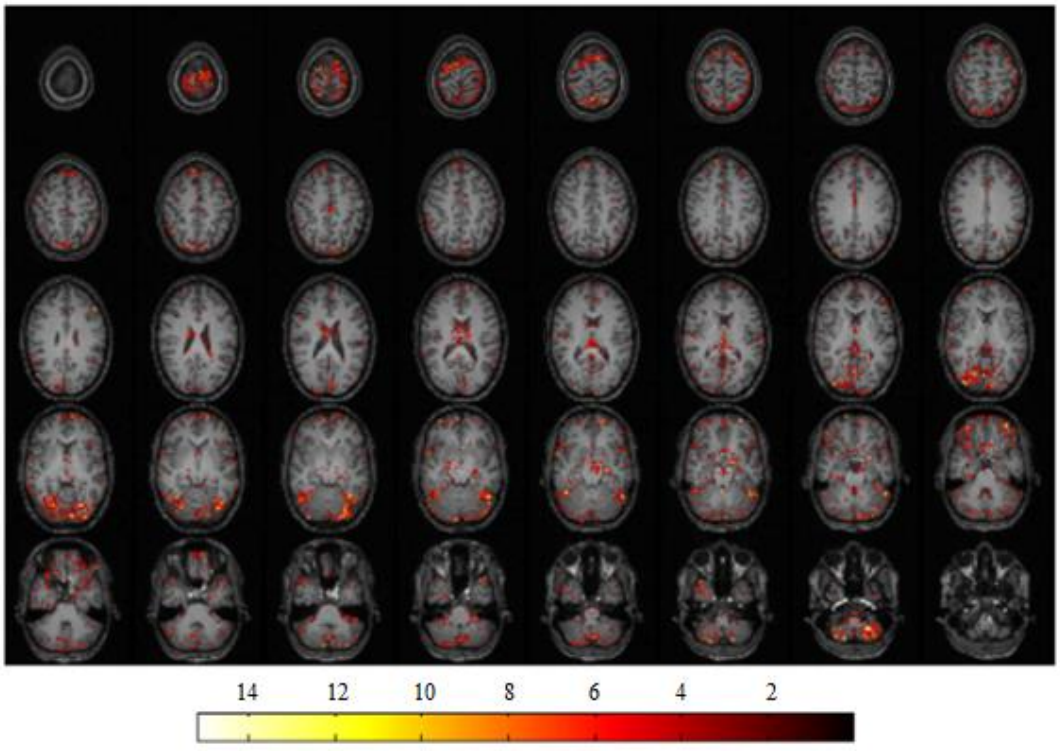


Figure 1.15 Subject movements noise can be identified near the edges of the brain (adapted from Lund et al., 2006).

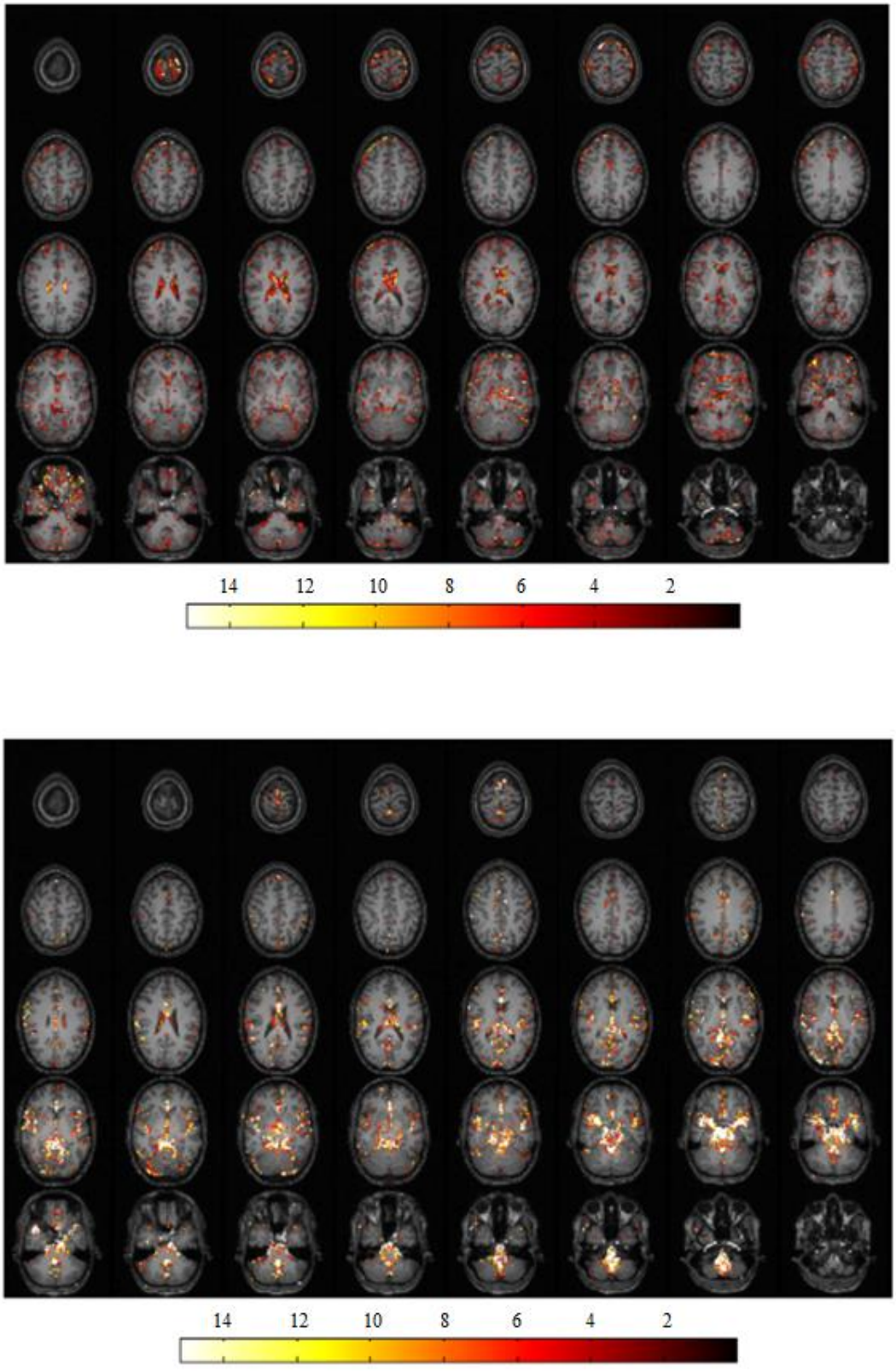


Figure 1.16 On top: respiratory noise is present near the edges of the brain as well as near in the larger veins and in the ventricles. On bottom: cardiac noise is present near larger vessels, comprising medial cerebral artery and Circle of Willis (adapted from Lund et al., 2006).

Physiological noise refers to various physiological processes, such as cardiac fluctuation and low-frequency components related to respiration (Birn et al., 2008a; Chang et al., 2009a), not directly associated with the functional region of the brain that is of interest. The partial overlap of the frequencies of these signals to BOLD related frequencies make these noises difficult to minimize in post-processing. These signals can be recorded during acquisition or can be estimate from fMRI data retrospectively, and they can be used to remove the physiological fluctuations from the fMRI time series (Birn et al., 2008b; Chang and Glover, 2009b). For example, respiratory-induced noise is mainly present near the edges of the brain as well as near the larger veins and in the ventricles, while cardiac-induced noise can be identified near larger vessels, such as medial cerebral artery and Circle of Willis. For these reasons, it is important to consider cautiously activations in these regions.

These sources of noise do not have the same impact on signal of interest. The noise in fMRI images has components that are either additive or multiplicative (Greve et al., 2011).

- The **additive component** of noise is often referred to as "***background noise***" and includes a scanner-dependent component related to thermal contributions from the subject or phantom, electronics, "spike" noise, and spurious RF interference signals injected through failure of the magnet room's RF shielding.
- The **multiplicative noise** can have two main sources. The first source is known as "***scanner instability***" related to the instrument-induced fluctuations in, for example, resistive shim currents or gradient, RF, or receiver amplifier gain or phase. The second source is produced by the ***physiological noise*** due to human subject.

Moreover, if we consider high field strength, it is important to know that there is an increase of raw signal to noise ratio (SNR), but also of spurious signals, such as thermal noise and physiological noise. So particular attention need when high fields are used.

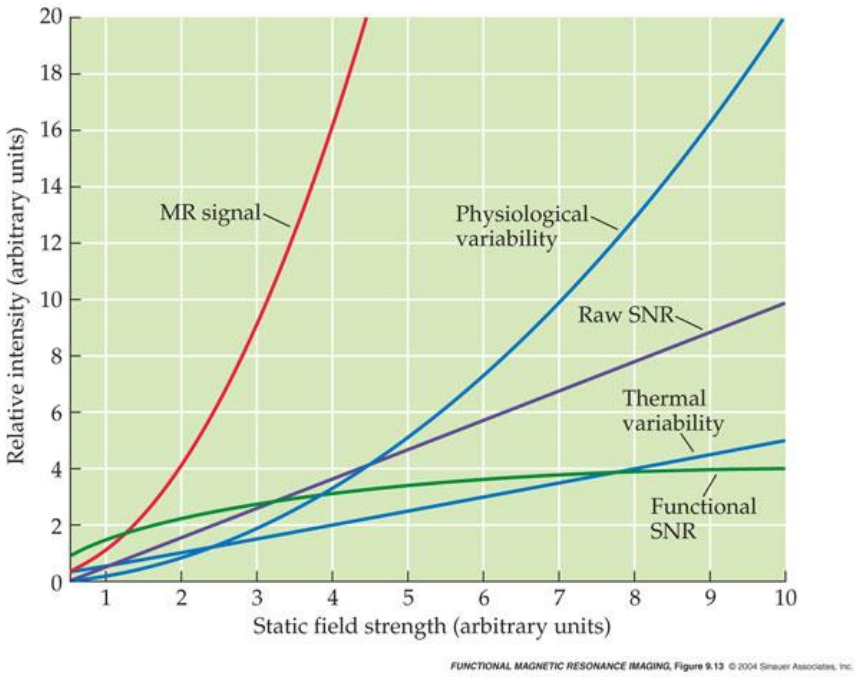


Figure 1.17 Relation between signal and field strength. (Source: Huettel, Song & Mc Carthy, 2004).

To reduce spurious noise, multiple post-processing steps are typically conducted to increase signal to noise and remove signal contributions from motion and physiological noise. Each of these steps raises potential interpretative issues and opportunities for methodological optimization.

References

Aguirre GK, Zarahn E, D'esposito M. (1998). The variability of human, BOLD hemodynamic responses. Neuroimage;8(4):360-9.

Arthurs OJ, Boniface S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? Trends Neurosci;25(1):27-31.

Birn RM, Smith MA, Jones TB, Bandettini PA. (2008a). The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. Neuroimage;40(2):644-54.

Birn RM, Murphy K, Bandettini PA. (2008b). The effect of respiration variations on independent component analysis results of resting state functional connectivity. Hum Brain Mapp; 29(7):740-50.

Bloch F., Hansen W., and Packard M. (1946). Nuclear induction experiment. Phys. Rev., 70(7):474–485.

Buckner RL, Bandettini PA, O'Craven KM, Savoy RL, Petersen SE, Raichle ME, Rosen BR. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proc Natl Acad Sci U S A;93(25):14878-83.

Buxton R. B., Wong E. C., Frank L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model.

Chang C, Cunningham JP, Glover GH. (2009a). Influence of heart rate on the BOLD signal: the cardiac response function. Neuroimage;44(3):857-69.

Chang C, Glover GH. (2009b). Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. Neuroimage;47(4):1448-59.

Devor A, Ulbert I, Dunn AK, Narayanan SN, Jones SR, Andermann ML, Boas DA, Dale AM. (2005). Coupling of the cortical hemodynamic response to cortical and thalamic neuronal activity.
Proc Natl Acad Sci U S A.

Frahm J, Baudewig J, Kallenberg K, Kastrup A, Merboldt KD, Dechent P. (2008). The post-stimulation undershoot in BOLD fMRI of human brain is not caused by elevated cerebral blood volume. Neuroimage;40(2):473-81. doi: 10.1016/j.neuroimage.2007.12.005.

Greve DN, Mueller BA, Liu T, Turner JA, Voyvodic J, Yetter E, Diaz M, McCarthy G, Wallace S, Roach BJ, Ford JM, Mathalon DH, Calhoun VD, Wible CG, Brown GG, Potkin SG, Glover G. (2011). A novel method for quantifying scanner instability in fMRI. Magn Reson Med;65(4):1053-61.

Handwerker DA, Ollinger JM, D'Esposito M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses.
Neuroimage;21(4):1639-51.

Handwerker DA, Gonzalez-Castillo J, D'Esposito M, Bandettini PA. (2012). The continuing challenge of understanding and modeling hemodynamic variation in fMRI. Neuroimage;62(2):1017-23. Review.

Huettel S., Song A. & Mc Carthy G. (2004). Functional Magnetic Resonance Imaging, Second Edition

Kasischke KA, Vishwasrao HD, Fisher PJ, Zipfel WR, Webb WW. (2004). Neural activity triggers neuronal oxidative metabolism followed by astrocytic glycolysis. Science;305(5680):99-103.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci USA;89: 5675–5679.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature;412(6843):150-7.

Logothetis NK. (2003).The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci;23(10):3963-71. Review

Lund TE, Madsen KH, Sidaros K, Luo WL, Nichols TE. (2006). Non-white noise in fMRI: does modelling have an impact? Neuroimage;29(1):54-66.

Mansfield P. (1977). Multi-planar image formation using NMR spin echoes. J Phys C: Solid State Phys; 10:L55-L58.

Mcintyre M, Richter W., Morden D.,Wennerberg A., Frankenstein U. (2003). Blood oxygenation level dependent functional magnetic resonance imaging. Journal Concepts in Magnetic Resonance; 16A(1):5.15.

McRobbie D., Moore E., Graves M., Prince M. MRI from picture to proton. 2006.

Ogawa S, Lee TM, Kay AR, Tank DW. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A;87(24):9868-72.

Ordidge R. (1999). The development of echo-planar imaging (EPI):1977-1982. Magn Reson Mater Phys Biol Med;9:117-121.

Purcell E. M., Torrey H. C., and Pound R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. Phys. Rev., 69(1):37–38.

Ridgway JP. (2010). Cardiovascular magnetic resonance physics for clinicians: part I. J Cardiovasc Magn Reson;12:71. Review.

Siesjö BK. (1978). Brain energy metabolism and catecholaminergic activity in hypoxia, hypercapnia and ischemia. J Neural Transm Suppl;(14):17-22.

Villringer A, Them A, Lindauer U, Einhäupl K and Dirnagl U. (1994). Capillary perfusion of the rat brain cortex. An in vivo confocal microscopy study.

Yan CG, Craddock RC, He Y, Milham MP. (2013). Addressing head motion dependencies for small-world topologies in functional connectomics. Front Hum Neurosci;7:910.

Zhong J, Kennan RP, Fulbright RK, Gore JC. (1998). Quantification of intravascular and extravascular contributions to BOLD effects induced by alteration in oxygenation or intravascular contrast agents. Magn Reson Med;40(4):526-36.

Chapter 2

Unsupervised learning method: clustering

2.1 Basic concept of clustering

Clustering and classification are two important fields in Data Mining. The former is a predictive method and is used mostly as a supervised learning approach, while the latter is a descriptive method and is applied as an unsupervised learning (Veyssieres and Plant, 1998).
The different goals of these two methods are related the different assessment of the output. In classification tasks, a known variable is exploited to predict unknown or future values of other variables. It is used to define an interpretable model of a given phenomenon, using an extrinsic assessment in which predicted groups are compared to a reference set of class label of a known variable. In opposition, clustering is used to discover a new set of categories, in which the new groups are of interest in themselves. Therefore, the assessment of these new groups is intrinsic, with an unknown label. In this thesis, we would like to discover unknown patterns, so we will focus on clustering algorithms.

The aim of a cluster analysis is to find groups of objects in a dataset such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups. It can be used to:

- discover: group of objects with similar characteristics;
- summarize: reduce the size of a large dataset.

2.1.1 Type and definition of clusters

According to definition of Tan et al., 2006 and Steinbach et al., 2003, clusters can be characterized by the following proprieties/characteristics:

A. ***Well-Separated***: a cluster is a set of points such that any point in a cluster is *closer* (or more similar) *to every other point* in the cluster *than to any point not in the cluster*. A threshold is

necessary to define how a point is sufficiently close to one other. This definition is satisfied only when cluster are far from each other.

B. ***Center-based***: a cluster is a set of objects such that an object in a cluster is closer (more similar) to the "*center*" of a cluster, than to the center of any other cluster. The "*center*" of a cluster is often identified with a ***centroid***, the average of all the points in the cluster, or a ***medoid***, the "most representative" point of a cluster.

C. ***Contiguity-based (Nearest Neighbor or Transitive)***: each point in a cluster is closer to *at least* one point in its cluster than to any point in another cluster.

D. ***Density-based***: a cluster is a dense region of points, which is separated by low-density regions, from other regions of high-density. Usually, clusters with noise or outliers are characterized by low density regions.

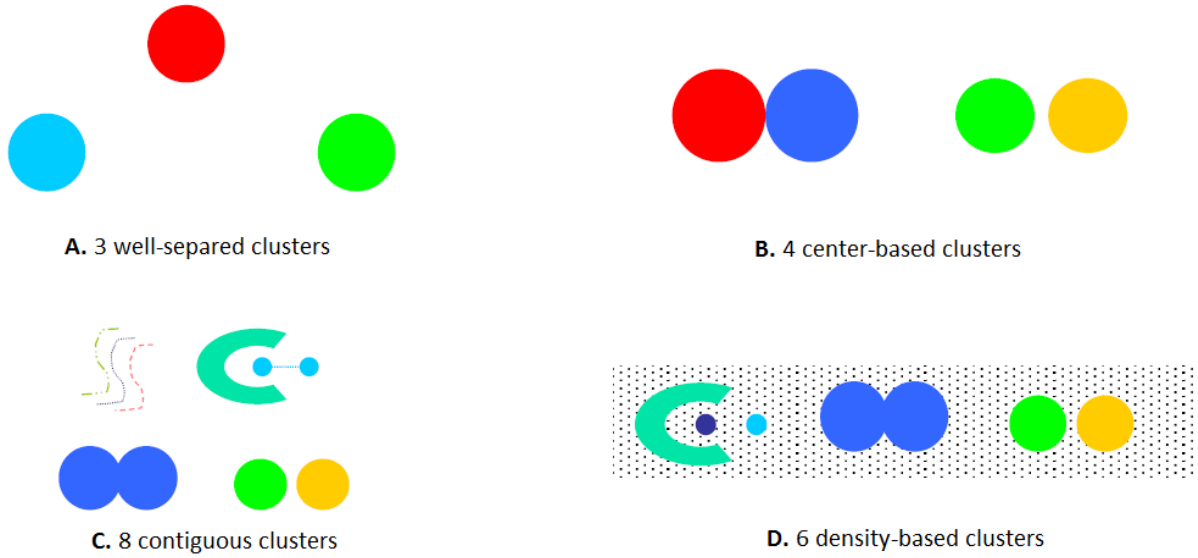


Figure 2.1 Type of clusters (2 dimensional points) (Source: Tan et al., 2006).

A cluster is a set of objects that together create a region with a uniform local property, such as ***size***, ***density***, s***hape*** (globular, non-globular).

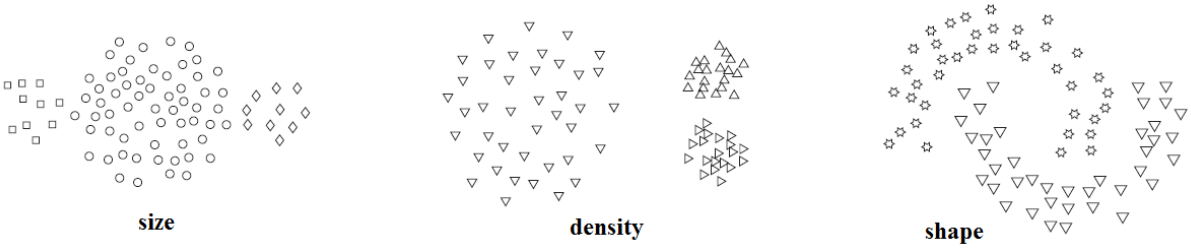


Figure 2.2 Proprieties of clusters (2 dimensional points) (Source: Tan et al., 2006).

2.1.2 Selection of variables and pre-processing

The selection of variables to include in dataset is really important to achieve a best result with clustering algorithm (Kaufman & Rousseauw, 1990; Lletı et al., 2004). Kaufman and Rousseauw 1990 said that:

> *in all of this [cluster analysis], it should be noted that a variable not containing any relevant information is worse than useless, because it will make the clustering less apparent. The occurrence of several such 'trash variables' will kill the whole clustering because they yield a lot of random terms in distances, thereby hiding the useful information provided by the other variables.*

Therefore, pre-processing is a fundamental step and it includes:

- features selection according to Kaufman and Rousseauw, 1990;
- features (variables) standardization/normalization if all variables a priori have the same importance;
- objects standardization/normalization if all objects a priori have the same importance.

Standardization/normalization can be performed according to the type of clustering algorithm chosen.

2.2 Clustering analysis

To group together objects that are similar in cluster, three different steps are necessary:

- ***Distance metrics or similarity measures***, which define the meaning of similarity/dissimilarity among points/objects;
- ***Clustering algorithm***, which defines the procedure to minimize distance of objects within groups and/or maximize the distance between groups using the distance metric chosen. According to the type of algorithm, it considers all possible clustering solutions and selects the one that has best inter- and intra-cluster distance metric. From this process, we can obtain $k^n/k!$ possible clustering solution, where *k* is the number of clusters and *n* the number of points/objects.
- ***Clustering evaluation or validation***, which assesses the goodness of fit of clustering solution.

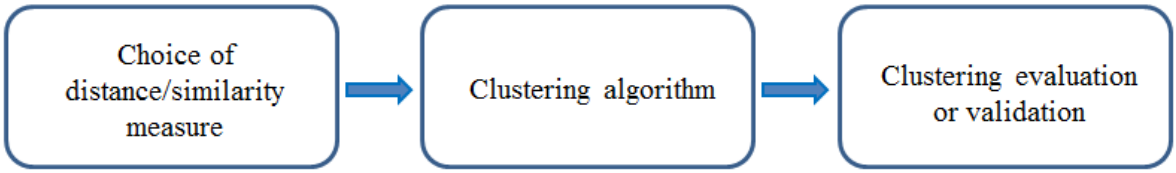


Figure 2.3 Clustering analysis steps.

2.2.1 Distance metrics and similarity measures

Since clustering is the grouping of similar instances/objects, some sort of measure, which can determines whether two objects are similar or dissimilar, is required. The distance measure, and consecutively the type of clustering used, depends on the ***attribute*** type and ***scale*** of the data. Usually we can have three different types of attributes (binary, discrete, continuous), while the common data scales can be more (qualitative, such as nominal or ordinal, or quantitative, such as interval with a unit of measurement or ratio).

There are two main types of measures used to estimate this relation: distance measures and similarity measures.

Among distance metrics there are:

- ***Minkowski*** (Han & Kamber, 2001), for numeric attributes:

$$d(x_i, x_j) = \left(|x_{i1} - x_{j1}|^g + |x_{i2} - x_{j2}|^g + \cdots |x_{iv} - x_{jv}|^g\right)^{1/g}$$

  where, *v* is the number of variables, $d(x_i, x_j)$ is the distance between two instances x_i and x_j. Given $g = 2$, we obtain the commonly used *Euclidean distance* between two objects, $g = 1$ the *Manhattan metric or city block metric*, and $g = \infty$ the *Chebychev metric*. The measurement unit used can affect the clustering analysis. To avoid the dependence on the choice of measurement units, the data should be standardized.
- ***contingency table*** for binary attributes: *simple matching coefficient* assesses dissimilarity between two objects, *Jaccard coefficient*, when positive or negative match are not equally important.

An alternative concept to that of the distance is the similarity function $s(x_i, x_j)$ that compares the two vectors x_i and x_j (Duda et al., 2001). Among these functions there are:

- Cosine measure: the normalized inner product between vectors when the angle between the two vectors is a meaningful measure;
- Pearson correlation measure;
- Dice coefficient.

The choice of distance/similarity measure should be based on the particular application and variables.

2.2.2 Clustering algorithms

The traditional algorithms for clustering can be divided in three main categories:

- Partitional Clustering;
- Hierarchical Clustering;
- Model-based Clustering (density-based).

In each clustering procedure, test the repeatability of results through different iterations of clustering algorithm, is necessary to avoid lack of inconsistency of a single iteration of the algorithm.

2.2.2.1 Partitional clustering

Partitional clustering aims to obtain directly a single partition of the dataset of objects into clusters. Such methods typically require that the number of clusters will be pre-set by the user and many of them are based on the iterative global optimization of ***objective function***.
Given the elements $x_i \in x$ with $i = 1,\ldots, N$ and a set of clusters Cj with $j = 1,\ldots, K$, in the optimization of objective function, the clustering problem consists in assigning each element x_i to a cluster Cj such that the intra-cluster distance is minimized and the inter-cluster distance is maximized.

K-means

Basic scheme:
Input: S (dataset), K (number of clusters)

1. Set K as the hypnotized number of clusters
   ➔ select K representative points in space, called centroids
2. Compute the ***distance metric*** of each point from all centroids
3. Assign all data points to the centroid with the minimum distance metric
4. Recompute the centroid for each cluster as the mean of the points belonging to each cluster
5. ***Repeat*** from 2. ***until*** centroid do not change and all data points are assigned to the same cluster with respect to the previous iteration

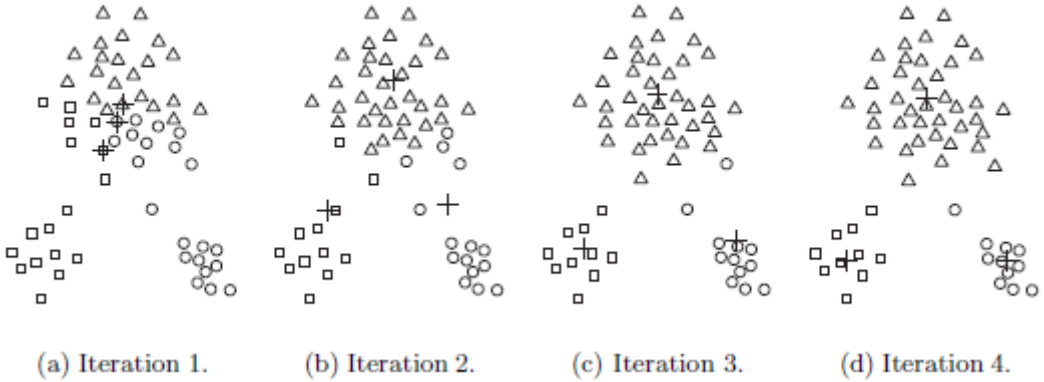


Figure 2.4 K-means with K= 3 (Source: Tan et al., 2006).

The initialisation of centroid are fundamental for the result obtained with this type of clustering algorithm. This process is influenced by these important characteristics:

- a different centroid initialization might produce a different clustering;
- different runs of the algorithm could produce different memberships of the input pattern.

In each clustering procedure, test the repeatability of results through different iterations of clustering algorithm, is necessary to avoid lack of inconsistency of a single iteration of the algorithm. This aspect is particularly important for K-means algorithm in which result is strictly linked to centroid initialization.

The complexity *O* of *I* iterations of the K-means algorithm performed on a sample size of *N* elements, each characterized by *V* variables, is: *O (I*N*V)* (Selim & Ismail, 1984).

According to these characteristics, it has the following advantages:

- the linear complexity: even if the number of instances is substantially large, this algorithm is computationally fast and with high speed of convergence;
- simple implementation.

Drawbacks of K-means are:

- cluster number, K, must be determined beforehand. The algorithm is very sensitive to this selection, which may make the difference between global and local minimum.
- it is not as versatile with clusters that differ for size, density and non-globular shape.
- it is sensitive to noisy data and outliers (a single outlier can increase the squared error dramatically) due to a weakness of arithmetic mean.
- it is sensitive to initial condition (centroid initialization), since different initial condition may lead to different result of cluster and identify a local optimum.
- it is difficult to determine the contribution that each variable makes to the grouping process, since it is assumed that each attribute has the same weight.

Therefore, considering the weaknesses of this algorithm some additional steps are necessary:

- in pre-processing: normalization the data and exclusion of outliers;
- in post-processing: exclusion of small clusters that may represent outliers and merge clusters with low SSE (close each other).

2.2.2.2 Hierarchical clustering

Hierarchical clustering aims to obtain hierarchy of clusters, forcing the data points into a strict hierarchy of nested subsets (Lance & Williams, 1967). The results of a hierarchical method can be displayed in a tree diagram, known as ***dendrogram***, representing the nested grouping of objects and similarity levels at which groupings change. A clustering of the data objects is obtained by cutting the dendrogram at the desired level.

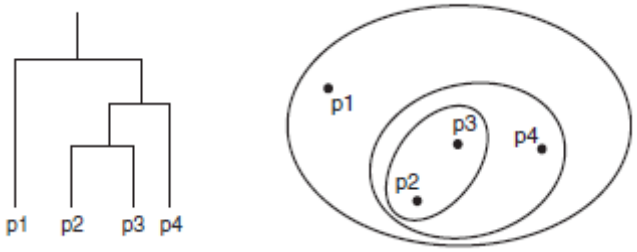


Figure 2.5 Dendrogram and nested cluster representation (Source: Tan et al., 2006).

In comparison to K-means (a partitional clustering method), hierarchical clustering does not require an initial value for *K*, so the data are not partitioned into a priori number of clusters. This can be considered an advantage. Moreover, the representation with dendrogram provides a meaningful taxonomies of dataset. In addition, hierarchical clustering cannot be viewed as a global optimization problem through an objective function, but it uses various criteria to decide locally, at each step, which cluster should be merged or split.

There are two well-known types of hierarchical clustering methods:

- ***Bottom-Up*** (***agglomerative***): Starting with each item in its own cluster, finds the best pair to merge into a new cluster. Repeat until all clusters are fused together.
- ***Top-Down*** (***divisive***): Starting with all the data in a single cluster, considers every possible way to divide the cluster into two. Choose the best division and recursively operate on both sides.

Divisive hierarchical clustering benefits from complete information about the global distribution when making top-level partitioning decisions, while agglomerative hierarchical clustering methods

make clustering decisions based on local patterns without initially taking into account the global distribution.

Both clustering algorithms use a ***similarity measure*** to merge or split each cluster in the process. The choice of similarity measure between cluster (inter-cluster similarity) may determines different results in agglomerative clustering.

There are different type of way to calculate inter-cluster similarity:

- ***Single linkage*** (minimum method or the nearest neighbor method): distance between the closest points in the different clusters; this two point can be linked by one link in the dendrogram (Sneath & Sokal, 1973);
- ***Complete linkage*** (the maximum method or the furthest neighbor method): distance between the most distant points in the different clusters (King, 1967);
- ***Central linkage***: distance of centers (centroids);
- ***Average linkage*** (minimum variance method): average distance of all pairwise proximity between points in the two clusters (Ward, 1963; Murtagh, 1984).

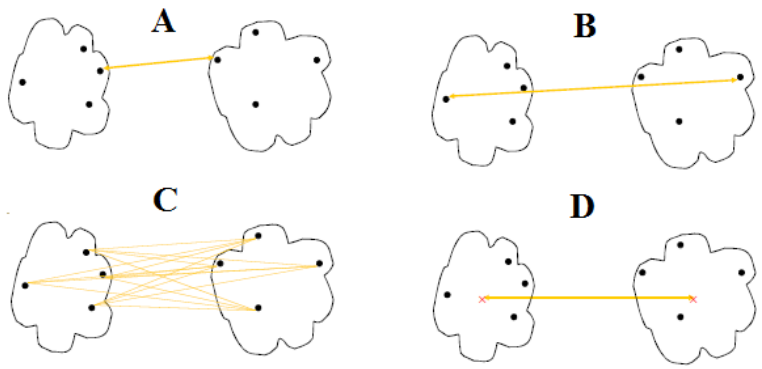


Figure 2.6 Inter-cluster similarity (Source: Tan et al., 2006).

The single linkage can handle non-isotropic cluster shapes, including well-separated, chain-like and concentric clusters, but is sensitive to noise and outliers. Compared to single linkage, complete linkage is less susceptible to noise and outliers, but tends to break large cluster and it is biased towards globular cluster shapes. Average linkage is a good compromise between single and complete linkage: it is less susceptible to noise, but it is biased towards globular shapes and is the most computationally demanding.

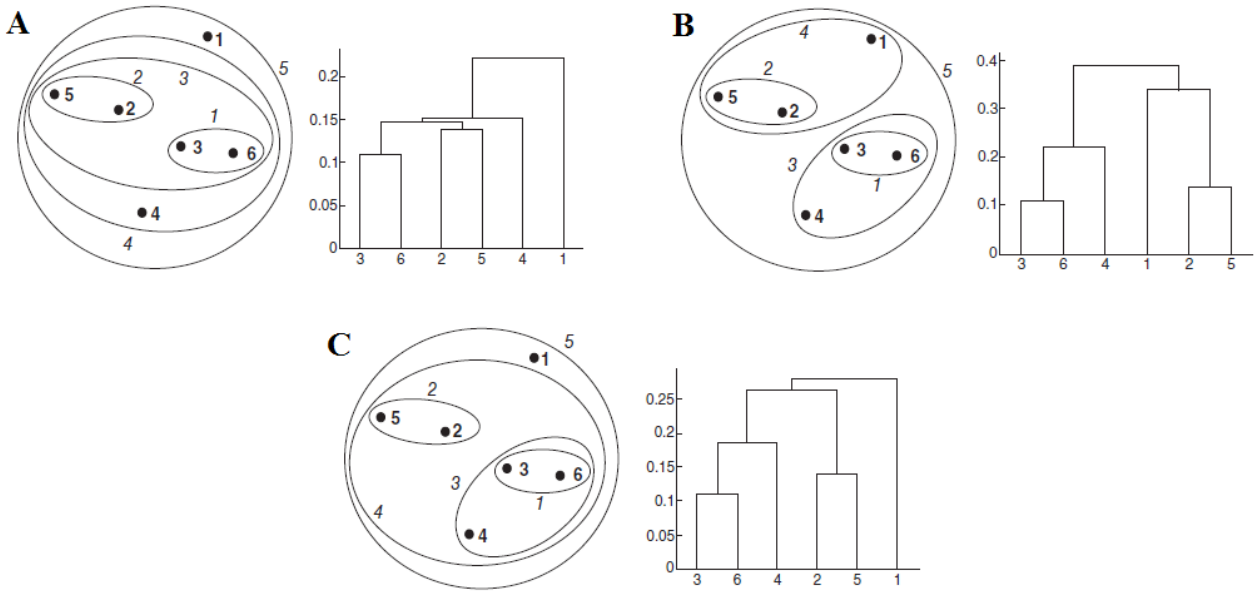


Figure 2.7 Different clustering results with different inter-cluster similarity measure (Source: Tan et al., 2006).

A re-iterations of the clustering and a comparison of the results is recommended to support the results.

Strengths of hierarchical methods are:

- versatility: the different similarity measures maintain good performance on different data sets on the basis of the specific requirement;
- visual hierarchical representation (dendrogram);
- multiple partitions: hierarchical methods produce not one partition, but multiple nested partitions, which allow to choose different partitions, according to the desired similarity level, by cutting the dendrogram at the desired level.

However, the main disadvantages of the hierarchical methods are:

- time consuming, storage requirement: the time complexity of hierarchical algorithms is non-linear, at least $O(I^3)$, because there are I steps and at each step the size, I^2, of the proximity matrix containing the similarity measure must be updated. Space and time complexity of hierarchical clustering severely limits the size of dataset that can be processed.
- hierarchical methods can never undo what was done previously. Namely there is no back-tracking capability.

Agglomerative

This method starts with as many clusters as data points and in each successive iteration, it agglomerates the closest two clusters that satisfy a predefined distance/similarity function.

Clusters are successively merged until only one cluster remains.

Basic scheme:

Input: *S* (dataset)

1. Calculate the distance metric between all data points
   ➔ Definition of a ***proximity matrix***
2. Each data point is considered a cluster
3. Calculate the similarity between all clusters (inter-cluster similarity)
   ➔ Update of a ***proximity matrix***
4. Merge the two most similar clusters into a higher level cluster
5. ***Repeat*** steps 3. and 4. for the new high-level clusters ***until*** only a single cluster remains.

The agglomerative process of object merging is displayed with dendrogram.

Divisive

This divisive clustering method is less commonly used. It works in a similar way to the agglomerative clustering but backwards. It starts with only one big cluster formed by all data points and, in each successive iteration, groups are continually divided until there are as many clusters as objects.

2.2.2.3 Density-based clustering

Density-based clustering aims to create clusters based on density functions. The main advantage of these algorithms is to create arbitrary shaped and sized clusters. However, one of it weakness is the difficult in identification of clusters with different density.

One of this algorithm is the DBSCAN. Density is estimated by counting the number of points within a specified *radius*. The choice of radius is important for this type of clustering.

2.3 Clustering evaluation or validation

The process of evaluating the results of a clustering algorithm is called cluster validity assessment. It is a fundamental part of the clustering process (Kovacs et al., 2005).

During cluster validation different aspects can be evaluated:

a) determining the clustering tendency of a set of data, i.e., distinguishing whether non-random structure actually exists in the data;

b) comparing the results of a cluster analysis to *externally* known results, e.g., class labels (*goodness of fit of clustering results*);
c) evaluating how well the results of a cluster analysis fit the data without reference to external information, using *only the data* (*goodness of fit of clustering results*);
d) comparing the results of two or more different number of clusters to determine which is better for selection of the *optimal number of clusters*.
e) comparing the results of two or more different clustering algorithms.

For b), c), and d), we can further distinguish whether we want to evaluate the entire clustering or just individual clusters.

Numerical measures (indices) that are applied for the evaluation of these various aspects of cluster validity, are classified into the following three categories (Theodoridis, 1999):

- ***External Criteria***: we evaluate the results of a clustering algorithm based on a pre-specified structure, which is imposed on a dataset, i.e. external information that is not contained in the dataset (external class labels);
- ***Internal Criteria***: measure the goodness of a clustering structure without respect to external information, but using intrinsic information of dataset;
- ***Relative Criteria***: compare two different clustering or clusters through internal or external indices.

Usually, internal criteria are considered more precise in the application to real dataset (Rendón, et al., 2011). Several validity indices have been developed and introduced, but most of them are based on the evaluation of (Berry & Linoff, 1996):

- ***Compactness (intra-cluster distance)***: the member of each cluster should be as close as possible to members of the same cluster. For example, compactness can be measured thought minimum, maximum or average distance between all pairs of point within cluster or between centroid and each point of the cluster.
- ***Separation (inter-cluster distance)***: the distance between any two cluster are maximized. The most common approaches measuring the distance between two different clusters are the same described before for inter-cluster similarity.

2.3.1 External Criteria

Purity

Purity and entropy are concept very similar. For a cluster *j*, the purity is

$$P_j = \frac{1}{n_j} max_j(n_{i,j})$$

where n_j is the number of objects in cluster j, $n_{i,j}$ is the number of objects in cluster j with class label i. The overall purity of the clustering solution is obtained as a weighted sum of the individual cluster purities and given as: $P_{tot} = \sum_{j=1}^{m} \frac{n_j}{n} p_j$

where m is the number of clusters, and n is the total number of objects.

Entropy

Entropy measures the purity of the clusters class labels. Thus, if all clusters consist of objects with only a single class label, the entropy is 0. However, as the class labels of objects in a cluster become more varied, the entropy increases. In each cluster the class distribution of the objects is:

$$E_j = \sum_i p_{ij} log(p_{ij})$$

The total entropy for the clustering solution is calculated as the weighted sum of the entropies of all clusters: $E_{tot} = \sum_{j=1}^{m} \frac{n_j}{n} E_j$

F-measure

F-measure combines the precision *pr* and recall *re*. For each cluster *j*, recall and precision of that cluster for each class label *i* are:

$$re(i,j) = n_{ij}/n_i \quad \text{and} \quad pr(i,j) = n_{ij}/n_j$$

where n_i is the number of objects in class i. The F-measure of cluster j and class i is given by the following equation:

$$F(i,j) = \frac{2 * re(i,j) * pr(i,j)}{re(i,j) + pr(i,j)}$$

$F(i,j)$ values are within the interval [0,1] and larger values indicate higher clustering quality.

2.3.2 Internal Criteria

Most internal indices are used as measure to assess intra-cluster similarity in comparison to inter-cluster dissimilarity (Kovács et al., 2005).

Cluster cohesion and Cluster separation: SSE

Sum of Squared Error (***SSE***) is a cluster validity measurement used for ***K-means clustering*** to assess the cohesion of cluster.

In each cluster the SSE is:

$$SSE_j = \sum_{i=1}^{n_j} dist^2(c_j, x_i)$$

where c_j is the centroid (or the most representative point) for cluster j.

The within cluster sum of square (***SSW***) for the clustering solution is:

$$SSW = \sum_{j=1}^{k} SSE_j$$

where k is the number of cluster. These measures can be used in an overall evaluation of a group of clusters.

A measure of separation between clusters in K-means is the between cluster sum of square (***SSB***), which is the sum of the squared distance of a cluster centroid, c_j, to the overall mean, c, of all the data points.

$$SSB = \sum_{j=1}^{k} m_j \, dist^2(c_j, c)$$

where m is a scale factor.

A good clustering solution is considered the one that minimizing the SSW and, equivalently, maximizing the SSB.

Silhouette

Silhouette index, normalized between -1 and 1, is obtained for each single element of dataset as ratio of a element's dissimilarity to its own cluster and to its nearest neighboring cluster. To evaluate the clustering solution, we can use the average of the silhouette scores of all the individual element of datasets. If the value is closer to 1, the clusters are well separated, if it is negative, the result of clustering does not reflect the correct separation of elements. Moreover, the Silhouette clustering evaluation criterion can be applied to any distance metric (Kaufman & Rousseeuw, 1990).

The Silhouette value for the entire dataset is defined as:

$$Sh = \frac{1}{k} \sum_{i=1}^{k} \frac{\left[\mathrm{D}_{i,j} - d_i\right]}{max_{j \neq i}\{d_i, D_{i,j}\}}$$

where k is the number of clusters, d_i (intra cluster similarity) is the average distance between each single point of the i-th cluster and the other points in the same i-th cluster, and $D_{i,j}$ (inter cluster similarity) is the maximum average distance between each single point of i-th cluster and points in j-th cluster (with $j = 1, \ldots, k; j \neq i$).

Davies-Bouldin index (DB)

Davies-Bouldin index is a non-negative value. It is a function of the ratio of the within cluster cohesion and between cluster separation: a lower value is associated to a better clustering result (Davies & Bouldin, 1979). By minimizing this index, clusters are the most distinct from each other, and therefore achieves the best partition. The Davies-Bouldin index is defined as:

$$DB = \frac{1}{k}\sum_{i=1}^{k} max_{j \neq i}\left[\frac{\left(d_i + d_j\right)}{D_{i,j}}\right]$$

with *j* comprised between 1 to *k*, where *k* is the number of clusters, d_i (intra cluster similarity) is the average distance between each point of *i*-th cluster and the centroid of *i*-th cluster, $D_{i,j}$ (inter cluster similarity) is the distance between the centroids of the *i*-th and *j*-th clusters. The maximum value of $D_{i,j}$ represents the worst-case of intra-to-inter cluster ratio for *i*-th cluster.

Cophenetic correlation coefficient (CPCC)

CPCC, ranging from 0 to 1, is a cluster validity measurements obtained for ***hierarchical clustering***. In a hierarchical cluster tree, any two objects in the original data set can be linked together at some level. The height of the link, also known as cophenetic distance, represents the distance between the two sub-clusters that contain those two objects. CPCC compares the cophenetic distances with the original distance between objects and gives a measure of how dendrogram preserves the pairwise distances between the original data points. The closer the value of the CPCC is to 1, the more accurately the clustering solution reflects your data (Saraçli et al., 2013).

Considering the original dataset *X* and the cophenetic distance matrix *C*, $x_{ij} = \left|X_i - X_j\right|$ is the distance, based on the metric choosen for clustering, between the *i*-th and *j*-th observations. c_{ij} is the height of the link and corresponds to the distance between the two sub-clusters that contain the *i*-th and *j*-th observations. $\bar{x}$ and $\bar{c}$ are the average of *X* and *C*, respectively.

$$CPCC = \frac{\sum_{i<j}\left(x_{ij} - \bar{x}\right)\left(c_{ij} - \bar{c}\right)}{\sqrt{\sum_{i<j}\left(x_{ij} - \bar{x}\right)^2 \sum_{i<j}\left(c_{ij} - \bar{c}\right)^2}}$$

Previous studies demonstrate that dataset with outliers have higher CPCC values than the data set without outliers (Johnson & Wichern, 2002; Saraçli et al., 2013).

2.4 Conclusion

Different studies evaluate the performance of these indices on different datasets (Rendon et al., 2011, Liu et al., 2010). Moreover, they reported that internal indexes are more accurate in group determining in a given clustering structure. Among internal indices, Slobodan (2006) reported that Silhouette index produces slightly more accurate results than the system that uses the Davies-Bouldin index. However, the computation of the Davies-Bouldin index is much less complex than the computation of Silhouette. For subsequent analyses, we choose both indices for a robust performance (Vendramin et al., 2010).

In Chapter 4, internal indices were used to evaluate two different aspects of clustering solution: selection of the optimal number of clusters and goodness of fit of clustering results.

Silhouette and Davies-Bouldin indices were used to assess within cluster similarity in comparison to between cluster dissimilarity and select the optimal number of clusters in dataset (Kovács, et al., 2005) by comparing the results obtained with different *k* number of clusters. Meanwhile, CPCC, only for hierarchical clustering, gave a measure of how accurately the clustering solution of dendrogram reflected original data.

References

Berry M. J. A. and Linoff G. (1996). Data Mining Techniques for Marketing, Sales and Customer Support, John Wiley & Sons, Inc.

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2), 224–227.

Duda, R., Hart, P., & Stork, D. (2001). Pattern Classification. John Wiley and Sons Inc.

Han J. & Kamber M. (2006). Data Mining: Concepts and Techniques, 2nd ed.Morgan Kaufmann Publishers, March 2006. ISBN 1-55860-901-6

Johnson, R. A., & Wichern, D. W. (2002). Applied Multivariate Statistical Analysis (5th editio). Prentice Hall, New York.

Kaufman L., Rousseauw P.J. (1990). Finding groups in data. An introduction to cluster analysis, Wiley, New York.

King, B. (1967). Step-wise Clustering Procedures, J. Am. Stat. Assoc. 69, pp. 86-101.

Kovács, F., Legány, C., & Babos, A. (2005). Cluster Validity Measurement Techniques. In Proceedings of the 6th International Symposium of Hungarian Researchers on Computational Intelligence, Budapest (pp. 18–19).

Lance G. N. and Williams W. T. (1967). A General Theory of Classificatory Sorting Strategies. Hierarchical Systems The Computer Journal; 9 (4): 373-380.

Liu Y , Li Z, Xiong H, Gao X, Wu J. (2010). Understanding of Internal Clustering Validation Measures, Proceedings of the 2010 IEEE International Conference on Data Mining, p.911-916.

Lletı R, Ortiz MC, Sarabia LA, Sánchez MS. (2004). Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. Analytica Chimica Acta 515:87–100.

Murtagh, F. (1984). A survey of recent advances in hierarchical clustering algorithms which use cluster centers. Comput. J. 26 354-359.

Rendón, E., Abundez, I., Arizmendi, A., & Quiroz, E. M. (2011). Internal versus External cluster validation indexes. International Journal of Computers and Communications, 5(1).

Saraçli, S., Dogan, N., & Dogan, I. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. Journal of Inequalities and Applications, 2013(203), 1–8.

Selim SZ, Ismail MA. (1984). K-means-type algorithms: a generalized convergence theorem and characterization of local optimality. IEEE Trans Pattern Anal Mach Intell;6(1):81-7.

Slobodan P. (2006). A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters. In Proceedings of the 11th Nordic Workshop on Secure IT-systems, pages 53–64.

Sneath, P., and Sokal, R. (1973). Numerical Taxonomy. W.H. Freeman Co., San Francisco, CA.

Steinbach M., Ertoz L. and Kumar V. (2003). The Challenges of Clustering High Dimensional Data.

Tan, P.N, Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. In Introduction to Data Mining (2006th ed., pp. 447–568). McGraw Hill.

Theodoridis S. and Koutroubas K. (1999). Pattern Recognition, Academic Press.

Veyssieres, M.P. and Plant, R.E. (1998). Identification of vegetation state and transition domains in California's hardwood rangelands. University of California.

Vendramin L, Campello R J. G. B. and Hruschka E R. (2010)Relative Clustering Validity Criteria: A Comparative Overview. Statistical Analysis and Data Mining: The ASA Data Science Journal Volume 3, Issue 4.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58:236-244.

Chapter 3

Quality data assessment and advanced data pre-processing

3.1 Background

Minimizing image noise is extremely critical in fMRI, because the signal changes related to the BOLD contrast are very small (1-2%) compared to the signal recorded. It is therefore crucial to separate and clarify the contributions of the various sources of noise from the scanner and the subject (see Chapter 1). While the noise from a subject will be inevitable intrinsic during the acquisition, we can attempt to control the noise from scanner (both *background noise* and *scanner instability*) through an appropriate periodic quality assurance protocol on phantom and acquisition parameters. However, despite careful control of the various noise sources some of them persist and different methods of image processing have been developed. During pre-processing steps, various image and signal processing techniques are applied to increase SNR in raw EPI images and to reduce *background and physiological noise*. These steps are crucial to ensure reliability and reproducibility of results of subsequent statistical analyses and greatly improve the statistical power of them.

3.2 Standard Data pre-processing

Usually, principal steps of ***standard pre-processing*** include (Friston et al., 1995):

1. slice-time correction (STC);
2. motion correction (realignment);
3. normalization to standard template;
4. spatial smoothing.

Slice time correction (Henson et al., 1999) minimize the differences due to temporal shift in sampling between the first and the last slice of the volume during a single TR. It consists in shifting the signal phase to a reference slice. Usually, the slice acquired in the middle of the sequence is used as a reference slice, because it correspond to maximum interpolation of TR/2.

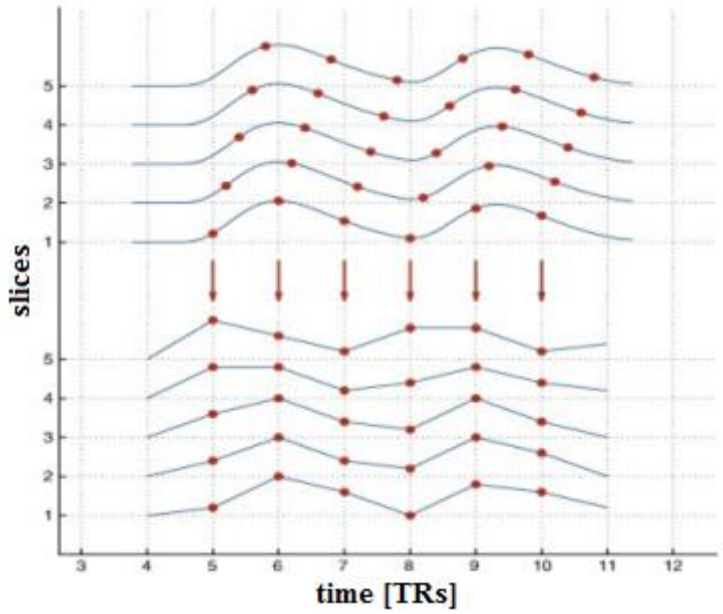


Figure 3.1 Effect of slice timing among different slices. (Source: Sladky et al., 2011).

Motion correction (realignment) (Hajanal et al., 1994). Mismatch between subsequent volumes in the time-series for head position, such as swallowing, breathing or little head adjustment, can introduce spurious BOLD signal changes and reduce the sensitivity of subsequent analyses. Realignment algorithms can reduce motion induce artefact through a rigid body transformation that minimises a cost function (define the difference between 2 images), as e.g. least square (SPM) or normalized correlation ratio (FSL). Rigid body transformation is defined by 3 translations in *x*, *y* and *z* directions and 3 rotations around the *x* (pitch), *y* (roll) & *z* (yaw) axes.

Normalization (Holden, 2008). The differences among subjects in head shape and size do not permit to compare them. Normalization procedure tries to minimize these differences with a transformation of single subject image in a standard space. Usually, associated to an EPI scan, a high resolution anatomical scan (T1 or T2 –weighted images, voxel size around 1 mm^3) is acquired. This high resolution anatomical scan is pre-processed, using noise reduction, bias correction for correction of broad intensity variation among the different brain regions, brain extraction to remove non brain tissues, and segmentation of grey matter, white matter and cerebrospinal fluid. In

addition, a common standard space is selected: Talairach (Talairach & Touroux, 1988) or the MNI space (Evans et al., 1992), obtained as average image of 152 T1-weighted images of different subjects. During normalization there are 3 steps:

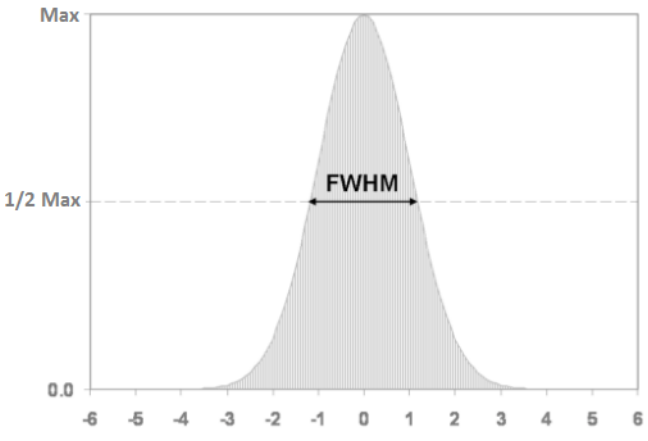
i) co-registration between T1-weighted and EPI images (fMRI),
ii) transformation of T1-weighted scan to match the template (i.e. normalization), and
iii) application of the normalization parameters obtained from ii) to the EPI images.

The ii) step of spatial normalisation uses a 12 parameter affine transformation (rotation and translation for a rigid body transformation, scaling and shears for change in size and shape) and a subsequent non-linear deformation. The data are resampled to a new spatial resolution, usually 2 mm isostropic voxel.

Spatial smoothing. Spatial smoothing is used to

- increase SNR by removing high frequency information and increasing sensitivity,
- improve inter-subject averaging (Poldrack et al., 2011) by reducing the anatomical variability among subject not fully corrected with normalization, and
- improve validity of the statistical tests making distribution of error term normal.

Smoothing is performed applying a Gaussian filter (also known as a kernel) to the pre-processed EPI image: the signal of each voxel being replaced by the weighted average of its neighbours (Ashby, 2011). It is the shape of the kernel that defines the weights that is applied to each voxel. A Gaussian kernel follows essentially a normal distribution: weight is always strongest in the centre of the kernel (i.e. the voxel being smoothed) and decreases with distance at a rate that depends on the width of the distribution. Wider distributions (wide kernel) result in greater smoothing, because distant voxels obtains higher weights. Usually, the shape of a smoothing kernel is usually described by the width of the distribution at half of its maximum value: the Full Width Half Maximum (FWHM).



However, the specifics of the width of kernel and the order in the pre-processing pipeline for smoothing step are critical, because it can impact on images with some drawbacks:

- reduction of spatial resolution of the data,
- edge artifacts, because brain voxels are smoothed with non-brain voxels, resulting in a dark ring around the brain,
- merging of activation peaks that are less than twice the FWHM or attenuation of small meaningful activation under statistical threshold;
- mis-localization of activation peaks (Mikl et al., 2008).

The width of smoothing kernel applied should be the minimum necessary to obtain a good compromise among these aspects, in particular avoiding smoothing away very small signals, or smoothing two independent signals into one (Ashby, 2011). As a general rule, a FWHM of twice the voxel dimension is a good compromise (Poldrack et al., 2011).

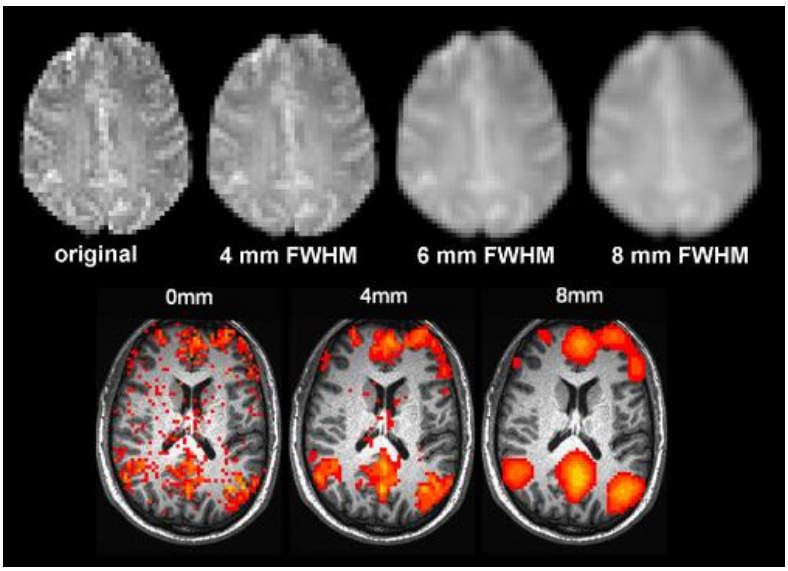


Figure 3.2 Smoothing effect of different FWHM values (Source: Poldrack et al., 2011).

Furthermore, Jo and others (2010) propose that smoothing should be restricted to the brain's gray matter mask to reduce the inclusion of unwanted BOLD and other physiological signals from surrounding large draining vessels.

3.3 Advanced data pre-processing

Subject-related noise during fMRI scanning introduces measurement inaccuracies as imaging voxels (Chapter 1). In particular, primary effects of participant head motion in fMRI data are corrected by realignment step. However, head motion induces secondary effects related to partial voluming, interpolation effects, magnetic field inhomogeneities, intra-volume motion, and spin-history effects (Friston et al., 1996), which cannot be corrected using realignment.

In literature, different indices were proposed to evaluate the subject-related noise impact on fMRI scan directly from images or indirectly from realignment parameters, and subsequent different approaches were used to reduce these sources of noise on fMRI data.

3.3.1 Indices of data quality assessment

In literature, different indices were proposed to assess the quality of fMRI data. Two different levels of evaluation are considered: single-subject scan in an investigated sample and volumes in a single-subject scan. In the former case, an assessment of the quality of single-subject scan is performed to evaluate if run can included because movements or artefacts are comprised in a range of tolerance that could be corrected with pre-processing steps. In the latter, an assessment of the volumes in a single-subject is performed to identify those corrupted by high movements and that need to be discard from the subsequent analyses.
For both levels, two different categories of indices were proposed: the first is based on ***realignment parameters*** and the second is computed on ***EPI images*** directly.

3.3.1.1 Indices derived from realignment parameters

In the first category, these indices are obtained directly from realignment parameters (*RP*), which are considered displacements relative to a single reference volume (***absolute displacements***), or from first derivative of realignment parameters (*RP d/dt*), which are obtained as displacements relative to the preceding volume (***relative displacements***) (Power et al., 2015).
As described from (Power et al., 2014), three principal patterns of movement in realignment parameters can be detected: still subject, subject who moved intermittently and returned to their original position, and subject who moved and remained displaced from the origin. These different patterns had not the same impact on fMRI signal: usually, the second pattern was considered more difficult to correct trough realignment procedure, if frequent high displacements occurred in less

than one TR. In all these case can be an additional drift in the realignment parameters that relative displacements minimize.

The most diffused metric for the quantification "single-subject scan motion" are based only on the inspection of ***absolute/relative displacements***, often applying a rule-of-thumb of "motion exceeding one voxel size" (Johnstone et al., 2006; Nemani et al., 2009; Churchill et al., 2012; Wilke et al., 2005). Different version of quality data indices using root mean square (RMS) are defined on absolute and relative displacements. The first formulation was proposed by Oakes and colleges (2005), as the average RMS of the percent signal difference between each volume and the first volume in translation (***average RMS of absolute displacements***). Wilke and colleges (2012) used the same approach, but considering translation and rotational parameters separately. Then, Power and collegues (2014) applied RMS on both absolute (***RMS of absolute displacements***) and relative displacements (***RMS of relative displacements***), but after a detrending of realignment estimates across all time points, that permits to reduce the influence of drift effect also in absolute displacements. Finally, Siegel and colleges (2014) obtained a summarized measure using RMS from both translational and rotational absolute realignment parameters, after a conversion of rotational estimates form degree to mm with a sphere of radius of 50 mm. Usually, RMS measure were used for the quantification of "single-subject scan motion" and the user-dependent threshold was defined in relation to the voxel size (e.g. half voxel's width).

Then, to evaluate the "volumes in a single-subject scan", other five metrics, based only on relative displacements, were proposed: mean motion, maximum motion, number of movements, rotation (VanDijk et al., 2012) and ***framewise displacement (FD)*** (Power et al., 2011). *Mean motion* is obtained as the mean of relative translational displacements, while maximum motion as the maximum of relative translational displacements. Similarly, the *number of movements* is defined as the number of volumes associated with a relative displacement > 0.1mm. *Rotation* is obtained as average of the absolute value of the Euler angle of relative displacements of rotation. Among them, the mean motion was used also for the evaluation of "single-subject scan motion" (VanDijk et al., 2012; Satterthwaite et al., 2012: threshold of 0.55 mm for the identification of gross motion). Finally, the FD has become popular and implemented in different pre-processing tools (e.g. FSL: fsl_motion_outliers; BRAMILA tools; AFNI: 3dDespike procedure). FD is obtained as the sum of relative displacements on the 6 realignment parameters, after the conversion of rotational displacements from degrees to millimetres (e.g. displacements on the surface of a sphere of radius 50 mm, which is approximately the mean distance from the cerebral cortex to the center of the head). Different thresholds were applied on the basis of investigate cohorts: 0.5 mm (Power et al.,

2012) 0.2 mm (Fair et al., 2013); 0.9 mm (Siegel et al., 2014). Moreover, the threshold depends also from the FD implementation (Power et al., 2015).

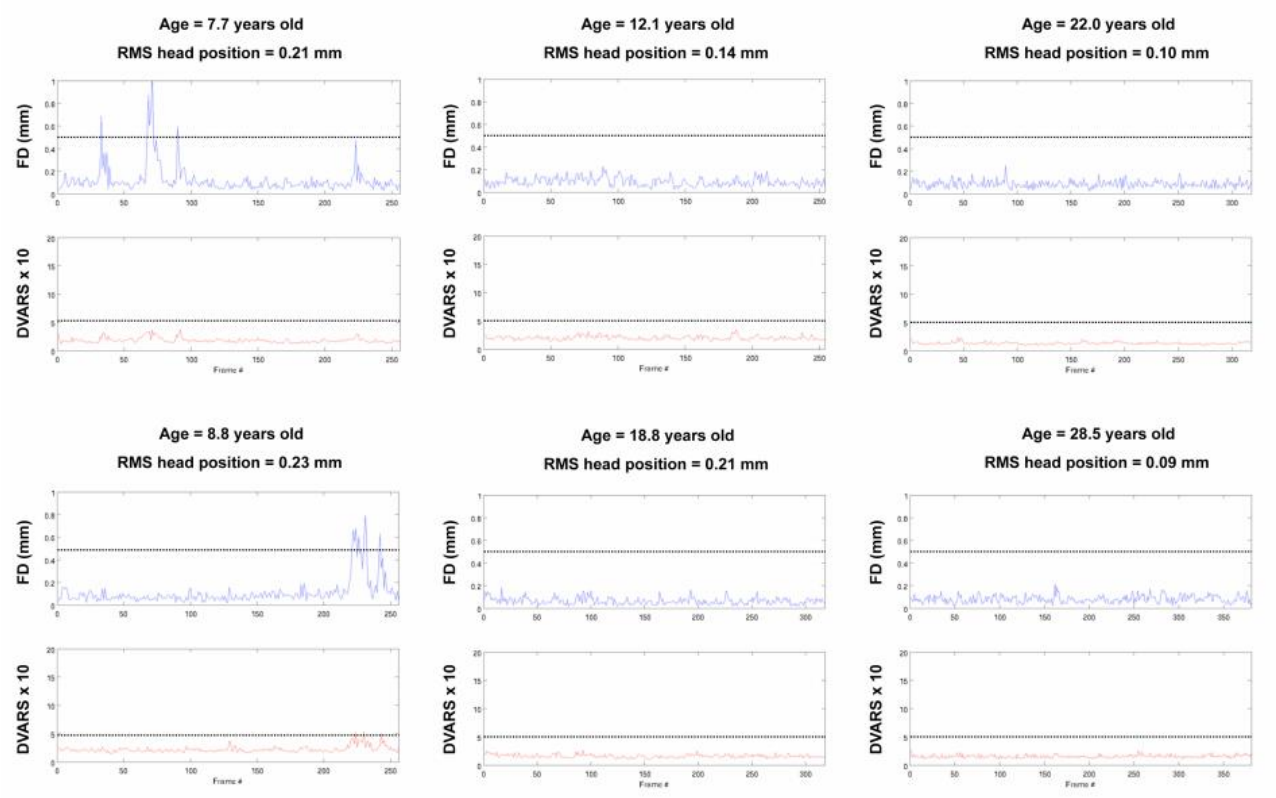


Figure 3.3 A comparison among FD, DVARS and mean RMS (Source: Power et al., 2012).

3.3.1.2 Indices derived from EPI images

In the second category, the derived indices are based on EPI images.

The first index is ***temporal SNR (tSNR)*** (Van Dijk et al., 2012). To obtain it, the mean signal across the EPI scan was calculated for each slice, and then the it was divided by the standard deviation of the signal intensity within the slice over time. The mean tSNR value across all voxels in an inclusive brain mask was derived as the measure of tSNR for the EPI data.

The second is the ***Derivative of rms VARiance over voxelS (DVARS)*** (D referring to temporal derivative of time-courses, VARS referring to RMS variance over voxels) (Smyser et al., 2010). It indexes the rate of change of BOLD signal across the entire brain for each volume. To calculate DVARS, the volumetric time-series is differentiated (by backwards differences - relative displacements) and RMS signal change is calculated over the whole brain for each volume. DVARS is a measure in time of how much the intensity of a brain image changes in comparison to the previous volume (as opposed to the global signal, which is the average value of a brain image at a volume). Commonly, used thresholds are comprised from 0.3 to 0.5% of BOLD. Importantly,

DVARS was applicable only to *resting state fMRI*, because in task-based fMRI it will indicate erroneously as possible artefacts signal changes related to paradigm.
When forming cohorts or identifying good volumes, the definition of a ***user-dependent threshold*** is necessary for all the above indices (based on realignment parameters or EPI-images): single-subject scans exceeding this threshold, for entire session or part of it, are discarded from the analyses, and similarly, volumes exceeding this threshold are excluded from analyses. The choice of these thresholds is strictly related to the type of population considered (Satterthwaite et al., 2013), the level of the assessment (single-subject scans or volumes), the length and sampling of the acquisition, and the amount of motion present. In fact, the more stringent the thresholds are, the more efficient is the removal of outlier subjects or volumes related to motion. As a drawback, stringent thresholds might increase the intrinsic variability and decrease the test-retest reliability of the data (Power et al., 2015). On the other side, permissive threshold could determine violation of assumptions and reduction of the power of subsequent statistical analyses.

Moreover, most often, these different metrics were summarized in a single value, using mean over time (Hallquist et al., 2013; Power et al., 2014). One of the drawback of summary statistics of motion is that they do not always distinguish between qualitatively different types of subject movement. If we analysed the head movements we can identify in two main categories: subject A who is perfectly still but moves suddenly once to reach a new head position (drift movements), and subject B who has frequent small to moderate movements. It is possible for these two subjects to have similar, or even identical, indices, despite the substantial qualitative differences in how they moved. Subject A will have data of acceptable quality throughout the scan except during and immediately after head motion, whereas the data of subject B will be somewhat compromised throughout much of the scan (Satterthwaite et al., 2013). This limit is particular evident when absolute displacements measures were used. Usually, metrics, derived from relative displacements highlights high movements (i.e. spike movements), while minimizes the effect of drift movements (Lemieux et al., 2014), that are detected in metrics derived by absolute displacements (Power et al., 2012).

3.3.2 Temporal filtering

Temporal filtering allow to

- improve the signal-to-noise ratio, as well as smoothing in the spatial domain, and, consequently, the power of statistical data analysis;

- assume stationary time courses.

To these aims, high-pass filters were used to remove low-frequency drifts from voxel time courses of fMRI data. This type of noise arises from *background noise* ("scanner drift": e.g., slowly-varying changes in ambient temperature) (Smith et al., 1999), from *physiological sources*, such as ~1Hz respiratory or ~0.25Hz cardiac cycles, that are aliased by the slower sampling rate, and from residual movement effects and their interaction with the static magnetic field (Turner et al., 1998).

Unfortunately, this pre-processing step is also one of the more dangerous ones, because condition-related signal changes may also be removed, if filter cut-off is not properly applied. It is important that the frequencies filtered out are below the frequencies contained in the stimulation task otherwise the experimentally induced signal fluctuations will be removed from the data or at least strongly reduced. From literature (Friston et al., 1995), a Gaussian filter of width 2.8 seconds, is a good approximation to the haemodynamic response function.

The use of low-pass filtering is much more controversial than high-pass filtering, in particular in some task designs also contain high frequency content in the model, e. g. dense event-related model.

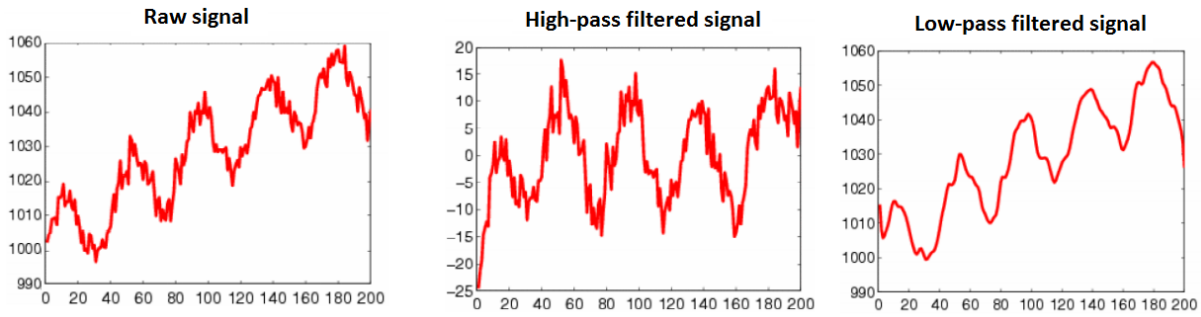


Figure 3.4 Examples of temporal filtering on time-courses.

In resting-state the use of temporal filtering is still debated. Correlation-based connectivity methods impose a temporal sample independence assumption. However, temporal filtering to address the high frequency noise may introduce sample dependence, violating this assumption (Davey et al., 2013).

3.3.3 Signal regression

Another approach to minimize secondary effects of head movements and impact of physiological noise is to regress out signals of no interest in GLM. Different signals can be considered: ***realignment parameters*** and derived measures, ***physiological signals***, and ***global signal***.
The regression of all these signals can be performed separately or simultaneously (Van Dijk et al., 2010).

3.3.3.1 Realignment parameters and derived measures

To reduce confounding effects of movement-related artefacts in fMRI time-series, the introduction of the 6 RP obtained by realignment step, in GLM as regressors of no interest, was proposed by Friston et al., 1996). This method has reported to explain 30-90% of the variance.
Subsequently, different temporally shifted versions of these waveforms (first temporal derivatives RP d/dt, squared values RP or RP d/dt) were proposed. For each of them different combination were considered:

- case 1: 12 motion related signals [RP RP d/dt];
- case 2: 24 motion-related signals [RP RP^2 RP d/dt RP^2 d/dt] (Satterthwaite et al., 2013; Yan et al., 2013; Power et al., 2014);
- case 3: 36 motion-related signals [RP RP^2 RP d/dt RP^2 d/dt RP d/dt^2 RP^2 d/dt^2] (Power et al., 2014).

However, there are different drawbacks for this method.

- First, this modelling of motion assumes that motion affects all brain regions equally, but in a real case the head motion is constrained by the head and neck anatomy. For example, rotations produce a pivot of the head around the neck, leading voxels at the edge of the brain to experience more motion than voxels near the pivot (Satterthwaite et al., 2013).
- Second, realignment parameters typically only model linear motion-induced signal variation while the underlying dynamics are non-linear (Fair et al., 2013), i.e. secondary effects of head motion are not necessarily captured.

3.3.3.2 Physiological signals

The heart and respiratory rate are not stationary across a typical time interval for fMRI scanning and shows small rate variations. Usually, heart rate fluctuations produce low-frequency contributions (< 0.1 Hz) that can affect resting-state networks, while respiratory rate at rest is around 0.2 Hz to 0.3 Hz. A low-pass filter will not be effective in removing signals faster than the Nyquist frequency

(equal to half of the sampling rate) and slower than the band-pass cut-off, which may be aliased into the retained frequency spectrum. Physiological noise, especially low-frequency components is a particular concern.

To reduce confounding effect due to physiological signals, two different approaches can be used. The most straightforward method is to acquire ***simultaneous*** physiological measurements during the scan (RETROICOR, Glover et al., 2000; Lund 2006), using chest straps for the respiratory (Birn et al., 2008) and a pulse oximeters for the cardiac rate (Chang et al., 2009).

If these measures are not available, physiological noise can be derived from EPI images, in particular from ***ventricles and white matter regions***. These regions are motivated by the fact that they contain a relatively high proportion of noise caused by the cardiac and respiratory cycles (Lund et al. 2006). Different approaches were proposed based on extraction of mean signal from region of interest (ROI) (Fox et al., 2005; Power et al., 2011, Satterthwaite et al., 2012) or from the segmentation of cerebrospinal fluid (CSF) and white matter (WM) (Hallquist et al., 2013, Weissenbacher et al., 2009). For ROIs extraction small sphere (radius around 4 mm) were centered in lateral ventricle, regions in deep cerebral white matter (Van Dijk et al., 2010), or in the anterior/posterior corpus callosum (Garrett et al., 2010). Moreover, to ensure that these ROIs does not contain any signal of interest, Garrett and others proposed to extract mean signal from unsmoothed fMRI data.

However, recent studies performed with MEG on influence of visceral signals (heart, stomach, …) on neural signal by the group of Tallon-Baudry highlighted a correspondence between fluctuation in neural responses and variation of heartbeats encoding self-relatedness of thoughts (Park et al., 2014, Babo-Rebelo et al., 2016). In particular, neural responses to heartbeats related to self-relatedness of thoughts occur in the default mode network (DMN). So, they arise the question related to the real meaning of physiological signal in resting state: noise or signal?

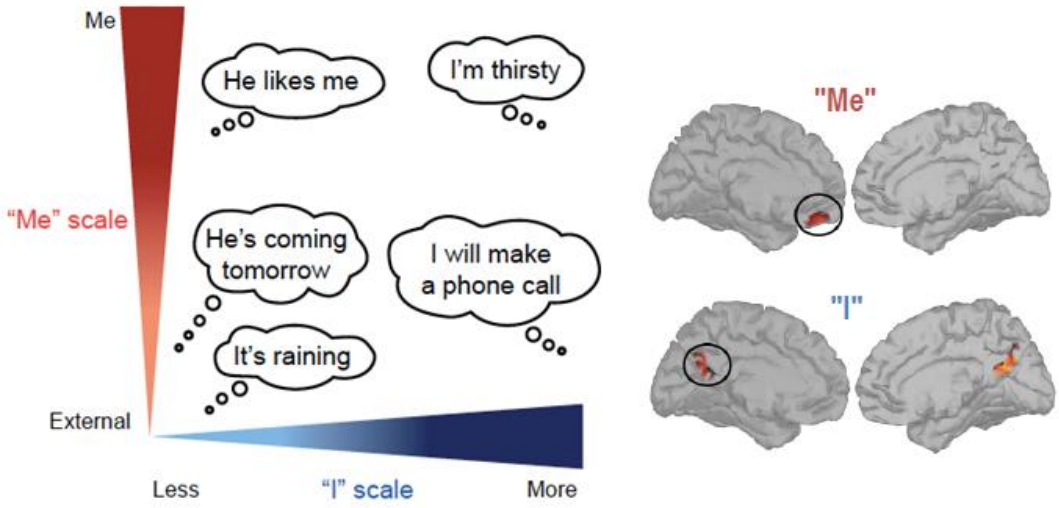


Figure 3.5 Neural responses to heartbeats encode self-relatedness of thoughts.
Experimental evidence for the distinction between "I" and "Me". Neural responses to heartbeats in: posterior cingulate cortex (PCC) encodes self-relatedness along the "I" dimension, while ventro-medial prefrontal cortex (vmPFC) encodes self-relatedness along the "Me" dimension (adapted from talk of C. Tallo-Baudry).

3.3.3.3 Global signal

Finally, to reduce general confounding effect, signal averaged over whole-brain (also referred to as ***global signal, "regression of the global signal," "global signal normalization," or "orthogonalization of the global signal"***) was obtained only for resting-state scan (Fox 2009, Power et al., 2011, Van Dijk et al., 2010). It has been found to correlate with the effects of carbon dioxide partial pressure variation (Chang et al., 2009). To regress it out, we assumes that global variations of BOLD signal and experimental conditions are uncorrelated: physiological sources will be distributed equally over time and over all in the most part of the voxels of the brain (although not necessarily at the same magnitude) (Macey et al., 2004). This assumption is not verified in task-based fMRI (Junghofer et al., 2005, Murphy et al., 2009) and its application can reduce sensitivity (Desjardins et al., 2001).

However, the application of this step is really debated, because, associated to the initial aim of noise removing, whole-brain signal regression can produce the emergence of robust negative correlations (Murphy et al., 2009; Weissenbacher et al., 2009). A reason for this effect is that the distribution of correlations is shifted (Murphy et al., 2009), because the method forces correlation strengths between a given source (e.g., a region of interest) and other voxels in the brain to be distributed around zero.

3.3.3.4 Independent Component Analysis (ICA)

To overcome this limit of the identification of motion and physiological signals, a data-drive approach on EPI images was proposed to identify and remove motion-related artefacts using Independent Component Analysis (ICA): ICA-based nuisance removal. ICA decomposes the data into a set of spatial independent component maps (ICs), and for each of them a corresponding time-courses was associated (McKeown et al., 1998; Beckmann & Smith, 2004). Among these ICs we can identify subsets of ICs corresponding to brain activity and subsets associated to noise (e.g. motion-related, physiological or scanner-induced noise). Time-courses of ICs representing noise can be regressed out from the fMRI data (Thomas et al., 2002; Kundu et al., 2012).
However, one of the big deal in this technique is the correct isolation and estimation of signals of interest form thus that are artefactual components. It can be performed manually or automatically. In this latter case, different methods have been proposed based on temporal and/or spatial features (i.e. CORSICA, PESTICA, ICA-AROMA) (Thomas et al., 2002; De Martino et al., 2007; Pruim et al., 2015).

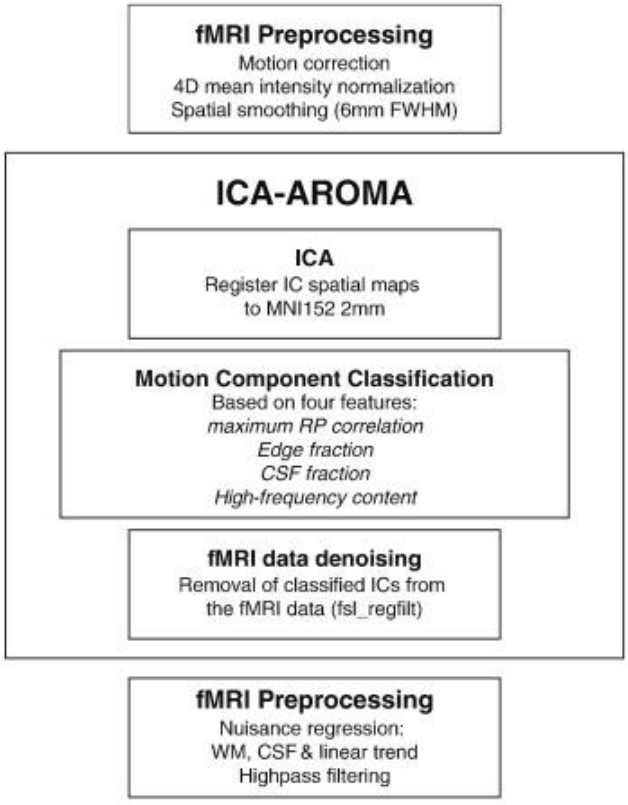


Figure 3.6 Example: Overview of the different steps in ICA-denoising embedded in fMRI pre-processing steam (Source: Pruim et al., 2015, ICA-AROMA)

However, overfitting of the data and removal of signal of interest can be some dis-vantages of the introduction of a large number of nuisance regressors (Yan et al., 2013; Satterthwaite et al., 2013).

3.3.4 Interpolation and Scrubbing

Two complementary strategies to regression of motion parameters are proposed:

- ***scrubbing*** (or "***censoring*** ") procedure: it is performed by removing the volumes during which significant movement occurred: in a first case, they can be regressed out ("spike regression"; Lemieux et al., 2007; Satterthwaite et al., 2013) or in a second case, they can be completely eliminated from the fMRI time-series (Power et al., 2012). In the first case, when a volume is identified as spike, it is classified as to discard in a binary regressor (or temporal mask). Therefore, that volume is censored from the residual time-series and replaced with a zero. Because the mean value of the residual time-series is also zero, such volumes do not influence the overall correlation value.
- ***interpolation***: when the volumes are not completely discarded, but they are replaced by interpolating values (i.e. from adjacent volumes, using cubic interpolation, Campbell et al., 2013).

To identify volumes contaminated by excessive motion, metric of quantification of spike movements are necessary (section 3.1): Power and colleagues (2011) proposed FD and DVARS measures (dual criteria), while Satterthwaite (2013) used relative RMS (single criterion). Thresholds can be applied on these metrics to censor the data (Figure 3.7). Selecting the optimal threshold for scrubbing is a trade-off. These thresholds are not prescriptive, and the optimal spike/scrubbing threshold selection will depend on both the number of volumes and the amount of motion present in a given dataset (Satterthwaite et al., 2013).

It is important to remember that removing time points from a continuous time series, can alterate the temporal structure of the data and precludes some analysis as frequency-based analyses such as Amplitude of Low Frequency Fluctuations (ALFF) and fractional ALFF (f/ALFF). For this reason, scrubbing have to be performed after functional pre-processing, because band-passing cannot be performed properly upon temporally discontinuous data.

However, scrubbing can effectively be used to minimize motion-related artefacts in resting state fMRI during seed-based correlation analyses (Power et al., 2011; 2012) or in task-based fMRI, during GLM (Siegel et al., 2014).

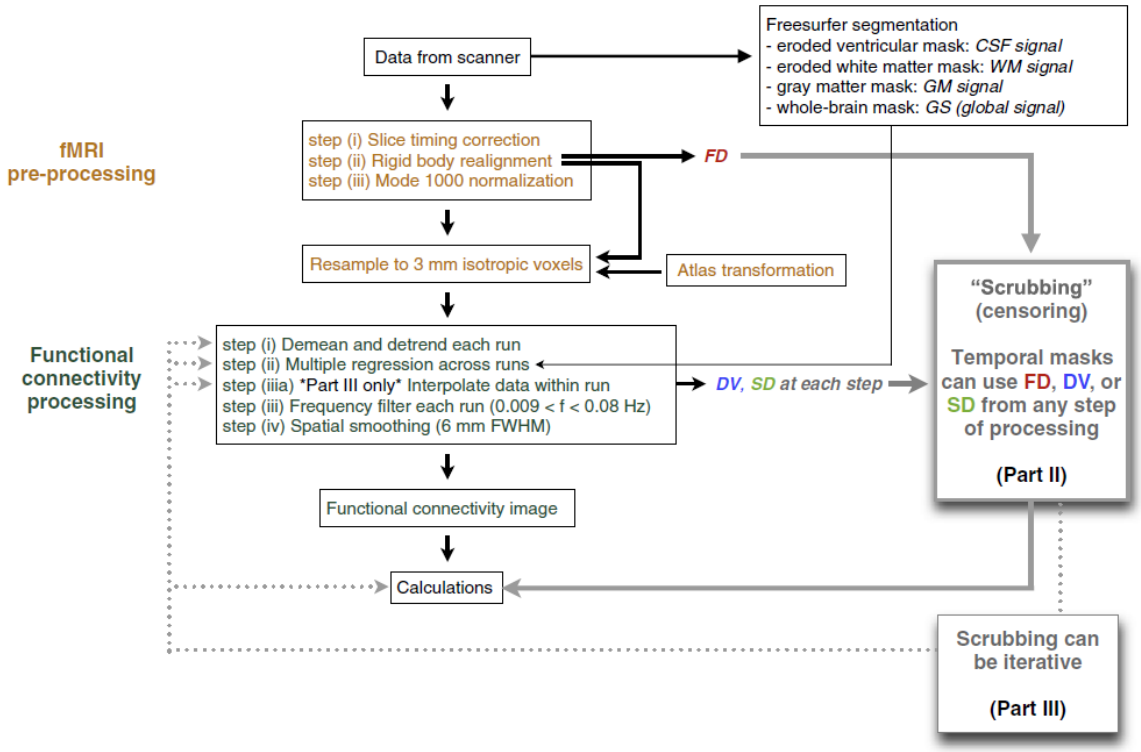


Figure 3.7 Example: Outline of scrubbing procedure (Source: Power et al., 2014)

Summary of processing recommendations.

	Why we do it	Possible drawbacks
Censoring	• Eliminates the influence of corrupted data • Conceptually and empirically most effective when implemented throughout a processing stream • Reduces dependence of correlations on motion	• Loss of data and possibly subjects • Unequal degrees of freedom across subjects But can trim to equal size
Interpolation	• Reduces amplitude of artifactual signal spread into adjacent TRs during frequency filtering	• Replacement data has synthetic characteristics Better than original characteristics But should probably not be treated as if it were original data (which is why we re-censor it)
Regression: global signal	• Strongly reduces dependence of correlations on motion Most effective at medium to long distances Positive relationships remain at short distances, which can be suppressed by censoring and interpolation • Eliminates post-motion influences on correlations Otherwise ~ 10 s post-motion influences • Reduces shared non-motion artifact Otherwise need some other type of artifact removal • Increases RSFC correspondence to ECoG	• Removal of shared neural signal • If composed of few signals Distorted correlation structure Possible misattributed group differences
Regression: white matter, CSF	• Modestly helpful at reducing artifact	
Regression: motion estimates	• Modestly helpful at reducint artifact • 24-parameter Volterra expansion increasingly used Superior to our old 12-parameter regression	• Even large expansions (36 parameters) are insufficient to remove motion artifact • 12 degrees of freedom lost per order of expansion

Table 3.1 Resume of advanced pre-processing steps. (Source: Power et al., 2014)

3.4 Impact and order of pre-processing steps

The order of these pre-processing steps is important to not reintroduce spurious signals.
A first observation is relative to temporal filtering and signal regression steps. If the fMRI time series are bandpass-filtered, but the regressors are unfiltered we obtain an unintentional reintroduction of nuisance-related variation into frequencies previously suppressed by the bandpass filter interest (Hallquist et al., 2013, Weissenbacher et al., 2009). So, the temporal filtering have to be performed after signal regression. Similarly, also smoothing is suggested to be applied after signal regression (Jo et al., 2010).

Different order of pre-processing steps has different impact on subsequent analyses. Although the chosen set of pre-processing and analysis steps (the "***pipeline***") significantly affects signal detection, pipelines are rarely quantitatively validated in the neuroimaging literature, due to complex pre-processing interactions. However, some validation were described in literature focalized on resting-state data and described the effect of various confound removal approaches (motion parameters, physiological signal, global signal, ICA-detrend), temporal faltering, and smoothing on multiple outcome measures of functional-connectivity estimates. These outcome measures comprised:

- signal-noise separation (SNS) (Shirer et al., 2015);
- test-retest reliability (TRT) of whole-brain connectivity (e.g. intra-class correlation (ICC)) (Shirer et al., 2015; Varikuti et al., 2016);
- group discriminability (GD) (Shirer et al., 2015);
- classification performance based on linear SVM (Vergara et al., 2016).

The most robust effect on resting state fMRI data quality metrics was obtained with selection of ***filtering frequency band*** that improved SNS, TRT, and GD. In comparison to temporal filtering, the effect of noise regression was generally weaker. Moreover, ***regression of signals*** of no interest provide a more biologically meaningful signal, but lead to a decrease of lead to better test–retest reliability (Yan et al., 2013; Shirer et al., 2015; Varikuti et al., 2016). Removing variance that is related to potentially confounding factors reduces reliability, pointing to the possibility that some structured noise may be improve the detectability of signal of interests. Among global and physiological signal regression, removing the mean WM and CSF signal obtained from segmentations seems to provide a good compromise, as this approach yields reliable estimates of within- and between-network connections (Yan et al. 2013, Varikuti et al., 2016).

In comparison to ***smoothing*** step, regression of head motion should be addressed before it to account for a better detection of small resting state networks (Vergara et al., 2016).
No studies reported the effect of ***normalization*** before or after advanced pre-processing. However, some software (i.e. DPARSF) speculate an improvement if data pre-processing if normalization step is performed at the end of pipeline, due to the important transformation that is applied to data by normalization.
Finally, ***temporal detrending*** was essential to remove low-frequency, reproducing time trends; however, higher-order polynomial warps compared to affine alignment had only a minor impact on the performance metrics (Strother et al., 2004).
Seed-based analyses showed that the ***definition of seed*** is important: the reliability of connectivity estimates is improved by use of seed obtained from intersection between average of grey matter segmentations of each subject and seed ROI at group analysis, and single-subject grey matter segmentation for within subject studies (Varikuti et al., 2016).

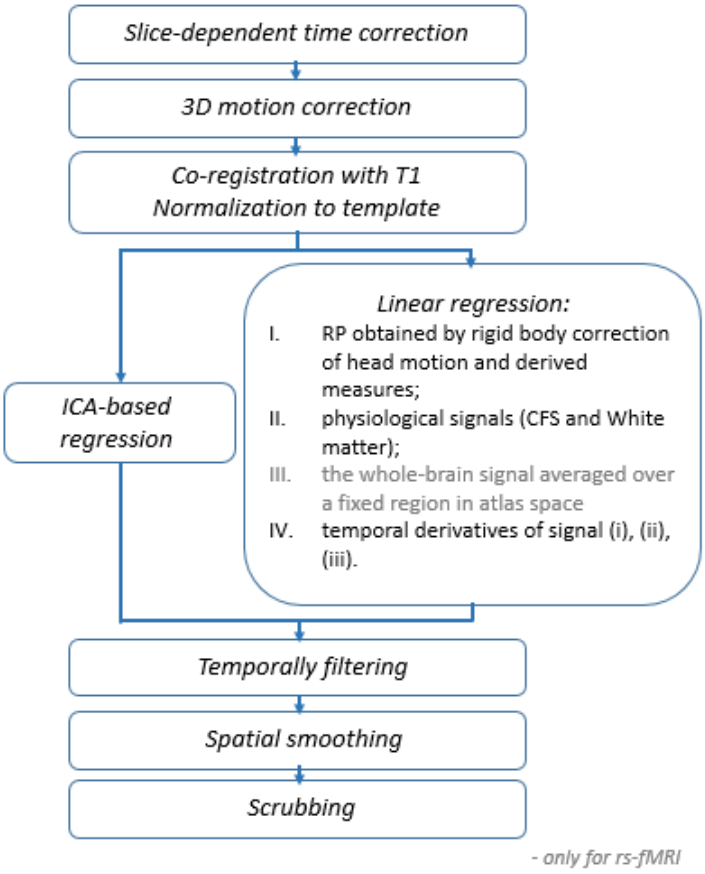


Figure 3.8 Pre-processing pipeline: optimized order of steps

Moreover, one potentially alternative to choosing such a rigid processing pipeline is the use of subject-specific processing pipelines (Churchill et al., 2012; Zhang et al., 2009). It is demonstrated

that pre-processing choices have significant, common effects on performance, across all subjects; however, there subject-specific pipelines that significantly improve on the optimal fixed-pipeline choice. Individual-subject optimization tends to increase the spatial extent of shared task activations, but with a trade-off of increased between-subject variance in voxel-wise signal amplitude of the SPM activation regions (Churchill et al., 2012).

3.5 General linear model (GLM)

The General Linear Model (GLM) is a multiple regression analysis suitable for both multiple qualitative and quantitative variables. Friston and colleagues (Friston et al. 1995) introduce this statistical method for estimation of systematic fluctuations of the BOLD signal induced by experimental stimulation in fMRI data.

GLM aims to "explain" or "predict" the variation of a dependent variable in terms of a linear combination (weighted sum) of several reference independent variables (predictors or regressors of interest). The dependent variable corresponds to the observed fMRI time course of a voxel and regressors of interest correspond to time courses of idealized fMRI responses for different conditions of stimulation during an experimental task. A predictor *x* is obtained by convolution of a condition of interest in task, modelled as a box-car time course, with a standard HRF. Different implementation of HRF were provided. The box-car time course of the condition of interest is defined as 1 in the time points were the corresponding stimuli was presented and 0 elsewhere.

For each voxel of the image a time course y_i (with i a generic voxel), with intensity of the BOLD signal at each time point (from 1 to n), is associated and it can be modelled as the sum of a number of known predictor variables ($x_1, \ldots, x_p$) each scaled by a parameter (β):

$$y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots x_{i,j}\beta_j + \cdots + x_{i,p}\beta_p + \varepsilon_i$$

where

- p number of predictors, n number of volumes in time series,
- β_j parameter estimate for generic predictor variable j ($j=1, \ldots, p$),
- ε_i error term for voxel i.

In GLM, the term “linear” refers to the additive relationship between the different weighted predictor variables; GLM is defined an “univariate” statistic, because it is referred to a single the dependent variable: a separate GLM is performed independently for each voxel time series. Spatial

covariance between neighbouring voxels is thus typically ignored at the model fitting stage and is considered subsequently.

Using the compact matrix notation we can reformulate this equation as:

$$Y = X\beta + \varepsilon$$

where

- Y is a n x 1 vector,
- X is the n x p design matrix with for each columns a predictor,
- β is a p x 1 vector with "coefficient" or "beta weight", quantifying potential contribution of each predictor in explaining the time course of voxel i,
- and ε is a n x 1 vector, an error value ("residual") associated with residual time course value that linear combination will not perfectly explain through predictors due to noise fluctuations.

Residual term can be expressed as difference between observed Y and predicted $\hat{Y}$ value of time course:

$$\varepsilon = Y - \hat{Y} = Y - X\beta$$

The aim of GLM is to estimate if, and to what extent, each predictor contributes to the variability observed in the voxel's time-course thought the estimation of beta values. Generally, we have more time points that predictors ($p << n$), which means that there is not a unique solution in the determination of beta weights. So, the beta weights are estimated as the value that minimize the sum of squared residuals (error term or unexplained variance),

$$\sum_{i=1}^{n} \left(Y_i - X_i\hat{\beta}\right)^2$$

i.e. the difference between the observed and predicted value of time course. This is obtained using different least square methods: ordinary least squares, (feasible) generalized least squares, and the so-called "smoothing and sandwich" approaches (Waldorp, 2009). The parameters are estimated as follow:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad var\left(\hat{\beta}\right) = \sigma^2 (X^T X)^{-1}$$

The square of the multiple correlation coefficient R provides a measure of the proportion of the variance of the data, which can be explained by the model:

$$R^2 = \frac{var(\hat{Y})}{var(Y)}$$

Usually, for completeness the first term of the design matrix is a constant term, to fit the base level of the signal time course, so *X* is a *n* x *p+1* matrix.

3.5.1 Assumptions

GLM operates correctly only under the following assumptions:

1. linearity: the relationship between the dependent variable and the recently freed independent variable is linear
2. predictors are not multicollinear, i.e. none of the explanatory dependent variable is perfectly correlated with any other explanatory variable, or any linear combination of. This assumption is important, because the $(X^T X)^{-1}$ does not exist, and infinite solutions for beta weights are found (Mumford et al., 2015).
3. residuals are assumed to be normal independent and identically distributed (normal i.i.d.). This implies four sub-assumptions: residuals
    a. must have an expected value of zero at each time point ($E[\varepsilon_i] = 0$),
    b. constant variance (homoscedasticity, $var[\varepsilon_i] = \varepsilon^2$),
    c. are assumed to be uncorrelated, ($Cov(\varepsilon_i, \varepsilon_j) = 0$) for all $i \neq j$, and
    d. are further assumed to be normally distributed, ($\varepsilon_i \sim N(0, \sigma^2)$).

If these assumption are verified, the solution obtained by the least squares method is optimal and it provides the most efficient unbiased estimation of the beta weights, otherwise non-conformity to the assumptions introduces bias into estimates of the variance, thus affecting test statistics, power, and false positive rates (Monti et al., 2011). In the implementation of GLM for fMRI these assumption are tested and corrected through different methods.

Although an initial evidence of linearity in fMRI signal (Boyton et al., 1996), different sources of non-linear signals are present in BOLD signal: vascular response, due to the vaso-elastic properties of the blood vessels and adaptive behaviour in neuronal-response (Logothetis, et al., 2003). These aspects can be limited by an accurate definition of the paradigm of stimulation (i.e. duration of single stimuli and inter-trial duration in block and event-related tasks).

Serial correlations violate the uncorrelation of the residuals and different approaches are used. In pre-coloring, temporal smoothing, acting as a low-pass filter, is applied and a known

autocorrelation structure is imposed on the data (Friston et al., 1995). In pre-whitening, through a two-pass procedure, autocorrelation are estimated and removed from GLM model, which is re-fitted (Bullmore et al.,1996). Moreover, these serial correlation can be reduced though an explicit noise-modelling, because residuals can often be taken as evidence of un-modeled, but potentially known, sources of variance. For example though "nuisance variable regression" approach, several factors that can induce autocorrelation, such as hardware related low-frequency drift, residual movement effects, and physiological noise (cardiac pulsation and respiration) (Lund et al., 2006) are modelled in design matrix. Indeed, inclusion of irrelevant factors in the model may have the effect of increase fit and increase power. On the other hand, an over-fitting of the BOLD signal can determine a mis-modeling and degrade the generalization of the results (Petersson et al., 1999).

3.5.2 Significance tests and contrasts of interest

To assess if an individual condition modelled from a predictor differs statistically from another is necessary to introduce a ***t test***. At this purpose, comparisons among beta weights can be performed as contrast *c* (Pernet et al., 2014) to identify significant differences from null hypothesis. For example, to test whether activation:

- in a single predictor *j* deviates significantly from baseline, the null hypothesis would be $H_o: \beta_j = 0$, where *c* = +1 and $\beta = \beta_j$ (value used to multiply the respective beta weight $c^T\beta = 0$);
- in predictor 1 is significantly different from activation in predictor 2, the null hypothesis would state that the beta weights of the two predictors would not differ, $H_o: \beta_1 = \beta_2$ or $H_o: \beta_1 - \beta_2 = 0$, where *c* = [+1 – 1] and $\beta = [\beta_1, \beta_2]$;
- in mean of predictor 1 and predictor 2 differs from predictor 3, the following null hypothesis could be specified: $H_o: (\beta_1 + \beta_2)/2 = \beta_3$ or $H_o: \beta_1 + \beta_2 - 2\beta_3 = 0$, where *c* = [+1 +1 -2] and $\beta = [\beta_1, \beta_2, \beta_3]$.

The constant term of design matrix represent the baseline o reference.

The following t test can be used:

$$t = \frac{c^T\beta}{\sqrt{var(\varepsilon)c^T(X^TX)^{-1}c}}$$

with *n* - *p* degrees of freedom. With the known degrees of freedom, all t value of all voxel for a specific contrast can be converted in a map thresholded using a *p-value* (usually tow-sided).

If value of test is positive and *p-value* is smaller than 0.05, the alternative hypothesis may be concluded, so significant differences can be assessed.

3.5.3 Multiple comparison

Testing around 100,000 voxels independently, for univariate statistics, at $p < 0.05$ means that on average, 5000 will be false positives. So, a correction for multiple comparison is necessary. Different methods can be used:

- Bonferroni correction: $p_{corrected} = p_{uncorrected}/n$; however if we consider that a spatial smoothing was performed in pre-processing and nearby voxels are correlated, Bonferroni correction appear to be overly conservative in fMRI.
- Familywise Error Rate (FWE), using Random Field Methods, controls for multiple comparisons in a similar way to Bonferroni correction, but using the number of "resels", that depend on the smoothness of the data, rather than the number of voxels to adjust the *p-value*.
- False Discovery Rate (FDR): proportion of false positives among rejected tests.

References

Ashby, F.G. (2011). Preprocessing. Statistical Analysis of MRI Data. Cambridge, MA: MIT Press.

Babo-Rebelo M, Richter CG, Tallon-Baudry C. (2016). Neural Responses to Heartbeats in the Default Network Encode the Self in Spontaneous Thoughts. J Neurosci;36(30):7829-40.

Beckmann CF, Smith SM. (2004). Tensorial extensions of independent component analysis for multisubject FMRI analysis. Neuroimage;25(1):294-311.

Birn RM, Smith MA, Jones TB, Bandettini PA. (2008). The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. Neuroimage;40(2):644-54.

Boynton GM, Engel SA, Glover GH, Heeger DJ. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci.;16(13):4207-21.

Bullmore E, Brammer M, Williams SC, Rabe-Hesketh S, Janot N, David A, Mellers J, Howard R, Sham P. (1996). Statistical methods of estimation and inference for functional MR image analysis.Magn Reson Med.;35(2):261-77.

Campbell KL, Grigg O, Saverino C, Churchill N, Grady CL. (2013). Age differences in the intrinsic functional connectivity of default network subsystems. Front Aging Neurosci;5:73.

Chang C, Cunningham JP, Glover GH. (2009). Influence of heart rate on the BOLD signal: the cardiac response function. Neuroimage;44(3):857-69.

Churchill NW, Oder A, Abdi H, Tam F, Lee W, et al. (2012) Optimizing preprocessing and analysis pipelines for single-subject fMRI. I. Standard temporal motion and physiological noise correction methods. Hum Brain Mapp 33: 609–627.

Davey CE, Grayden DB, Egan GF, Johnston LA. (2013). Filtering induces correlation in fMRI resting state data. Neuroimage;64:728-40. doi: 10.1016/j.neuroimage.2012.08.022. Epub 2012 Aug 25.

De Martino F, Gentile F, Esposito F, Balsi M, Di Salle F, Goebel R, Formisano E. (2007). Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers. Neuroimage;34(1):177-94. Epub 2006 Oct 27.

Desjardins AE, Kiehl KA, Liddle PF. (2001). Removal of confounding effects of global signal in functional MRI analyses. Neuroimage;13(4):751-8.

Evans A. C., Collins D. L. and Milner. B. (1992). An MRI-based stereotactic atlas from 250 young normal subjects", Journal Soc. Neurosci. Abstr. 18: 408.

Fair DA, Nigg JT, Iyer S, Bathula D, Mills KL, Dosenbach NU, et al. (2013). Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. Front Syst Neurosci;6:80.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A.;102(27):9673-8.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI timeseries revisited. NeuroImage, 2(1), 45–53.

Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. (1996). Movement-related effects in fMRI time-series. Magn Reson Med;35(3):346-55.

Garrett DD, Kovacevic N, McIntosh AR, Grady CL. (2010). Blood oxygen level-dependent signal variability is more than just noise. J Neurosci;30(14):4914-21.

Glover GH, Li TQ, Ress D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med;44(1):162-7.

Hajnal JV., Myers R., Oatridge A., Schwieso JE., Young IR., Bydder GM. (1994). Artifacts due to stimulus-correlated motion in functional imaging of the brain. Magn Reson Med, 3, p283–291.

Hallquist MN, Hwang K, Luna B. (2013). The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. Neuroimage;82:208-25.

Henson R., Buechel C., Josephs O., Friston K. (1999). The slice-timing problem in event-related fMRI NeuroImage, p. 125-125.

Holden M. (2008). A review of geometric transformations for nonrigid body registration. IEEE Trans Med Imaging, 27, p 111-128

Jo HJ, Saad ZS, Simmons WK, Milbury LA, Cox RW. (2010). Mapping sources of correlation in resting state FMRI, with artifact detection and removal. Neuroimage;52(2):571-82.

Johnstone T, Ores Walsh KS, Greischar LL, Alexander AL, Fox AS, et al. (2006) Motion correction and the use of motion covariates in multiple-subject fMRI analysis. Hum Brain Mapp 27: 779–788.

Junghöfer M, Schupp HT, Stark R, Vaitl D. (2005). Neuroimaging of emotion: empirical effects of proportional global signal scaling in fMRI data analysis. Neuroimage;25(2):520-6.

Kundu P, Inati SJ, Evans JW, Luh WM, Bandettini PA. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. Neuroimage;60(3):1759-70

Lemieux, L., Salek-haddadi, A., & Lund, T. E. (2007). Modelling large motion events in fMRI studies of patients with epilepsy. Magn Reson Imaging., 25(6), 894–901.

Lund TE, Madsen KH, Sidaros K, Luo WL, Nichols TE. (2006). Non-white noise in fMRI: does modelling have an impact? Neuroimage;29(1):54-66.

Macey PM, Macey KE, Kumar R, Harper RM. (2004). A method for removal of global effects from fMRI time series. Neuroimage;22(1):360-6.

McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ. (1998). Analysis of fMRI data by blind separation into independent spatial components. Hum Brain Mapp;6(3):160-88.

Mikl M, Mareček R, Hluštik P, Pavlicová M, Drastich A, Chlebus P, Brázdil M, Krupa P. (2008). Effects of spatial smoothing on fMRI group inferences. Magn Reson Imaging; 26: 490 – 503

Monti MM. (2011). Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. Front Hum Neurosci.;5:28.

Mumford JA, Poline JB, Poldrack RA. (2015). Orthogonalization of regressors in FMRI models. PLoS One;10(4):e0126255.

Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? Neuroimage;44(3):893-905.

Nemani AK, Atkinson IC, Thulborn KR (2009) Investigating the consistency of brain activation using individual trial analysis of high-resolution fMRI in the human primary visual cortex. NeuroImage 47: 1417–1424.

Oakes TR, Johnstone T, Ores Walsh KS, Greischar LL, Alexander AL, Fox AS, Davidson RJ. (2005). Comparison of fMRI motion correction software tools. Neuroimage. 2005 Nov 15;28(3):529-43.

Park HD, Correia S, Ducorps A, Tallon-Baudry C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. Nat Neurosci;17(4):612-8.

Pernet CR. (2014). Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers. Front Neurosci.;8:1.

Petersson KM, Nichols TE, Poline JB, Holmes AP. (1999). Statistical limitations in functional neuroimaging. II. Signal detection and statistical inference. Philos Trans R Soc Lond B Biol Sci.;354(1387):1261-81. Review.

Poldrack, R. A., Mumford, J., & Nichols, T. (2011). Preprocessing fMRI data. In Handbook of Functional MRI Data Analysis. Cambridge: Cambridge University Press.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Anne, K., Church, J. A., … Miezin, F. M. (2011). Functional network organization of the human brain. Neuron, 72(4), 665–678.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect , characterize , and remove motion artifact in resting state fMRI. NeuroImage, 84, 320–341.

Power JD, Schlaggar BL, Petersen SE. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage. 2015 Jan 15;105:536-51. Review.

Pruim RH, Mennes M, Buitelaar JK, Beckmann CF. (2015). Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. Neuroimage;112:278-87.

Razavi M., Grabowski T. J., Vispoel W. P., Monahan P., Mehta S., Eaton B., Bolinger L. (2003). Model assessment and model building in fMRI. Hum. Brain Mapp. 20, 227–23810.1002/hbm.10141

Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Elliott MA, Hakonarson H, Gur RC, Gur RE. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage. 2012 Mar;60(1):623-32.

Satterthwaite, T. D., Elliott, M. a., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., … Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. NeuroImage, 64(1), 240–256.

Shirer WR, Jiang H, Price CM, Ng B, Greicius MD. (2015). Optimization of rs-fMRI Pre-processing for Enhanced Signal-Noise Separation, Test-Retest Reliability, and Group Discrimination. Neuroimage;117:67-79.

Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Jessica, A., Schlaggar, B. L., … Louis, S. (2014). Statistical Improvements in Functional Magnetic Resonance Imaging Analyses Produced by Censoring High-Motion Data Points. Human Brain Mapping, 35(5), 1981–1996.

Sladky R, Friston KJ, Tröstl J, Cunnington R, Moser E, Windischberger C. (2011). Slice-timing effects and their correction in functional MRI. Neuroimage;58(2):588-94.

Smith, AM, Lewis, BK, Ruttimann, UE, Ye, FQ, Sinnwell, TM, Yang, Y, Duyn, JH, & Frank, JA. (1999). Investigation of low frequency drift in fMRI signal. Neuroimage, 9, 526–33.

Smyser CD, Inder TE, Shimony JS, Hill JE, Degnan AJ, Snyder AZ, Neil JJ. (2010). Longitudinal analysis of neural network development in preterm infants. Cereb Cortex. 2010 Dec;20(12):2852-62.

Strother S, La Conte S, Kai Hansen L, Anderson J, Zhang J, Pulapura S, Rottenberg D. (2004). Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. Neuroimage. 2004;23 Suppl 1:S196-207.

Talairach J. and Tournoux P. (1988). Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging, Thieme Medical Publishers, New York, NY.

Thomas CG, Harshman RA, Menon RS. (2002). Noise reduction in BOLD-based fMRI using component analysis. Neuroimage;17(3):1521-37.

Turner R Howseman A Rees GE Josephs O and Friston K. (1998). Functional magnetic resonance imaging of the human brain: data acquisition and analysis. Exp. Brain Res.
123:5-12.

Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J Neurophysiol;103(1):297-321.

Van Dijk, K. R. A., Sabuncu, M. R., & Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. NeuroImage, 59(1), 431–438.

Varikuti DP, Hoffstaedter F, Genon S, Schwender H, Reid AT, Eickhoff SB. (2016). Resting-state test-retest reliability of a priori defined canonical networks over different preprocessing steps. Brain Struct Funct

Vergara VM, Mayer AR, Damaraju E, Hutchison K, Calhoun VD. (2016). The effect of preprocessing pipelines in subject classification and detection of abnormal resting state functional network connectivity using group ICA. Neuroimage. pii: S1053-8119(16)00243-3.

Waldorp L. (2009). Robust and unbiased variance of GLM coefficients for misspecified autocorrelation and hemodynamic response models in fMRI. Int J Biomed Imaging;2009:723912.

Weissenbacher A, Kasess C, Gerstl F, Lanzenberger R, Moser E, Windischberger C. (2009). Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. Neuroimage;47(4):1408-16.

Wilke M, Lidzba K, Staudt M, Buchenau K, Grodd W, et al. (2005) Comprehensive language mapping in children, using functional magnetic resonance imaging: what's missing counts. Neuroreport 16: 915–919. doi: 10.1097/00001756-200506210-00008

Wilke, M. (2012). An alternative approach towards assessing and accounting for individual motion in fMRI timeseries. NeuroImage, 59(3), 2062–2072.

Yan CG, Craddock RC, He Y, Milham MP. (2013). Addressing head motion dependencies for small-world topologies in functional connectomics. Front Hum Neurosci;7:910.

Zhang J, Anderson JR, Liang L, Pulapura SK, Gatewood L, Rottenberg DA, Strother SC. (2009). Evaluation and optimization of fMRI single-subject processing pipelines with NPAIRS and second-level CVA.
Magn Reson Imaging;27(2):264-78.

Chapter 4

Motion data assessment using auto-calibrated methods to detect outlier subjects and volumes in fMRI studies

4.1 Introduction

In the recent years, the field of signal denoising in functional resonance imaging (fMRI) is rapidly evolving (Power, Schlaggar, & Petersen, 2015) with the implementation of specific post-processing pipelines. Notably, subject motion was demonstrated to cause spurious signals introducing false-positive activations in subsequent statistical analysis (Bullmore et al., 1999; Field, Yen, Burdette, & Elster, 2000).

One of the mandatory pre-processing step in fMRI data for the subject motion correction is the realignment procedure of each single scan volume to a reference volume (the first or the mean volume of the entire scan) through a rigid body transformation, minimizing the difference among volumes through different cost functions (K. J. Friston et al., 1995; Jenkinson, Bannister, Brady, & Smith, 2002). During this step, the position of the head in the space is estimated at each time point and described with six realignment parameters: translational displacements along the x-, y-, and z-axis, and rotational displacements of pitch, yaw, and roll. These parameters are considered a measure of motion obtained from data themselves (Power et al., 2015). However, even after realignment procedure, a residual motion related variance is still present (Andersson, Hutton, Ashburner, Turner, & Friston, 2001). Different patterns of movement in realignment parameters can be detected: subjects who moved intermittently and returned to their original position (i.e. spike movements), and subject with head drift movements and who remained displaced from the origin (Power et al., 2014). These different patterns had not the same impact on fMRI signal: usually, the first pattern had a negative impact on the quality of the signal detected, in particular if frequent high displacements occurred in less than one volume (Christodoulou et al., 2013; Power et al., 2014).

In specific populations of subjects that tend to move during scanning session, such as children (M Wilke, Holland, Myseros, Schmithorst, & Ball, 2003) or patients with disorder of consciousness (Harrison & Connolly, 2013), the evaluation of the impact of motion related artefacts on fMRI data is extremely important: 1) to exclude outlier subjects for movements using an "outlier subjects detection" method and 2) to minimize, whenever possible, the impact of motion related artefacts in included subjects using an "outlier volumes detection" method.

In group studies, it is important to ensure homogeneity of fMRI data quality for motion of all subjects of the population under investigation. To discriminate in the investigated sample outlier subjects for head movements (point 1), an "outlier subjects detection" is performed to identify these participants. Outliers could determine violation of assumptions for subsequent statistical analyses, increase of error variance, and reduction of the power of subsequent statistical analyses (Rasmussen, 1988; Schwager & Margolin, 1982; Zimmerman, 1994).

The problem of outliers identification in a dataset containing outliers is well illustrated by Roussseuw and Leroy (1987). For example, within the context of classical multiple regression, they point out that outliers cannot always be discovered by looking at least squares residuals: some outliers are leverage points which greatly distort the regression lines so that they (the leverage-point outliers) yield relatively small least squares and some "good" points yield large least squares. Roousseuw and Leroy state that "… there are many multivariate data sets … where the outliers remain invisible even through a careful analysis of the least squared residuals."

Currently, different indices, derived from the six realignment parameters, were proposed to detect outlier subjects for resting state and task-based fMRI: root mean square (RMS) (Oakes et al., 2005; Siegel et al., 2014), RMS of first derivative (RMS d/dt) (Power et al., 2014; Van Dijk, Sabuncu, & Buckner, 2012), and absolute translation and rotation displacements of realignment parameters (Power et al., 2015; Van Dijk et al., 2012). Subjects exceeding, for the entire session of part of it, different threshold values (mm/degrees or voxel's width - (Churchill et al., 2012; Johnstone et al., 2006; Siegel et al., 2014) are considered outliers. The choice of these thresholds on these metrics are user-dependent, because they are strictly related to the type of subjects included (Power et al., 2015): studied populations differ for clinical or demographic (for example age) characteristics.

In each single-subject, it is important to ensure homogeneity of data quality among volumes in time. To minimize the impact of motion related artefacts in low-quality data (point 2), an "outlier volumes detection" with the identification of volumes corrupted by excessive motion (Power et al., 2015) is performed. Different methods have been proposed and, among them, Siegel et al. (Siegel et

al., 2014) showed that the so-called “scrubbing” (or “censoring”) procedure can improve the detectability of task-based fMRI signal, even in subjects with mild amounts of head movements. Importantly, they verified that this approach performed better than including motion related nuisance regressors in GLM as usually done (K. Friston, Williams, Howard, Frackowiak, & Turner, 1996). To implement the scrubbing procedure, a binary temporal mask that identifies volumes corrupted by motion is created and included in GLM as a regressor of no interest. To define this mask, framewise displacement (FD) metric (Power et al., 2011) is suggested (Siegel et al., 2014). However, analogously to “outlier subjects detection”, the identification of an user-dependent motion threshold on this metric is necessary to identify volumes associated to movements.

In outlier subjects and outlier volumes detection the choice of thresholds is crucial (Satterthwaite et al., 2013). It depends from the length and sampling of the acquisition and the amount of motion. In fact, the more stringent the thresholds are, the more efficient is the removal of outlier subjects or motion related volumes. As a drawback, stringent thresholds might increase the intrinsic variability and decrease the test-retest reliability of the data (Power et al., 2015). On the other side, permissive threshold could determine violation of assumptions and reduction of the power of subsequent statistical analyses. Therefore, user-dependent thresholds might be sub-optimal, in particular in specific categories of patients (e.g. children and disorder of consciousness patients).

The aim of this study is to overcome the limitations of user-dependent thresholds in fMRI data motion assessment: we propose auto-calibrated procedures, tested in three fMRI datasets, to detect outlier subjects/volumes for movements. We assessed the presence of outlier subjects and volumes using two separate auto-calibrated procedure based on clustering analyses and Mahalanobis distance (MD) (Mahalanobis, 1936) applied to realignment parameters. Two levels were implemented: i) “outlier subjects detection” to detect subjects that can be considered outliers for large movements and ii) “outlier volumes detection” to identify volumes corrupted by motion on which to apply the scrubbing procedure. Clustering is an unsupervised learning method that groups data instances into subsets, called clusters, based on a distance measure of similarity/dissimilarity among instances (Tan, Steinbach, & Kumar, 2006), discovering automatically hidden patterns targeted on each study cohort (Tan et al., 2006). For clustering method, in each level, a two-steps architecture was implemented: 1) application of clustering algorithms and 2) validation of clustering results (Tan et al., 2006). MD, is a robust multivariate outlier detection method, able to account for the difference of variance in each variable and the covariance between variables (Filzmoser, 2004).

The performance of these methods were compared to the results obtained with previous indices described in literature (outlier subjects detection: absolute and relative displacements, and RMS and

RMS d/dt of realignment parameters; outlier volumes detection: FD). The evaluation among methods was performed on GLM residuals and specific indices.

4.2 Materials and Methods: participants and acquisition details

4.2.1 Participants

Three different datasets of healthy participants, previously acquired at the Fondazione IRCCS Istituto Neurologico "Carlo Besta" (from 2011 to 2014), were used in this study to assess the implemented motion exclusion procedures and to compare the results with previous methods. Participants of dataset 1 were collected as control group for a study on children with early focal brain lesions, while subjects of datasets 2 and 3 were recruited as control group during CRC-Start Up Coma Research Centre project. Local Ethics Committee approved the study. Written informed consent was obtained from each healthy participant or the legally authorized representative of child prior to their inclusion in the study. MRI acquisition did not reveal brain abnormalities and no history of neurological or psychiatric diseases were reported by all healthy subjects. Demographic information of the subjects included in the three datasets were reported in Table 4.1.

Dataset	n. subjects	Age (mean ± SD)	Sex (M/F)	Task
1	17	13 ± 3.76	10/7	Language task: verbal fluency task
2	31	38 ± 12.32	18/13	Visual task: faces and houses
3	28	37 ± 10.98	14/14	Olfactory task: unpleasant odors

Table 4.1 Demographic and task information.

4.2.2 MRI and fMRI imaging

Imaging data of dataset 1 were obtained with a 1.5T MR scanner (Magnetom Avanto, Siemens AG, DE), equipped with an 8-channel head matrix coil, whereas imaging data of dataset 2 and 3 were acquired using an 3T MR scanner (Achieva , Philips Healthcare BV, NL), equipped with a 32-channel head coil.

In each protocol a high resolution 3D T1-weighted gradient-echo sequence (dataset 1: 160 sagittal slices, TR = 1640 ms, TE = 2 ms, FOV = 256 x 256 mm, no gap, voxel size = 1 x 1 x 1 mm^3, flip angle = 12°; dataset 2 and dataset 3: 185 sagittal slices, TR = 9.781 ms, TE = 4.6 ms, FOV = 240 x 240 mm, no gap, voxel size = 1 x 1 x 1 mm, flip angle = 8°) was acquired for each participant.

For each participant, functional images sensitized to blood–oxygen level-dependant (BOLD)

contrast were obtained using whole brain T2*-weighted echo-planar imaging (EPI) (dataset 1: 25 axial slices, TR = 3950 ms, TE = 50 ms, FOV = 128 x 128 mm, no gap, voxel size= 2 x 2 x 4 mm3, dynamic scans = 100; dataset 2: 40 axial slices, TR = 2500 ms, TE = 30 ms, FOV = 240 x 240 mm, gap = 0.5 mm, voxel size = 3 x 3 x 3 mm3, flip angle = 90°, dynamic scans = 245, SENSE = 2.5; dataset 3: 40 sagittal slices, TR = 2500 ms, TE = 20 ms, FOV = 240 x 240 mm, gap = 0.5 mm, voxel size = 3 x 3 x 3 mm3, flip angle = 90°, dynamic scans = 408, SENSE = 2.5).

4.2.3 fMRI tasks

Three datasets were chosen for this study: two datasets were characterized by high head movements unrelated (dataset 1) and related (dataset 3) to the execution of the task and by low head movements (dataset 2).

E-Prime (EPrime2 Professional, Psychology Software Tools, PA) was used for stimuli administration. Healthy participants were fitted with MRI-compatible headphones. Visual stimuli were presented using googles (VisuaStim, Resonance Technology Inc., Northridge CA, USA) at 3T scanner, whereas using a back-projector at 1.5T scanner. Odour stimuli were presented using an olfactometer MRI-compatible (Lundström, Gordon, Alden, Boesveldt, & Albrecht, 2010) capable of delivering square-shaped odor stimuli using a maximum birhinal air flow of 3.5 l/m (1.75 l/m per nostril) (A. Nigri et al., 2016).

4.2.3.1 Dataset 1: Verbal fluency task

An overt verbal fluency (VF) block-design task was administered to a cohort of children. Three 20-s long conditions in a pseudorandomized order were presented: a black fixation cross for the rest condition, a letter of Italian alphabet for the phonemic fluency condition, and a semantic category (fruits, animals, colours, …) for the semantic fluency condition. The participants were asked to overtly produce as many words as possible beginning with the presented letter for the phonemic fluency condition and belonging to the semantic category for the semantic fluency condition. All stimuli were presented in black on a grey background. All subjects performed the task correctly.

4.2.3.2 Dataset 2: Visual task

All participants performed a block design task. Five blocks showing a set of faces were alternated with 5 blocks showing a set of houses. Face or house blocks were spaced out by 10 rest blocks showing a white fixation cross on a black background. Each block lasted 30 seconds. Face and house blocks included 10 stimuli, each shown for 3 s. Details of the task were reported in previous study (Bertolino, Ferraro, Nigri, Bruzzone, & Ghielmetti, 2014).

4.2.3.3 Dataset 3: Olfactory task

Odor stimuli were presented within a blocked design, alternating two 30-s-long conditions in 14 cycles in a pseudo-randomized order: rest condition, in which clean air was presented, and odor condition. In latter condition, 6 odor deliveries, lasting 2 s each one, were interspaced with 3 s of clean air flow to minimize habituation and adaptation. Two different odor stimuli (odorants 1-octenol-3-ol like mushroom; n-butanol like white-board marker) were administered to both nostrils in a random and balanced order during odor block. Details of the task were reported in previous studies (A. Nigri et al., 2016; Anna Nigri et al., 2013).

For each single dataset (1: VF task, 2: visual task, 3: olfactory task), we performed, separately, the following analyses procedures.

4.3 Materials and Methods: pre-processing and algorithm implementations

4.3.1 Data pre-processing

Analyses were performed using the following softwares and in-house scripts: MATLAB R2014b (Statistics and Machine Learning Toolbox, and Signal Processing Toolbox; Mathworks Inc., Natick, MA, USA) and Statistical Parametric Mapping (SPM8) (Wellcome Department of Cognitive Neurology, London, UK).

fMRI data standard pre-processing included realignment to the mean image producing 6 realignment parameters (3 translational RPt and 3 rotational RPr parameters), normalization to the MNI template (Evans et al., 1997) with the unified model framework procedure, spatial smoothing (6-mm FWHM Gaussian isotropic kernel), temporal low-pass filtering with a hemodynamic response kernel, and temporal high-pass filtering with a cut-off period of 128 s (K. J. Friston et al., 1995). T1 image was co-registered to fMRI mean image and normalized to MNI template.

4.3.2 Quality data assessment of movements

The clustering and MD methods were applied to obtain an “outlier subjects detection” and an “outlier volumes detection”. Realignment parameters were used as input.

In clustering method, for both outlier subjects/volumes detections, a modular structure was implemented: a first phase of clustering algorithm and a subsequent phase of validation of clustering results. The choice of the type clustering algorithm, hierarchical (HCl) and k-means (KM) clustering, was defined based on the type of data input and a-priori hypotheses (Tan et al.,

2006). In the validation phase, different measures were used to ensure reliable clustering results (Tan et al., 2006): Silhouette index (Sh), Davies -Bouldin index (DB), and Cophenetic correlation coefficient (CPCC); CPCC was used only for "outlier volumes detection". Sh and DB indices assessed within-cluster similarity in comparison to between-clusters dissimilarity and allowed to select the optimal number of clusters in the sample (Kovács, Legány, & Babos, 2005). Meanwhile, CPCC, only for HCl, gives a measure of how accurately the clustering solution of dendrogram reflects original data. Details of clustering algorithms (HCl, KM) and validation indices (Sh , DB, and CPCC) are explained in Chapter 2.

An additional method based on MD was implemented, because MD is a robust method for multivariate outlier detection (Filzmoser, 2004) and is able to account for the difference of variance in each variable and the covariance between variables. Therefore, this metric provides a way to measure distances that resume the different scales of the variables and quantify size and distribution of them in a covariance matrix (Filzmoser, 2004).

To test normality assumption of multivariate dataset, Mardia test (Mardia, 1980) is applied. The two derived measures, skewness and kurtosis, allow to test two hypotheses that are compatible with the assumption of multinormality ($p<.05$). Although MD is originally thought to be used for multivariate normal distributions, there has been statistical-deductive efforts to support its application beyond the set of normal distribution (Ekstrom, 2011).

For a normal multivariate sample (*m*-by-*n* matrix) of *m* variables and *n* observations x_i ($i = 1, \dots n$), the squared MD (Filzmoser, 2004; Mahalanobis, 1936) is defined as

$$MD_i^2 = (x_i - \bar{x})' C^{-1} (x_i - \bar{x})$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$C = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})\,(x_i - \bar{x})'$$

where $\bar{x}$ is the multivariate arithmetic mean and C the estimated covariance matrix.

To identify outliers, squared MD was compared with a critical value of the χ^2 distribution ($\alpha = .05$) for multivariate normal distributions (Rousseeuw & Van Zomeren, 1990).

We compared clustering and MD results, for "outlier subjects detection" and "outlier volumes detection", to previous indices described in literature (RMS, RMS d/dt, and absolute displacements for "outlier subjects detection"; FD for "outlier volumes detection").

4.3.2.1 Outlier subjects detection

Clustering method

For each single subject, first derivatives (*d/dt*) and subsequent RMS of first derivatives were computed separately for the 3 translation (RP_t) and the 3 rotation (RP_r) realignment parameters of each single subject, in order to identify absolute displacement and exclude drift during time acquisition. Figure 4.1 reports a schematic representation of the "outlier subjects detection" workflow in the clustering method. For all subjects *S* of each dataset from realignment parameters were produced the following RMS:

$$RP_t\,[S, 3, V] \rightarrow RP_t\,d/dt\,[S, 3, V] \rightarrow RMS_t\,[S, V]$$

$$RP_r\,[S, 3, V] \rightarrow RP_r\,d/dt\,[S, 3, V] \rightarrow RMS_r\,[S, V]$$

where *S* = number of subjects, *V* = number of volumes.

RMS_t and RMS_r (*S*-by-*V* data matrices) of all subjects were used as separate inputs of two HCl algorithm, producing two dendrograms. An agglomerative HCl algorithm was chosen based on three strengths of this type of clustering: no assumptions of any particular number or distribution (size, density and shape) of clusters, less susceptibility to noise, and no need of input data normalization.

In the validation phase, each HCl was re-iterated 40 times to avoid cluster instability and to test repeatability of clustering results (Tan et al., 2006). For each iteration of HCl, separate Sh and DB indices were obtained to determine the optimal number of clusters (Chapter 2). Then, the optimal number of clusters for the two different inputs (i.e. RMS_t and RMS_r) was identified for Sh and DB based on the highest number of iterations producing that clustering results.

For both separate inputs (i.e. RMS_t and RMS_r), we compared the optimal numbers of clusters identified with Sh and DB indices. if the number of clusters of the two indices were different, no outliers were identified in the datasets and all subjects were included in 4.3.2.2 step ("outlier volumes detection"). If the number of clusters of Sh and DB indices was equal, the common clustering solution was considered reliable and the data were partitioned according with the number of clusters identified. In this case, it was equal to 2, we hypothesized that clusters were well separated. Therefore, individuals fallen in the cluster with the highest mean of RMS_t (or RMS_r) were considered outlier subjects. If the number of clusters was 3 or 4, we hypothesized that clusters were separated, but with more uncertainty than the previous case; so subjects in cluster with the highest mean of RMS_t (or RMS_r) were considered as tending to be outlier subjects, but included in subsequent analyses.

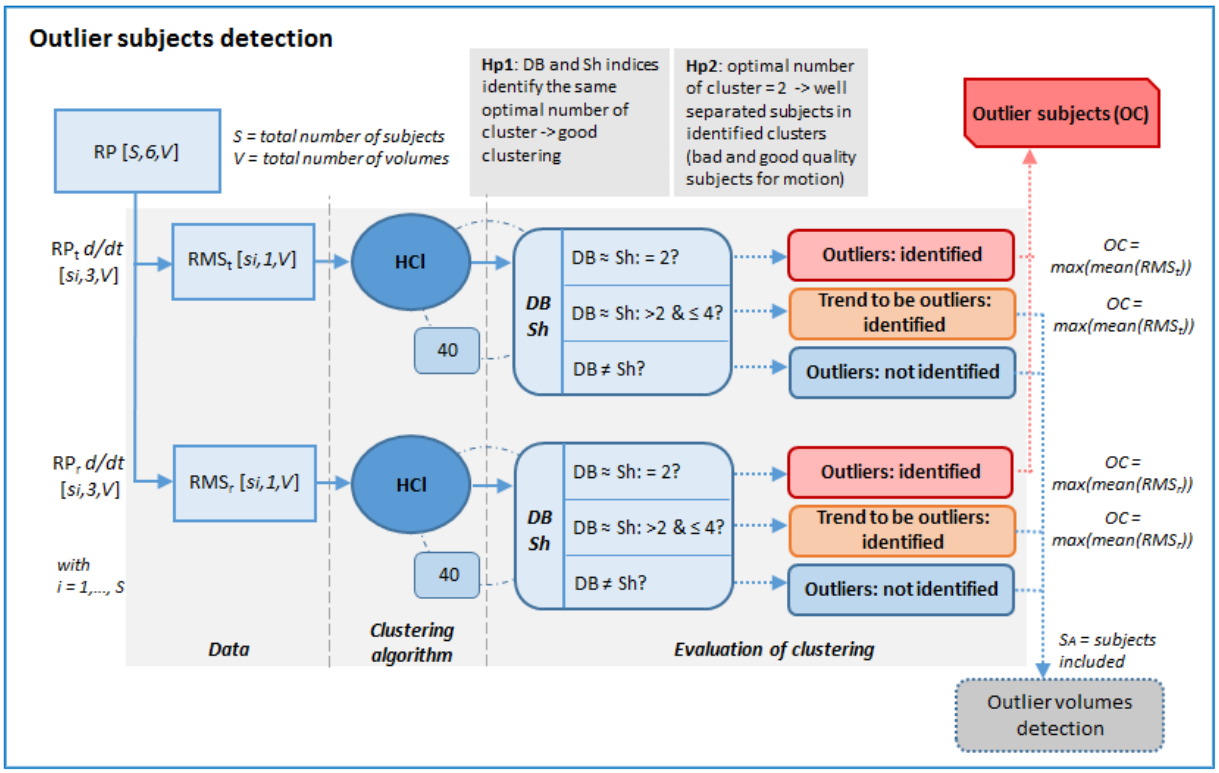


Figure 4.1 "Outlier subjects detection": workflow of clustering method

All the subjects, excepted the outliers, (subsample S_A) were considered as inputs for subsequent "outlier volumes detection" assessment.

Mahalanobis distance

We compute the means of RP$_t$ *d/dt* [*S*, 3, *V*] and RP$_r$ *d/dt* [*S*, 3, *V*] along time axis (*V*), respectively. These measure were used to ensure a number of variables (i.e. x, y, z or pitch, roll, yaw) << number of observations (i.e. subjects). Multinormality assumption was verified (Mardia test - α = .05), and, if the result was significant, MD was obtained for each subject. We performed a chi-square test on squared MD (p = .05), to identify outlier subject.

Other methods: RMS, absolute displacements

In addition, as described in previous studies, different metrics were considered for each subject: RMS (Oakes et al., 2005; Siegel et al., 2014), RMS of first derivatives (Power 2011, Power 2014) and translation and rotation absolute displacements (Power et al., 2014; Marko Wilke, 2012) of realignment parameters. RMS and RMS d/dt were obtained as the root mean squared of realignment parameters, with rotational estimates converted to translational at radius of 50 mm, and first derivatives of them, (i. e. differences relative to the preceding volume), respectively (Power 2011, Power 2014). Absolute displacements for translation (and rotation) were obtained sum of the

absolute values of the x, y, and z (pitch, roll, yaw) estimates for a given volume (Power 2014).

To identify outliers, different user-dependent thresholds were defined for each of these metrics based on the previous literature (RMS: 1 and 1.5 mm; RMS d/dt: 0.1 and 0.05 mm; translational and rotational absolute displacements: 2 mm and 0.3 degree). Subjects exceeded these thresholds for 50%, or 25% session were identified as outliers (Churchill et al., 2012; Kirwan, Shrager, & Squire, 2009; Siegel et al., 2014).

4.3.2.2 Outlier volumes detection

Clustering method

First derivatives (*d/dt*) of the 3 translation (RP_t) and the 3 rotation (RP_r) parameters of each single subject comprised in S_A subsample were considered separately as inputs in this level. See Figure 4.2 for a schematic representation of the "outlier volumes detection" workflow in the clustering method. For each subject s_i of each dataset:

$$RP_{t,i}\,[V, 3] \rightarrow RP_{t,i}\,d/dt\,[V, 3]$$

$$RP_{r,i}\,[V, 3] \rightarrow RP_{r,i}\,d/dt\,[V, 3] \qquad \text{with } i = 1, \ldots, A$$

$RP_{t,i}\,d/dt$ ed $RP_{r,i}\,d/dt$ of each single subject were used as separate inputs of two HCl algorithm, producing two dendrograms. An agglomerative HCl algorithm was chosen as clustering algorithm, because we hypothesized that clusters corresponding to corrupted or not corrupted volumes can differ in size (i.e. the number of volumes corrupted by motion << the number of volumes not corrupted by motion).

For validation phase, each HCl was re-iterated for 40 times and 3 different measures were obtained: CPCC (i.e. $CPCC_{t,i}$ and $CPCC_{r,i}$), as measure of how appropriately the data has been clustered (goodness of fit of clustering), and Sh (i.e. $Sh_{t,i}$ and $SH_{r,i}$) and DB (i.e. $DB_{t,i}$ and $DB_{r,i}$), to identify the optimal number of clusters in dataset). Then, the optimal number of clusters for the two different inputs (i.e. $RP_{t,i}\,d/dt$ ed $RP_{r,i}\,d/dt$) was identified for Sh and DB based on the highest number of iterations producing that clustering results.

CPCC index was used for the identification of the sub-sample of subjects to be scrubbed, while Sh and DB indices for the identification of volumes to be discarded in this sub-sample.

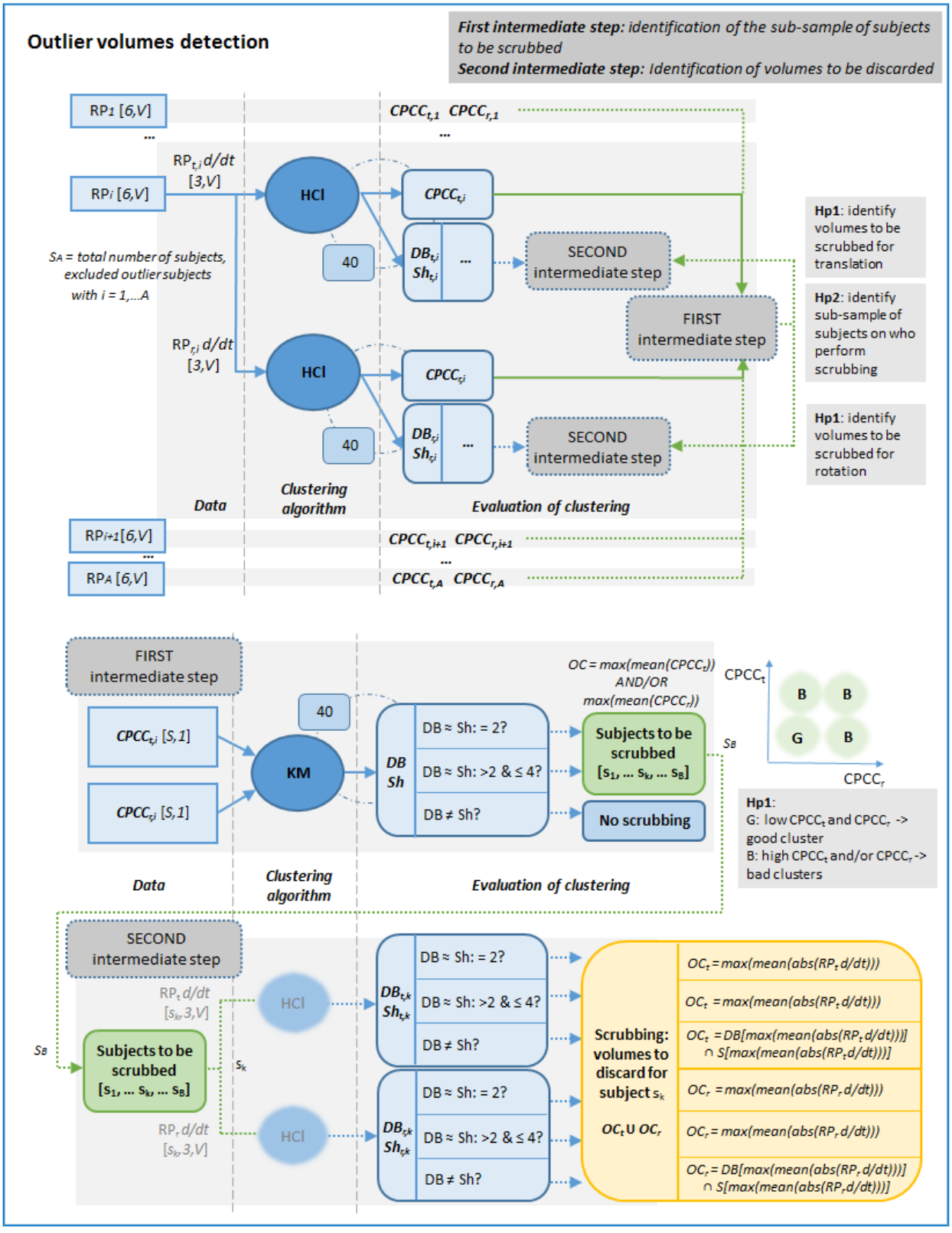


Figure 4.2 "Outlier volumes detection": workflow of clustering method

A. First intermediate step: Identification of the sub-sample of subjects to be scrubbed

To identify the subjects who required scrubbing, $CPCC_t$ (S_A-*by-1* vector data) and $CPCC_r$ (S_A-*by-1* vector data) of all subjects were used as inputs for a KM clustering. The choice of initial centroid has a great impact on final clustering result (Tan et al., 2006), so a user-dependent initialization was performed with the following coordinates: 2 cluster [0.75, 0.75; 0.95, 0.95], 3 clusters [0.75, 0.75; 0.95, 0.8; 0.8, 0.95], 4 clusters [0.75, 0.75; 0.75, 0.95; 0.95, 0.75; 0.95, 0.95] (Figure 4.2). For the validation phase, KM was re-iterated 40 times and a Sh and DB optimal number of cluster was defined based on the maximum number of iterations producing that clustering results. If the number of clusters of the two indices was different, no subjects to be scrubbed were identified in the datasets. Otherwise, if it was equal and comprised in the range 2 to 4, the clustering solutions were considered reliable and the data were partitioned in clusters according to the number of clusters identified. In this case, subjects in the cluster with the highest mean of $CPCC_t$ and/or maximum of mean of $CPCC_r$ were considered as inputs (subsample S_B) for subsequent second intermediate step.

B. Second intermediate step: Identification of volumes to be discarded

For each single subject comprised in the subsample S_B, Sh ($Sh_{t,i}$ and $SH_{r,i}$) and DB ($DB_{t,i}$ and $DB_{r,i}$) indices, identified during previous HCl, were considered for validation phase. If the number of clusters of Sh and DB indices was equal and comprised between 2 to 4, clustering solutions were considered reliable and the data were partitioned in clusters according to the number of clusters identified: volumes in cluster with the highest mean of absolute value of $RP_{t,i}$ *d/dt* (similarly for the $RP_{r,i}$ *d/dt*) were considered to be discarded from the analyses ("bad cluster"). If it was different, two separated clustering results were obtained. . For each of these solutions, volumes in cluster with the highest mean of absolute value of RPt,i d/dt (similarly for the RPr,i d/dt) were identified as "bad clusters". Then, the volumes comprised in the intersection between the two "bad clusters" were discarded. The realignment parameters of each discarded volume were replaced with the mean of realignment parameters of the previous and subsequent volumes (interpolation) in the subsample S_B. They constituted the realignment parameters of new subsample S_{Bn}

C. Re-iteration of clustering method

To determine the presence of possible residual volume to be additionally discarded in the subsample S_{Bn}, "outlier volumes detection", as described above, was repeated again on sample S_A, with the replacement of S_B with S_{Bn} subsample. The first intermediate step produced a new sub-sample of subjects to be scrubbed $S_{C..}$ The second intermediate step for the identification of additional volumes to be discarded were applied only to subject comprised in intersection between S_{Bn} and S_C.

Mahalanobis distance (MD)

For each subject of subsample S_B, we performed MD computation with the same input used for "outlier volumes detection" of clustering method (i.e. $RP_{t,i}$ *d/dt* ed $RP_{r,i}$ *d/dt*). In this case, the observations were the volumes. Multinormality assumption was verified (Mardia test - $\alpha = .05$). Squared MD for each volume was computed. A chi-square test on squared MD of each subject was performed: volumes characterized by squared MD above the identified threshold were considered outliers and considered to be discarded.

Other methods: framewise displacement (FD)

FD (Power et al., 2011), was implemented for each time point as the sum of the absolute values of the first derivatives of six realignment parameters, after converting rotational parameters to mm with a sphere of 50 mm radius. All volumes, whose FD exceeded a user-dependent threshold (FD02 = 0.2 mm, FD05 = 0.5 mm, FD08 = 0.8 mm, and FD1 = 1 mm) (Power, Anne, Snyder, Schlaggar, & Petersen, 2013; Power, Barnes, Snyder, Schlaggar, & Petersen, 2012; Siegel et al., 2014) were considered to be discarded.

Temporal masks and GLM analyses

For each single-subject, temporal masks (Siegel et al., 2014) (1-by-V vector) with 0 for outlier volumes and 1 for the remaining volumes were produced for clustering, MD, FD (i.e. FD02, FD05, FD08 and FD1) methods. To preserve temporal continuity of BOLD signal changes, segments of remaining volumes (mask = 1) lasting one volume were codified as 0. Finally, subjects with less than 75% of remaining volumes (mask = 1) were excluded from the study sample (Siegel et al., 2014).

Separate GLMs (SPM8) for each single-subject and for each temporal mask (Siegel et al., 2014), obtained by the different methods (clustering, MD, FD at different thresholds), were performed. To this aim, the pre-processed functional data were entered in to a single-subject analysis (K. J. Friston et al., 1995) with regressors of interest convolved with hemodynamic response (K. J. Friston et al., 1995), and as nuisance regressor the relevant temporal mask. The formers were specific for each dataset (VF task: phonemic fluency, semantic fluency, rest; visual task: faces, houses; olfactory task: odour, rest). Pre-whitening correction for autocorrelation (Monti, Simpson, & Forest, 2011; Woolrich, Ripley, Brady, & Smith, 2001) was performed. For each subject con-images of conditions of interest were produced (Pernet, 2014) (VF task: phonemic and semantic fluency > rest; visual task: faces vs. houses; olfactory task: odour > rest).

The relevant con-images for each participant were entered into a group analysis (random-effects analysis). At level-group, t-maps were produced and thresholded with the same voxel-level and cluster-level thresholds, for all the used methods in each single dataset (Figure 4.3).

To perform a GLM residual analysis (step 4.4.1), "outlier volumes detection" assessment, without the application of the first intermediate step for clustering, was applied to outlier subjects, excluded with "outlier subjects detection".

4.4 Materials and Methods: statistical analyses for comparison between clustering and other methods

4.4.1 Outlier subjects detection

Residual analysis

To verify the goodness of "outlier subjects detection" in clustering and MD methods, we hypothesized that subjects with un-modelled high head-movements in GLM, presented high squared single-subject GLM residuals (Diedrichsen & Shadmehr, 2005; Siegel et al., 2014), and, therefore, in comparison to other subjects, they can be classified also as outlier subjects for squared residuals. To this aim, for each subject of the sample S (outliers subjects included), we computed a whole-brain mean of the squared residuals obtained by single-subjects GLM computed with temporal mask of FD. To detect possible outlier subjects in squared residuals, a generalized extreme studentized deviate (ESD) test (Rosner, 1983) (α=0.05, upper bound on the suspected number of outliers = 10% of *S*) was implemented. Then, we observed if the outlier subjects identified for "outlier subjects detection" with previous methods (clustering, MD, FD at different thresholds) based on realignment parameters were the same outlier subjects identified with this residual analysis.

4.4.2 Outlier volumes detection

Comparison of temporal masks

To assess differences in the discarded volumes among clustering method, MD and FD (at different thresholds), we performed two evaluations. The first evaluation was based on Hamming distance (HD) of the temporal masks between clustering method and other methods. HD is a measure ranging from 0 (identical) to 1 (completely dissimilar) and indicates the distance between binary

vectors. The second evaluation was based on a one-way ANOVA with repeated measures performed on the numbers of discarded volumes (i.e. within-subjects factor) for each method (i.e. between-subjects factor) (SPSS - Version 17.0. Chicago: SPSS Inc.). Descriptive statistics were derived for HDs and number of discarded volumes.

Residual analysis

For each subject of the sample SA (outliers subjects excluded), we computed a whole-brain mean of the squared residuals obtained by single-subjects GLM computed with temporal mask of clustering method, MD, and FD. One repeated measures one-way ANOVA was performed on whole-brain mean squared residuals of single-subject GLM (i.e. within-subjects factor), obtained from the different methods (i.e. between-subjects factor).

4.5 Results

Results of GLM single-subject and group analyses for all temporal masks showed significant activations for contrasts of interest (Figure 4.3).

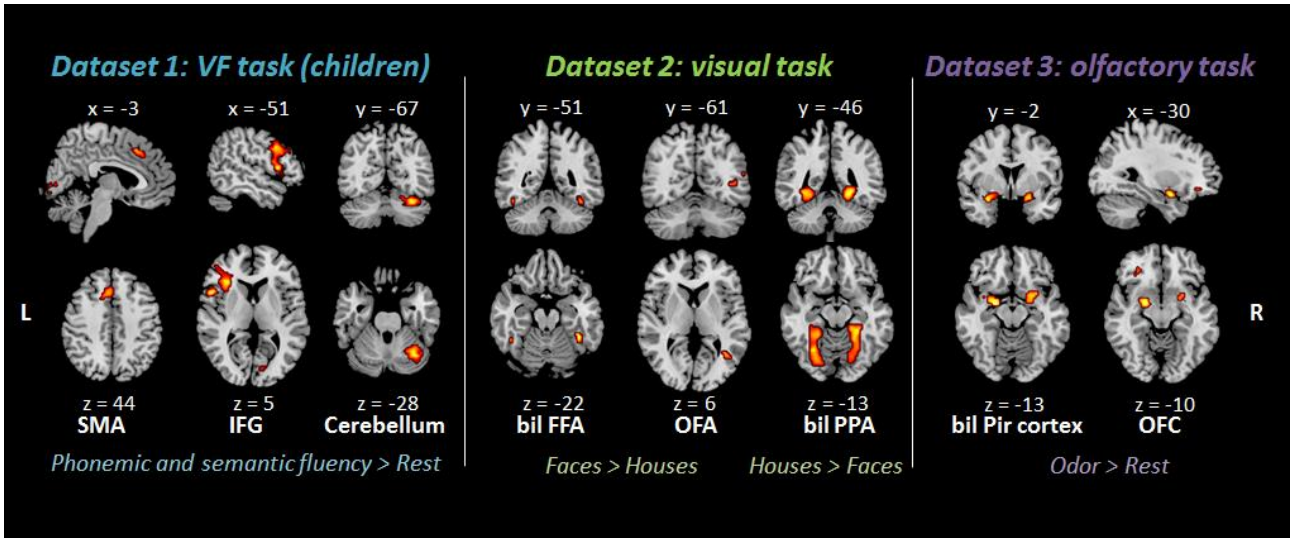


Figure 4.3. Statistical t-maps of random-effect analyses for contrast of interest using clustering temporal mask as regressor of no interest in single-subject GLM. Inferences at group-level were drawn from activated clusters obtained with specific voxel-level threshold and cluster-level threshold: dataset 1 $p < 0.05$ FDR, minimum 5 voxels for cluster; dataset 2: faces > houses $p < 0.01$ FDR, minimum 5 voxels for cluster, houses > faces: $p < 0.01$ FWE, minimum 5 voxels for cluster; datasets 3 $p < 0.05$ FDR, minimum 5 voxels for cluster. Abbreviations: FDR = false discovery rate, FWE = family-wise error; SMA = supplementary motor area; IFG = inferior frontal gyrus; FFA = fusiform face area; OFA object face area; bil = bilateral; Pir = piriform; OFC = orbitofrontal cortex.

4.5.1 Outlier subjects detection

For each dataset, descriptive statistics RMS, RMS d/dt and absolute displacements of realignment parameters were obtained (Table 4.2).

	Dataset1: VF task	**Dataset2: visual task**	**Dataset3: olfactory task**
Descriptive statistics	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***
Ab dis T	0.6398 (0.4336)	0.2398 (0.1616)	0.8168 (1.4393)
Ab dis R	0.0136 (0.0115)	0.0330 (0.0020)	0.0103 (0.0148)
RMS	0.1814 (0.1268)	0.0699 (0.0514)	0.2492 (0.4465)
RMS d/dt	0.0117 (0.0071)	0.0115 (0.0059)	0.0218 (0.0088)

Table 4.2 Descriptive statistics on realignment parameters and derived measures for each dataset. Abbreviations: SD = standard deviation; RMS = root mean squared; Ab dis = Absolute displacement; T = translation; R = rotation; VF = verbal fluency.

Outlier subjects identified with the different methods (clustering, MD, RMS, RMS d/dt and absolute displacements) were reported in Table 4.3.

			Dataset1: VF task		**Dataset2: Visual task**		**Dataset3: Olfactory task**	
Outlier subjects detection	**thr**	**session**	**outlier subjects**	**tending to be outlier**	**outlier subjects**	**tending to be outlier**	**outlier subjects**	**tending to be outlier**
Clustering	-	-	**ID7**	ID12; ID17	**ID14**	n.o.	**ID12**	n.o.
MD	-	-	**ID7**; ID17	-	**ID14**	-	**ID12**	-
Ab dis T (mm)	> 2	50%	n.o.	-	n.o.	-	ID8; ;ID11; **ID12**	-
		25%	ID10	-	n.o.	-	ID8; ID11; **ID12**; ID22	-
Ab dis R (°)	> 0.3	50%	n.o.	-	n.o.	-	n.o.	-
		25%	n.o.	-	n.o.	-	n.o.	-
RMS (mm)	> 1.5	50%	n.o.	-	n.o.	-	ID11; **ID12**	-
		25%	n.o.	-	n.o.	-	ID11; **ID12**	-
	> 1	50%	n.o.	-	n.o.	-	ID11; **ID12**	-
		25%	n.o.	-	n.o.	-	ID11; **ID12**; ID22	-
RMS d/dt (mm)	> 0.1	50%	n.o.	-	n.o.	-	n.o.	-
		25%	n.o.	-	n.o.	-	n.o.	-
	> 0.05	50%	n.o.	-	n.o.	-	n.o.	-
		25%	n.o.	-	n.o.	-	**ID12**	-
ResMS			**ID7**; ID17		ID16		**ID12**	

Table 4.3 "Outlier subjects detection": comparison of quality data assessments for identification of outlier subjects on motion parameters. **Abbreviations**: thr = user-dependent threshold; Ab dis = Absolute displacement, T = translation, R = rotation, ResMS = whole-brain mean squared residuals, n.o. = no outliers.

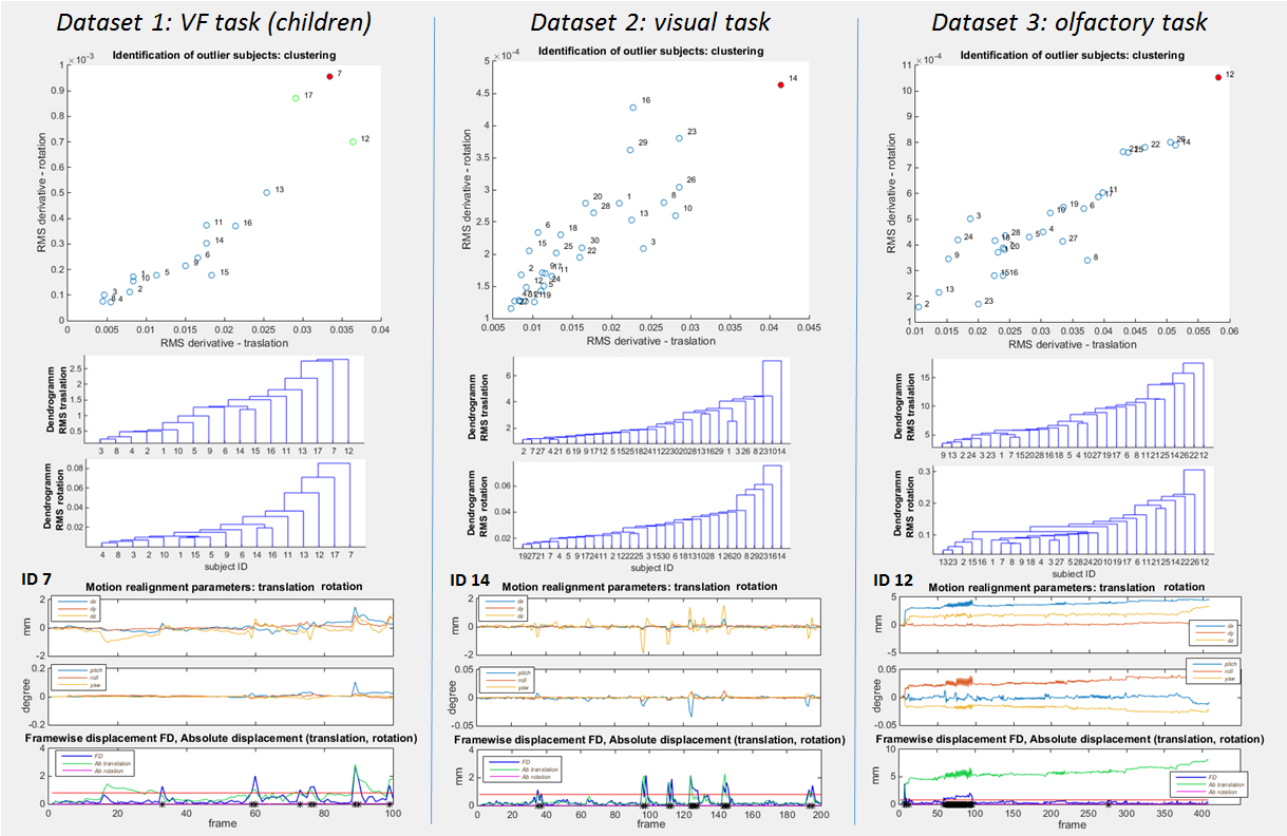


Figure 4.4 "Outlier subjects detection": results of outlier subjects obtained with clustering and MD methods in the 3 datasets.

A qualitative evaluation of the movements of outlier subjects showed that, the auto-calibrated method (clustering and MD) and partially the RMS d/dt measures mainly identified subjects with spike movements, while other metrics (RMS and absolute displacements) characterized subjects with slight drifts (Figure 4.5).

First dataset (verbal fluency task; high head movements unrelated to the execution of the task): MD method identified two outlier subjects (ID7; ID17); the same subjects were identified as outlier and as tending to be outlier in clustering method, respectively (Table 4.3). Importantly, the ESD test on GLM squared residuals characterized as outliers the same detected subjects (ID7; ID17). Among the user-dependent threshold methods, absolute displacements in translation identified one different outlier subject (ID10), while the remaining metrics (RMS, RMS d/dt) did not detect any outliers.

Second dataset (visual task; low head movements): auto-calibrated methods (clustering and MD algorithms) identified the same outlier subjects (ID14). Interestingly, the ESD test on GLM squared residuals characterized a different outlier (ID16). However, as showed in Figure 4.6, the ID14 was characterized also by high squared residuals. No user-dependent threshold methods was able to detect any outlier.

Third dataset (olfactory task; high head movements related to the execution of the task): auto-calibrated methods (clustering and MD algorithms) identified the same outlier (ID12). ESD test on GLM squared residuals confirmed this result. This subject was also identified as outlier by RMS d/dt using the most restrictive threshold (0.05mm on 25% of session). Importantly, the user-dependent threshold methods, absolute displacement in translation and RMS, identified from 2 to 4 outliers, including the ID12 subject.

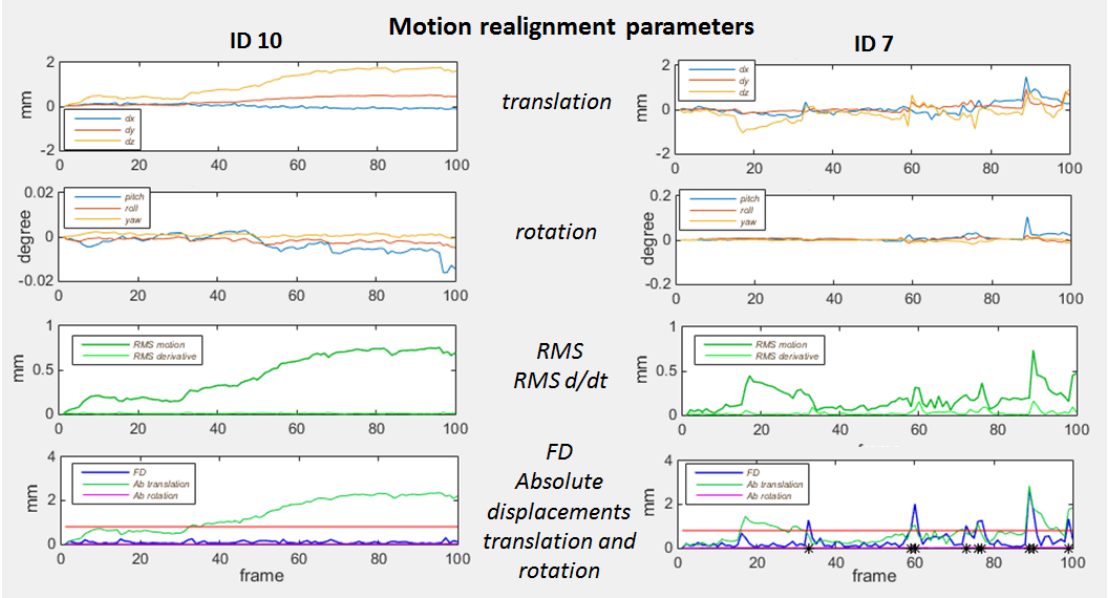


Figure 4.5 Dataset 1: movement parameters of outlier subject identified with RMS and absolute displacement (on the left, ID10) and with clustering and MD procedures (on the right, ID7).

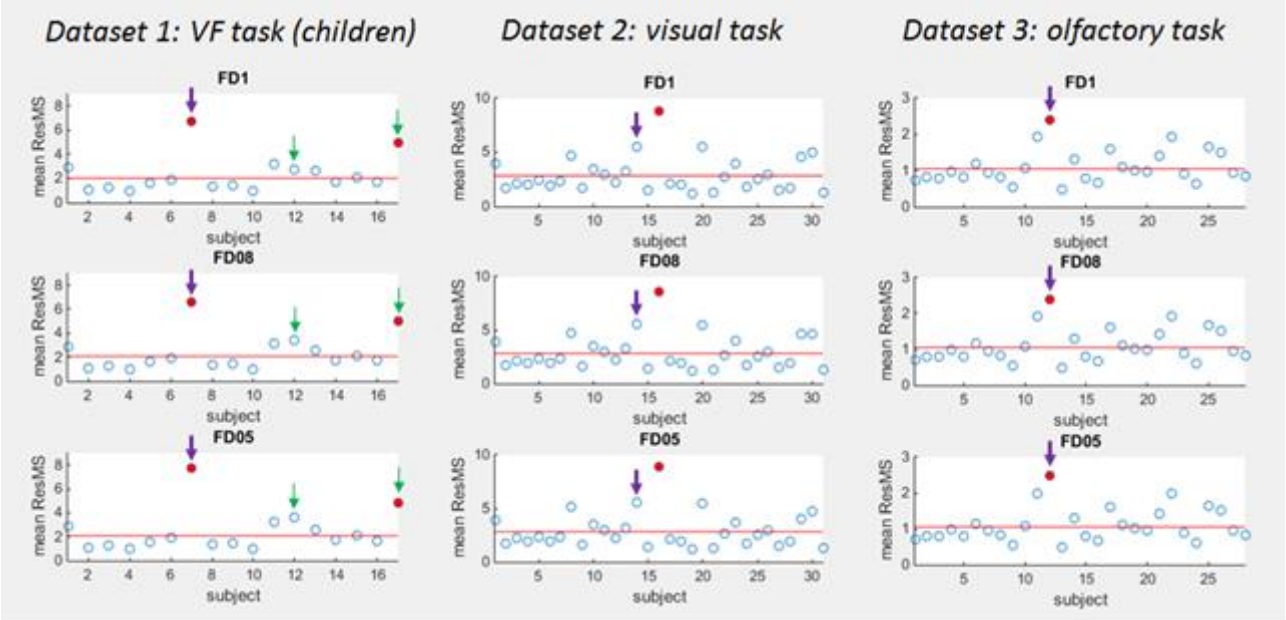


Figure 4.6 "Outlier subjects detection": single-subject GLM whole-brain residuals plots to identify possible mis-modelling and outliers subjects. Outlier subjects identified with clustering method (big arrows) and with ESD test on squared residuals (red point). Subjects tending to be outlier identified with clustering method (small arrows).

4.5.2 Outlier volumes detection

4.5.2.1 Temporal masks

FD02 determined a large number of subjects with less than 75% of remaining volumes (mask = 1) (VF task: 8/17 subjects excluded; visual task: 7/31 subjects excluded; olfactory task: 10/28 subjects excluded), who were excluded from the study sample: no GLMs were performed with this temporal mask due to the small the sample size.

Descriptive statistics of HDs and number of discarded volumes of temporal masks for each dataset were reported in Table 4.4.

n. of discarded volumes (%)	Statistics	cl	MD	FD1	FD08	FD05	FD02
Dataset 1: VF task	*Mean*	0.750	1.500	0.563	0.813	3.563	31.750
	SD	1.065	1.897	1.209	1.515	5.379	30.493
Dataset 2: visual task	*Mean*	1.400	3.300	0.235	0.667	1.883	15.250
	SD	2.191	3.547	0.610	1.093	2.648	16.345
Dataset 3: olfactory task	*Mean*	0.236	0.527	0.472	0.735	2.233	20.389
	SD	0.657	1.427	0.654	1.195	2.810	13.441

Temporal mask: HDs	Statistics	cl-MD	cl-FD1	cl-FD08	cl-FD05	cl-FD02
Dataset 1: VF task	*Mean*	0.009	0.006	0.008	0.036	0.313
	SD	0.012	0.010	0.012	0.051	0.305
Dataset 2: visual task	*Mean*	0.028	0.014	0.015	0.024	0.146
	SD	0.031	0.021	0.021	0.028	0.163
Dataset 3: olfactory task	*Mean*	0.007	0.008	0.012	0.041	0.411
	SD	0.019	0.014	0.019	0.052	0.271

Table 4.4 "Outlier volumes detection": descriptive statistics of HDs and number of discarded volumes for the different temporal masks for each dataset. **Abbreviations**: SD = standard deviation, cl = clustering method; MD = Mahalanobis distance method; FDxx = method with FD thresholded at xx (i.e. 1, 0.8, 0.5, 0.2 mm); HD = Hamming distance.

HDs, the distance between the temporal masks of clustering and other methods, were small (range: 0.006-0.041). despite of HDs between clustering and FD02 methods showed the highest dissimilarity (HD > 0.1). Similar results were obtained for number of discarded volumes, confirming that FD threshold equal to 0.2 mm was too conservative in these datasets. Moreover, we can observe that a lower number of volumes was discarded in clustering method compared to MD.

The ANOVA on number of discarded volumes showed no significant differences between temporal masks obtained with clustering and other methods, excepted for temporal mask at FD02 in all datasets and FD05 and FD08 for olfactory task (Table 4.5).

Dataset 1: VF task (children)						
	cl	**FD1**	**FD08**	**FD05**	**FD02**	**MD**
cl	-	ns	ns	ns	0.016	ns
FD1		-	ns	ns	0.015	ns
FD08			-	ns	0.014	ns
FD05				-	0.01	ns
FD02					-	0.021
MD						-
	F(1.022, 15.330) = 16.247; p = 0.001					
Dataset 2: visual task						
	cl	**FD1**	**FD08**	**FD05**	**FD02**	**MD**
cl	-	ns	ns	ns	0.001	ns
FD1		-	ns	ns	0.000	ns
FD08			-	ns	0.000	ns
FD05				-	0.000	ns
FD02					-	0.003
MD						-
	F(1.120, 32.472) = 22.097; p = 0.000					
Dataset 3: olfactory task						
	cl	**FD1**	**FD08**	**FD05**	**FD02**	**MD**
cl	-	ns	0.018	0.004	0.000	ns
FD1		-	ns	0.005	0.000	ns
FD0.8			-	0.006	0.000	ns
FD0.5				-	0.000	0.006
FD0.2					-	0.000
MD						-
	F(1.027, 26.694) = 62.221; p = 0.001					

Table 4.5 "Outlier volumes detection": results of repeated measures ANOVA performed on the numbers of discarded volumes for each method. Greenhouse-Geisser correction for sphericity and post-hoc tests using the Bonferroni correction for multiple comparison were applied. Abbreviations: cl = clustering method; MD = Mahalanobis distance method; FDxx = method with FD thresholded at xx (i.e. 1, 0.8, 0.5, 0.2 mm).

4.5.2.2 Residual analysis

Mean and standard deviation of whole-brain mean squared residuals obtained with clustering method were smaller than those obtained with other methods, except for some FD thresholds in which were similar (Table 4.6). Repeated measures one-way ANOVA on whole-brain mean squared residuals (dataset1: $F(1.22, 18.326) = 0.550$, $p < 0.502$; dataset2: $F(1.596, 46.273) = 0.338$, $p < 0.666$; dataset3: $F(1.831, 47.593) = 0.485$, $p < 0.602$) did not show significant differences among the different methods.

		cl	MD	FD1	FD08	FD05
VF task: **ResMS**	Mean	2.037	2.056	2.033	2.075	2.074
	SD	1.047	1.075	1.057	1.098	1.089
visual task: **ResMS**	Mean	2.815	2.859	2.833	2.827	2.829
	SD	1.614	1.620	1.627	1.599	1.621
olfactory task: **ResMS**	Mean	1.051	1.052	1.050	1.050	1.053
	SD	0.393	0.396	0.394	0.395	0.405

Table 4.6 "Outlier volumes detection": mean and standard deviation (SD) of single-subject GLM whole-brain mean squared residuals (ResMS) (outlier subjects excluded).

4.6 Discussion

One of the main limitation of the most used head movements assessments in fMRI datasets is the application of user-dependent thresholds (Power et al., 2015; Satterthwaite et al., 2013) To overcome this drawback, we implemented two auto-calibrated procedures, clustering and MD. The auto-calibrated and user-dependent threshold methods were applied at sample and single-subject level in 3 different datasets characterized by high head movements unrelated and related to the execution of the task and by low head movements.

At sample-level, our results showed that both auto-calibrated methods identified the same outlier subjects, but one in all the 3 datasets. These results were supported by GLM residuals analysis at group level. Notably, the most part of these outlier subjects were not detected by user-dependent threshold methods (i.e. RMS, RMS d/dt, and absolute displacements), except in the sample of the participants showing the highest movements related to task (dataset 3). Importantly, at single-subject level, the clustering method discarded a small number of volumes, but it decreased GLM residuals in the dataset with high movements (dataset 1). This pattern was not assessed in dataset 3 due to task related movements. In the dataset with lowest movements (dataset 2) the number of discarded volumes was in line with other methods, but residuals were smaller.

Three major points characterized the clustering and MD implementation. First, the computational approach between auto-calibrated and user-dependent thresholds methods is different: in the former, all the subjects are grouped together and the variability of a single subject is considered in relation to other subjects of sample, whereas, in the latter methods, each subject is individually analysed. This ensures the reduction of variance in data for motion at both sample and single-subject levels. Secondly, translation and rotation parameters are used as separate inputs. This approach is chosen for two reasons. This avoids a transformation of rotation parameters from degrees/radians in mm using a priori definition of a the average distance from the origin around which rotation is

performed (Power et al., 2012; Marko Wilke, 2014); for example this approximation could introduce a bias in children cohort. Moreover, it allows a more complete characterization of the motion process, through a comprehensive approach, taking into account all parameters, but separately (translation and rotation) (Marko Wilke, 2014). Third, the clustering and MD methods had as inputs the RMS of first derivatives or first derivatives over time of realignment parameters. Notably, these relative metrics, obtained from the differences between consecutive volumes highlights high displacement (i.e. spike movements), while minimizes the effect of drift movements (Lemieux, Salek-haddadi, & Lund, 2007; Marko Wilke, 2014), that are detected in absolute metrics (i.e. RMS and absolute displacements) (Power et al., 2012).

4.6.1 Outlier subjects detection

For "outlier subjects detection", the outlier subjects identified with both auto-calibrated methods presented relevant high movements, also confirmed through a visual inspection of realignment parameters in these subjects (Figure 4.4). MD detected an additional outlier subjects, that was identified as tending to be outlier in the clustering method. This additional level of the identification of subjects tending to be outliers in clustering method allows a conservative approach: in this case is possible to verify after GLM computation if the exclusion of these subjects is necessary or not.

The check for outliers in residuals values after GLM estimation was used by other method as ArtRepair (Mazaika, Whitfield-Gabrieli, & Reiss, 2007). Notably, residual values could include other sources of signal of no interest in addition to movements. In our study, GLM residual analyses were used only to confirm the results obtained during the initial data-driven assessment. Outlier subjects detected in all the datasets presented the highest (VF and olfactory tasks) or high (visual task) whole-brain mean squared residuals in single-subject analyses (Figure 4.5). Moreover, outlier test (ESD test) performed on mean squared residuals identified as outliers the same outlier subjects or tending to be outlier subjects detected with the clustering and MD methods for VF task and olfactory task (Figure 4.5). In visual task, the different outlier subject detected with this test could be due to other sources of signal variability not fully modelled in single-subject GLM (Diedrichsen & Shadmehr, 2005).

When RMS d/dt was thresholded (i.e. with user-dependent threshold), it allows to identify only one outlier subject in the third dataset. If the same measure was used as input in the clustering method it allows to identify an outlier for each dataset, also reported in residual analyses. This strengthen the use of an algorithm that auto-calibrates the identification of the outlier subject based on the sample

population.

The other metrics, RMS and absolute displacements, identified different outlier subjects compared to clustering and MD methods (Table 4.3). However, these different results can be justified, considering that displacements relative to a single reference volume (i.e. RMS and absolute displacements), detect both spike movements and drifts (Power et al., 2012) (see Figure 4.5 for a comparison).

4.6.2 Outlier volumes detection

At single-subject level, we observed different patterns between number of discarded volumes and GLM residuals among the high and low movements datasets. In particular, clustering method discarded a small number of volumes, and it slight decreased GLM residuals in the dataset with high movements (dataset 1), while, at similar number of discarded volumes, it slight reduced GLM residuals in the dataset with low movements (dataset 2). This last pattern was not observed in the third dataset due to task-related movements.

All these results taken together highlighted two positive features of clustering method: a higher preservation of temporal information associated to a robust GLM computation. The re-iteration of the assessment was performed to ensure a conservative approach in the detection of possible outlier volumes.

For " outliers volumes detection", significant differences were identified in number of volumes between temporal masks obtained with a threshold of FD equal to 0.2 mm and all other methods for all datasets (Table 4.5). Similarly, high value of HDs were identified between temporal masks of clustering and FD thresholded at 0.2 mm (Table 4.4). This suggests that this temporal mask was too restrictive in these datasets.

References

Andersson, J. L. R., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling Geometric Deformations in EPI Time Series, 919, 903–919.

Bertolino, N., Ferraro, S., Nigri, A., Bruzzone, M. G., & Ghielmetti, F. (2014). A neural network approach to fMRI binocular visual rivalry task analysis. PloS One, 9(8), e105206.

Black, P. E., & Pieterse, V. (2006). Manhattan distance. In Dictionary of Algorithms and Data Structures (31 May 200).

Bullmore, E. T., Brammer, M. J., Curtis, V. A., Morris, R. G., Williams, S. C. R., Sharma, T., & Mcguire, P. K. (1999). Methods for Diagnosis and Treatment of Stimulus-Correlated Motion in Generic Brain Activation Studies Using fMRI. Human Brain Mapping, 48, 38–48.

Christodoulou, A. G., Bauer, T. E., Kiehl, K. a., Feldstein Ewing, S. W., Bryan, A. D., & Calhoun, V. D. (2013). A quality control method for detecting and suppressing uncorrected residual motion in fMRI studies. Magnetic Resonance Imaging, 31(5), 707–717.

Churchill, N. W., Oder, A., Abdi, H., Tam, F., Lee, W., Ween, J. E., … Strother, S. C. (2012). Optimizing Preprocessing and Analysis Pipelines for Single- Subject FMRI. I. Standard Temporal Motion and Physiological Noise Correction Methods, 33(3), 609–627.

Churchill, N. W., Spring, R., Afshin-pour, B., Dong, F., & Strother, S. C. (2015). An Automated, Adaptive Framework for Optimizing Preprocessing Pipelines in Task- Based Functional MRI. PLoS ONE, 1–25.

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2), 224–227.

Diedrichsen, J., & Shadmehr, R. (2005). Detecting and adjusting for artifacts in fMRI time series data. NeuroImage, 27(3), 624–634.

Ekstrom, J. (2011). Mahalanobis' distance beyond normal distributions ¨. UCLA Stat (preprint), 1–16.

Evans, A., Collins, L., Paus, C. H. T., MacDonald, D., Zijdenbos, A., Toga, A., … Mazziota, J. (1997). A 3d probabilistic atlas of normal human neuroanatomy. In Third International Conference on Functional Mapping of the Human Brain (pp. 5, S349).

Field, A. S., Yen, Y., Burdette, J. H., & Elster, A. D. (2000). False Cerebral Activation on BOLD Functional MR Images : Study of Low-amplitude Motion Weakly Correlated to Stimulus. Am J Neuroradiol, 21, 1388–1396.

Filzmoser, P. (2004). A multivariate outlier detection method. In Proceedings of the Seventh International Conference on Computer Data Analysis and Modeling (pp. 18–22).

Friston, K. J., Holmes, a P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. NeuroImage.

Friston, K., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-Related Effects in fMRI Time-Series, 35(3), 346–355.

Harrison, A. H., & Connolly, J. F. (2013). Neuroscience and Biobehavioral Reviews Finding a way in : A review and

practical evaluation of fMRI and EEG for detection and assessment in disorders of consciousness. Neuroscience and Biobehavioral Reviews, 37(8), 1403–1419.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage, 17(2), 825–841.

Johnson, R. A., & Wichern, D. W. (2002). Applied Multivariate Statistical Analysis (5th editio). Prentice Hall, New York.

Johnstone, T., Walsh, K. S. O., Greischar, L. L., Alexander, A. L., Fox, A. S., Davidson, R. J., & Oakes, T. R. (2006). Motion Correction and the Use of Motion Covariates in Multiple-Subject fMRI Analysis. Human Brain Mapping, 27(10), 779–788.

Kaufman, L., & Rousseeuw, P. J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. (2003rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Kirwan, C. B., Shrager, Y., & Squire, L. R. (2009). Medial temporal lobe activity can distinguish between old and new stimuli independently of overt behavioral choice. Proc Natl Acad Sci U S A, 106(34), 14617–14621.

Kovács, F., Legány, C., & Babos, A. (2005). Cluster Validity Measurement Techniques. In Proceedings of the 6th International Symposium of Hungarian Researchers on Computational Intelligence, Budapest (pp. 18–19).

Lemieux, L., Salek-haddadi, A., & Lund, T. E. (2007). Modelling large motion events in fMRI studies of patients with epilepsy. Magn Reson Imaging., 25(6), 894–901.

Lundström, J. N., Gordon, A. R., Alden, E. C., Boesveldt, S., & Albrecht, J. (2010). Methods for building an inexpensive computer-controlled olfactometer for temporally-precise experiments. International Journal of Psychophysiology, 78(2), 179–189.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In Proceedings of the National Institute of Sciences of India (p. 2(1): 49–55).

Mardia, K. V. (1980). Tests of univariate and multivariate normality. In P. R. Krishnaiah (Ed.), Handbook of statistics (pp. 1:279–320). Amsterdam: North Holland.

Mazaika, P., Whitfield-Gabrieli, S., & Reiss, A. (2007). Artifact Repair for fMRI Data from High Motion Clinical Subjects. In presentation at Human Brain Mapping conference.

Monti, M. M., Simpson, S. L., & Forest, W. (2011). Statistical analysis of fMRI time-series : a critical review of the GLM approach. Frontiers in Human Neuroscience, (March), 5–28.

Nigri, A., Ferraro, S., Bruzzone, M. G., Nava, S., D'Incerti, L., Bertolino, N., … Lundström, J. N. (2016). Central olfactory processing in patients with disorders of consciousness. European Journal of Neurology, 23(3), 605–612.

Nigri, A., Ferraro, S., D'Incerti, L., Critchley, H. D., Bruzzone, M. G., & Minati, L. (2013). Connectivity of the amygdala, piriform, and orbitofrontal cortex during olfactory stimulation: a functional MRI study. Neuroreport, 24, 171–5.

Oakes, T. R., Johnstone, T., Walsh, K. S. O., Greischar, L. L., Alexander, A. L., Fox, A. S., & Davidson, R. J. (2005). Comparison of fMRI motion correction software tools. NeuroImage, 28, 529–543.

Pernet, C. R. (2014). Misconceptions in the use of the General Linear Model applied to functional MRI : a tutorial for junior neuro-imagers. Frontiers in Human Neuroscience, 8(1).

Power, J. D., Anne, K., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2013). Steps toward optimizing motion artifact removal in functional connectivity MRI ; a reply to Carp. NeuroImage, 76, 439–441.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage, 59(3), 2142–2154.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Anne, K., Church, J. A., … Miezin, F. M. (2011). Functional network organization of the human brain. Neuron, 72(4), 665–678.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect , characterize , and remove motion artifact in resting state fMRI. NeuroImage, 84, 320–341.

Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. NeuroImage, 105, 536–551.

Rasmussen, J. L. (1988). Evaluating outlier identification tests: Mahalanobis D Squared and Comrey D. Multivariate Behavioral Research, 23(2), 189–202.

Rendón, E., Abundez, I., Arizmendi, A., & Quiroz, E. M. (2011). Internal versus External cluster validation indexes. International Journal of Computers and Communications, 5(1).

Rosner, B. (1983). Percentage points for a generalized ESD many-outlier procedure. Technometrics, 25, 165–172.

Rousseeuw, P. J., & Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85(411), 633–651.

Saraçli, S., Dogan, N., & Dogan, I. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. Journal of Inequalities and Applications, 2013(203), 1–8.

Satterthwaite, T. D., Elliott, M. a., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., … Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. NeuroImage, 64(1), 240–256.

Schwager, S. J., & Margolin, B. H. (1982). Detection of multivariate outliers. The Annals of Statistics, 10, 943–954.

Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Jessica, A., Schlaggar, B. L., … Louis, S. (2014). Statistical Improvements in Functional Magnetic Resonance Imaging Analyses Produced by Censoring High-Motion Data Points. Human Brain Mapping, 35(5), 1981–1996.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. In Introduction to Data Mining (2006th ed., pp. 447–568). McGraw Hill.

Van Dijk, K. R. A., Sabuncu, M. R., & Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. NeuroImage, 59(1), 431–438.

Wilke, M. (2012). An alternative approach towards assessing and accounting for individual motion in fMRI timeseries. NeuroImage, 59(3), 2062–2072.

Wilke, M. (2014). Isolated Assessment of Translation or Rotation Severely Underestimates the Effects of Subject Motion in fMRI Data. PLoS ONE, 9(10).

Wilke, M., Holland, S. K., Myseros, J. S., Schmithorst, V. J., & Ball, W. S. (2003). Functional Magnetic Resonance Imaging in Pediatrics. Neuropediatrics, 34(5), 225–233.

Wilson, J. L., Jenkinson, M., De Araujo, I., Kringelbach, M. L., Rolls, E. T., & Jezzard, P. (2002). Fast, fully automated global and local magnetic field optimization for fMRI of the human brain. NeuroImage, 17(2), 967–976.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. NeuroImage, 14(6), 1370–86.

Zimmerman, D. W. (1994). A note on the influence of outliers on parametric and nonparametric tests. Journal of General Psychology, 121(4), 391–401.

Chapter 5

Effects of pre-processing pipeline on GLM analysis of task-based fMRI data

5.1 Introduction

In the recent years, there is a growth of literature for the impact of different combination and order of pre-processing steps on resting state fMRI data. Although it has been established that the chosen pre-processing steps may significantly affect fMRI results, it is not well understood how pre-processing choices interact among them and with fMRI experimental design in task-based fMRI.

Until few years ago, most fMRI analyses are performed under the implicit assumptions that either the results are relatively insensitive to the chosen set of pre-processing steps (e.g. the pre-processing "pipeline"), or that the standard pre-processing defined as default setting in in software packages reached near-optimal results. In recent years, it has been repeatedly shown that both standard and advanced data pre-processing choices may have significant effects (negative or positive) on the quality of results (e.g.: Murphy et al., 2009; Strother et al., 2004; Zhang et al., 2009). To draw reliable conclusions from fMRI results, it is thus necessary to evaluate the interactions of pre-processing and data analysis choices with statistical indices and parameters of interest obtained from analyses. This is particularly relevant for studies comprising samples with different demographic and clinical characteristics, where difference in variability of BOLD signals can be weaker, and head motion and physiological noise have a greater impact on fMRI data than control subjects.

In advanced pre-processing, additional steps were introduced to correct for subject head movement and physiological noise. Among the different methods for the correction of these noise signals, there is the regression of nuisance signals from fMRI data (Chapter 3).

For reduction of head movements, usually the regression of realignment parameters were performed (Friston et at., 1996; Johnstone et al., 2006). The effectiveness of this technique is also debated: motion parameters regression has been found to both improve reliability of fMRI results (Freire and Mangin, 2001), and reduce noise variance (Lund et al., 2005), particularly in young children compared to adults (Evans et al., 2010). Moreover, it has been shown that the impact of motion parameters regression is also related to patters of head motion: it is a more important denoising step in cases of task-correlated motion (Johnstone et al., 2006). However, other research has shown that applying it in block designs reduces the strength of task activation (Johnstone et al., 2006), and the procedure may remove both fMRI signal and artifact indiscriminately.

For reduction of physiological noise, the regression of signal obtained from simultaneous cardiac and respiratory recordings during fMRI acquisition, has been generally shown to improve results (Glover et al., 2000; Jones et al., 2008; RETROICOR). However, these physiological recordings are not always available and other alternative methods were proposed in resting state fMRI: regression of mean white matter (WM) and cerebrospinal fluid (CSF) signal obtained from segmentations of these regions in anatomical scan. This choice derived from the assumptions that neural activation is localized to grey matter and fluctuations in WM and CSF regions should primarily reflect signals of non-neural origin, such as cardiac and respiratory fluctuations (Behzadi et al., 2007).

Another open discussion point is related to which signals are necessary to regress out for improvement in BOLD signal detection. It was demonstrated that the removal of variance related to potentially confounding factors reduces reliability, pointing to the possibility that some structured noise may be improve the detectability of signal of interests (Yan et al. 2013, Varikuti et al., 2016).

Finally, different studies have shown that the effectiveness of motion and physiological noise correction depends on its order in the pre-processing pipeline (Jones et al., 2008; Vergara et al., 2016), and interactions with other parameters, such as temporal filtering and spatial smoothing (Shaw et al., 2003, Zhang et al., 2008). Temporal filtering and smoothing were suggest as steps after signal regression, because the signal regression can reintroduce frequencies previously suppressed by the bandpass filter (Hallquist et al., 2013, Weissenbacher et al., 2009) or distribute noise signal spatially (Jo et al., 2010), with a lower detection of small resting state networks (Vergara et al., 2016).

The impact of different type of nuisance regression and the interaction with other pre-processing steps suggested by recent advanced in resting-state fMRI is not well described for task-based fMRI.

On the basis of recent advances in resting state fMRI on the order of steps in pre-processing pipeline, the following order was tested: temporal detrend, nuisance regression, temporal filtering and smoothing (Chapter 3). Four different variation of pre-processing pipelines were proposed using different nuisance signals in regression:

- none
- regression of 6 realignment parameters (***RP***)
- regression of mean WM and CFS signals (***PhyP***)
- regression of 6 realignment parameters and mean WM and CFS signals (***RPPhysP***)

According to these elements, the aim of these studies was to compare the impact of different pre-processing pipelines on two important task-design issues:

i. strength of task contrast and goodness of model fitting in GLM using beta contrast estimates and residuals in ROIs;
ii. improvement in variability of BOLD signal using temporal SNR.

To perform this evaluation a sample of relative still healthy participants that undergone to a visual block task with visualization of faces and houses during fMRI were considered.

5.2 Materials and Methods: participants and acquisition details

The ***dataset 2*** of healthy participants described in Chapter 4 (4.2.3.2) was used to assess these hypotheses: participants were submitted to fMRI during the administration of a block visual task with faces and houses images. Acquisition parameters were previously described in details.

5.3 Materials and Methods: pre-processing and algorithm implementations

5.3.1 Data pre-processing

Analyses were performed using the following software and in-house scripts: MATLAB R2014b and SPM8. The following steps were applied to each single subject. A initial pre-processing comprised the following steps: realignment to the mean image, normalization to the MNI template (Evans et al., 1997) with the unified model framework procedure and resampling to 2-mm isotropic voxels. Moreover, T1 image was co-registered to fMRI mean image and normalized to MNI template. Temporal detrending was performed with a third degree polynomial function (Strother et al., 2004). All these steps were defined as "***standard pre-processing***" hereafter.

Sources of spurious signals, which were considered in the subsequent pre-processing pipelines and removed by linear regression, included: 1) six parameters obtained by rigid body head motion correction, 2) mean WM timecourse, 3) mean CSF timecourse. From fMRI data, WM and CSF signals were extracted as average time courses over the voxels included in the mask of thresholded segmentation of WM and CFS maps (voxel with intensity > 0.75), obtained from unified model of normalization step, respectively (Halliquist et al 2013; Weissenbacher et al., 2009).

After the standard pre-processing, by using different combination of these regression signals, the impact of four different ***pipelines*** were tested:

1. **wf_o**: no signal regression,
2. **wf**: realignment parameters regression (*RP*),
3. **nwf**: physiological parameters regression (*PhyP*: mean WM and CFS time courses)
4. **mwf**: realignment and physiological parameters regression (*RPPhyP*).

After signal regression, spatial smoothing (6-mm FWHM Gaussian isotropic kernel), temporal low-pass filtering with a hemodynamic response kernel, and temporal high-pass filtering with a cut-off period of 128 s (Friston et al., 1995) were applied. For each pre-processing pipeline we obtained the following four categories of fMRI data:

1. **swf_o**: no signal regression,
2. **swf**: realignment parameter regression (*RP*),
3. **snwf**: physiological parameter regression (*PhyP*)
4. **smwf**: realignment and physiological parameters regression (*RPPhyP*).

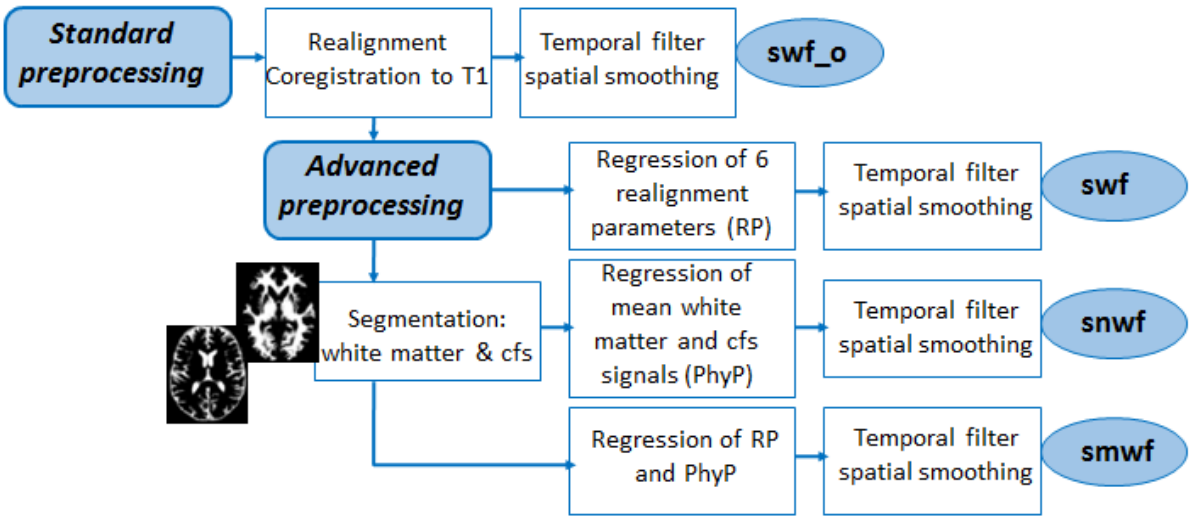


Figure 5.1 Scheme of pre-processing pipelines

5.3.2 Data analyses

Pipeline effects were measured on:

- GLM results using beta contrast estimations and squared residuals in ROIs (region of interest);
- fMRI data before and after the step comprising spatial smoothing using tSNR.

To perform these evaluations the following analyses were applied. For each pipeline, a whole-brain single-subject analysis in the framework of the GLM were applied. The pre-processed functional data were entered in to a single-subject analysis (K. J. Friston et al., 1995) with regressors of interest convolved with hemodynamic response (K. J. Friston et al., 1995). The following condition were modelled: faces, houses. For each subject con-images of conditions of interest were produced (Pernet, 2014): faces vs. houses.

To extend the results at group level, for each of the four pipelines, a GLM group analysis (random effect analysis) was obtained using the relevant con-images of previous step. Subsequently, for each group analysis, 6 regions of interest (ROIs) were obtained as significant cluster of activation at the same cluster and voxel thresholds (voxel-level threshold family-wise error FWE $p<0.05$; cluster-level threshold false discovery rate FDR $p<0.05$). To obtain a common set of ROIs among the different pipelines, the intersection of each common cluster of activation in the group analyses were used. If a ROI did not present the corresponding region on the other side, a flip L/R or R/L of the identified ROI was performed. The following set of ROIs was considered (Figure 5.2): bilateral parahyppocampal place area (PPA), bilateral fusiform face area (FFA), and bilateral object face area (OFA).

To assess the strength of task contrast and goodness of fit in GLM, mean beta contrast estimates and GLM mean of the squared residuals were extracted in each common ROI by single-subject GLM analyses for each pipeline, as first validation. Descriptive statistics were derived for both measures.

To assess differences in the mean beta contrast estimates and mean GLM squared residuals for each single ROI among pipelines, two one-way ANOVA with repeated measures was performed on these parameters, separately, in each ROI (i.e. within-subjects factor) for each pipeline (i.e. between-subjects factor) (SPSS - Version 17.0. Chicago: SPSS Inc.). Post-hoc tests using the Bonferroni correction were applied to identify significant differences between pipelines.

Moreover, to additional evaluate the strength of the task, clusters of activation were reported for each group analysis to assess if additional regions were identified comparing the different pipelines.

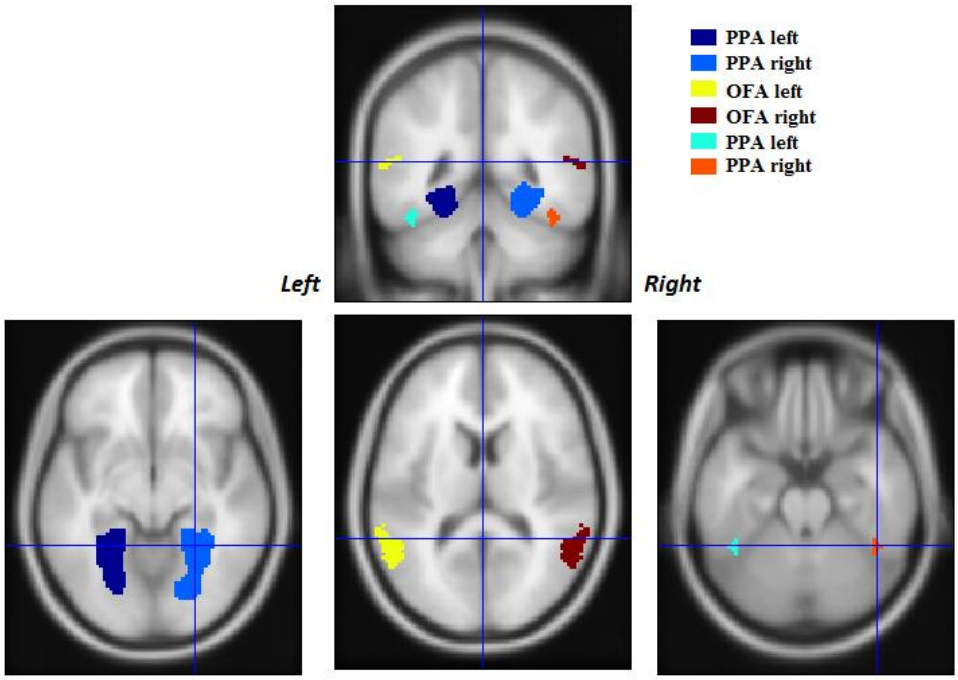


Figure 5.2 Region of interest obtained as intersection of each single common cluster of interest in the map of group analyses for the different pipelines.

We hypothesized that single-subject BOLD signal variability will increase with a more accurate pipeline. To this aim, tSNR was used as second validation. It was obtained for each voxel as the mean of the voxel time series divided by its standard deviation, on fMRI data before and after smoothing step, to be compared with other studies of tSNR (e.g. Kruger and Glover, 2001; Triantafyllou et al., 2005). For each single subjects and common ROI previous defined, mean tSNR value was extracted. In addition, to evaluate the general impact of these pipelines, three additional ROIs were defined for grey matter (GM), WM and CFS. These ROIs were obtained as mask of thresholded segmentations obtained from unified model of normalization step.

To assess differences in the mean tSNR in ROIs among pipelines, a one-way ANOVA with repeated measures was performed on mean tSNR in each ROI (i.e. within-subjects factor) for each pipeline (i.e. between-subjects factor). Post-hoc tests using the Bonferroni correction were applied to identify significant differences between pipelines.

5.4 Results

5.4.1 Evaluation of task strength

In all group GLM analyses, significant cluster of activation were detected in bilateral FFA, with a

higher signal and extension on the right side, and right OFA for faces > houses contrast; in bilateral PPA for houses > faces contrast.

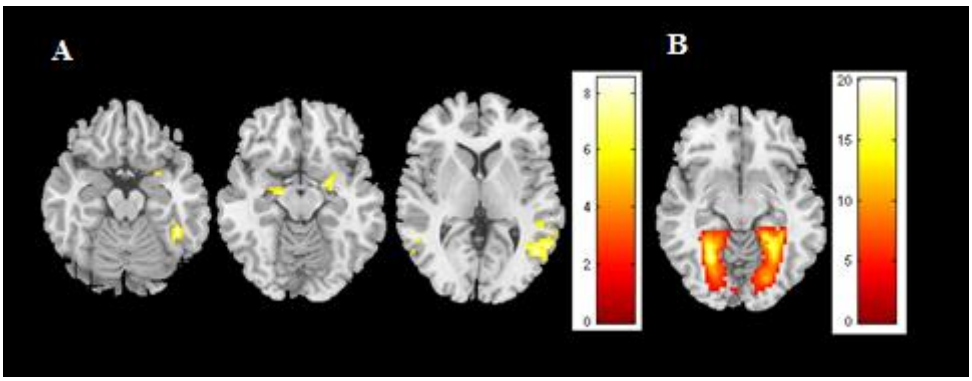


Figure 5.3 Group analysis results for 4) RPPhyP pipeline: A. Faces>Houses and B. Houses>Faces.

Clusters of activation for each group analysis in the different pipelines were reported in Supplementary Data (at end of this Chapter). In both contrast, there was a slight improvement in z-score e number of voxels in task ROIs from 1) none to 4) pipeline and the identification of an additional region in amygdala in the contrast faces > houses with the introduction of nuisance regressors: right amygdala in 2) RP pipeline and bilateral amygdala in 3) PhyP and 4) RPPhyP pipelines.

To evaluate the contrast task strength, boxplot of beta contrast estimates were reported for each common ROI. We can observe a decrease of variability from "none" pipeline to RPPhyP pipeline.

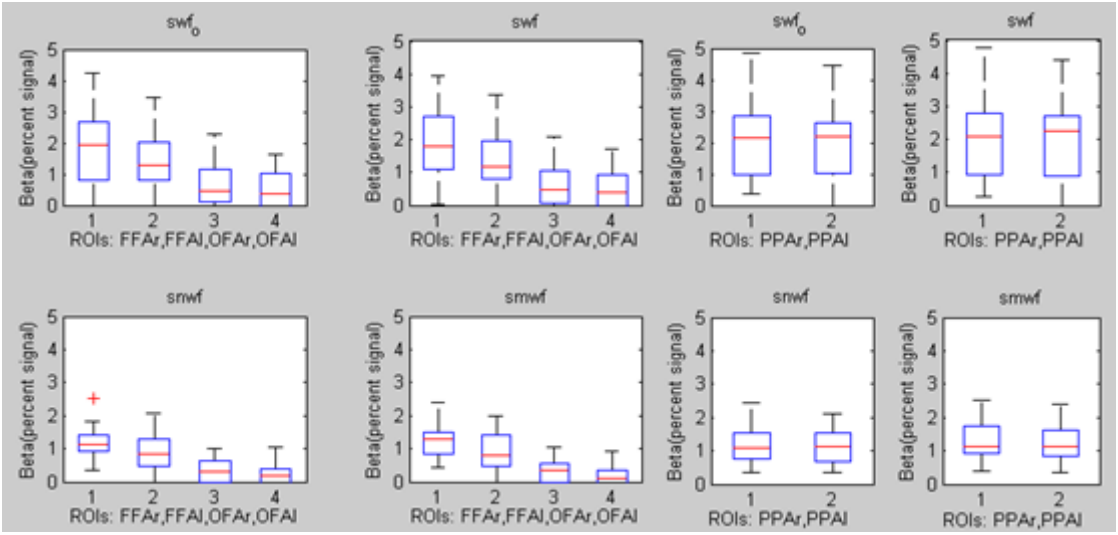


Figure 5.4 Boxplot of beta contrast estimates in ROIs for the four pipelines (swf_o: no signal regression; swf: RP; snwf: PhyP; smwf: RPPhyP): bilateral FFA and FFA for faces > houses contrast and bilateral PPA for the houses > faces contrast.

To evaluate the goodness of fit of GLM, GLM mean of the squared residuals were showed for the four pipelines compared among them. Similar pattern of beta contrast estimates was observed.

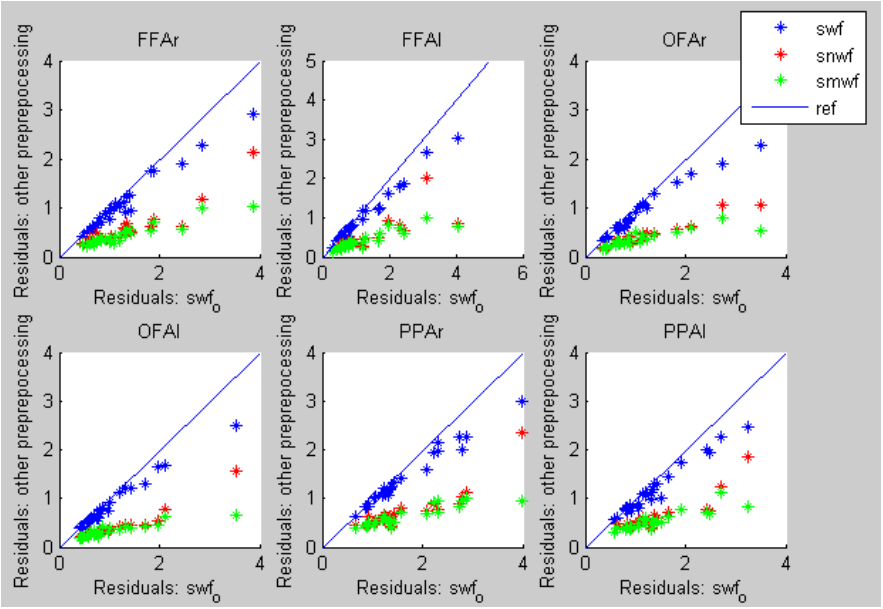


Figure 5.5 The introduction of nuisance regressors reduced the signal assigned to error terms in GLMs. Residuals almost always decrease upon application of additional signal regression, indicating that the GLM was better able to model the variance in the data when nuisance signal were modelled. 1) (none: swf_o), 2) (RP: swf), 3) (PhyP: snwf), and 4) (RPPhyP: smwf) pipelines.

From post-hoc tests on repeated measures one-way ANOVA on mean of beta contrast estimates and squared residuals in ROIs, significant differences were highlighted between the following pipelines: none-PhyP; none-RPPhyP; RP-PhyP; RP-RPPhyP. Moreover, while significant differences were detected between none-RP pipelines on mean squared residuals for each ROIs, no significant differences were identified on beta contrast estimates.

	Faces> Houses: FFA right				**Faces> Houses: FFA left**			
	none	**RP**	**PhyP**	**RPPhyP**	**none**	**RP**	**PhyP**	**RPPhyP**
none	-	ns	0.014	0.022	-	ns	0.028	0.035
RP	-	-	0.013	0.019	-	-	0.05	0.05
PhyP	-	-	-	ns	-	-	-	ns
ANOVA	F(1.15, 27.595) = 10.779; p = 0.002				F(1.122, 26.935) = 8.487; p = 0.006			
	Faces> Houses: OFA right				**Faces> Houses: OFA left**			
	none	**RP**	**PhyP**	**RPPhyP**	**none**	**RP**	**PhyP**	**RPPhyP**
none	-	ns	ns	ns	-	-	-	-
RP	-	-	ns	ns	-	-	-	-
PhyP	-	-	-	ns	-	-	-	-
ANOVA	F(1.096, 26.301) = 4.704; p = 0.036				ns			

	Houses> Faces: PPA right				Houses> Faces: PPA left			
	none	**RP**	**PhyP**	**RPPhyP**	**none**	**RP**	**PhyP**	**RPPhyP**
none	-	ns	0.000	0.001	-	ns	0.001	0.002
RP	-	-	0.001	0.001	-		0.001	0.002
PhyP	-	-	-	0.05	-			ns
ANOVA	F(1.099, 26.375) = 20.501; p = 0.000				F(1.089, 26.127) = 18.586 p = 0.000			
	Squared residuals: FFA right				**Squared residuals: FFA left**			
	none	**RP**	**PhyP**	**RPPhyP**	**none**	**RP**	**PhyP**	**RPPhyP**
none	-	0.002	0.000	0.000	-	0.001	0.000	0.000
RP	-	-	0.000	0.000	-	-	0.000	0.000
PhyP	-	-	-	ns	-	-	-	ns
ANOVA	F(1.230, 20.524) = 57.794; p = 0.000				F(1.124, 26.976) = 33.497; p = 0.000			
	Squared residuals: OFA right				**Squared residuals: OFA left**			
	none	**RP**	**PhyP**	**RPPhyP**	**none**	**RP**	**PhyP**	**RPPhyP**
none	-	0.019	0.000	0.000	-	0.007	0.000	0.000
RP	-	-	0.000	0.000	-	-	0.000	0.000
PhyP	-	-	-	ns	-	-	-	ns
ANOVA	F(1.064, 25.529) = 37.257; p = 0.000				F(1.103, 26.482) = 42.220; p = 0.000			
	Squared residuals: PPA right				**Squared residuals: PPA left**			
	none	**RP**	**PhyP**	**RPPhyP**	**none**	**RP**	**PhyP**	**RPPhyP**
none	-	0.000	0.000	0.000	-	0.000	0.000	0.000
RP	-	-	0.000	0.000	-	-	0.000	0.000
PhyP	-	-	-	ns	-	-	-	ns
ANOVA	F(1.291, 30.991) = 68.539; p = 0.000				F(1.257, 30.126) = 73.059; p = 0.000			

Table 5.1 Results of repeated measures one-way ANOVA on mean beta contrast estimates and squared residuals with a Greenhouse-Geisser correction and post-hoc tests using the Bonferroni correction. Significant differences were highlighted (grey). ns = not significant.

5.4.2 Evaluation of BOLD signal variability: tSNR

Descriptive statistics for tSNR in ROIs were reported in Figure 5.6 and Figure 5.7.

We can identify an increase of tSNR between the 1) pipeline (swf_o) and the others in grey matter, white matter and CFS before smoothing step. After smoothing, there was an increase in tSNR compared with data without smoothing for each pipeline and a gradual positive trend was highlighted in all regions from the 1) pipeline (swf_o) to the 4) pipeline (smwf).

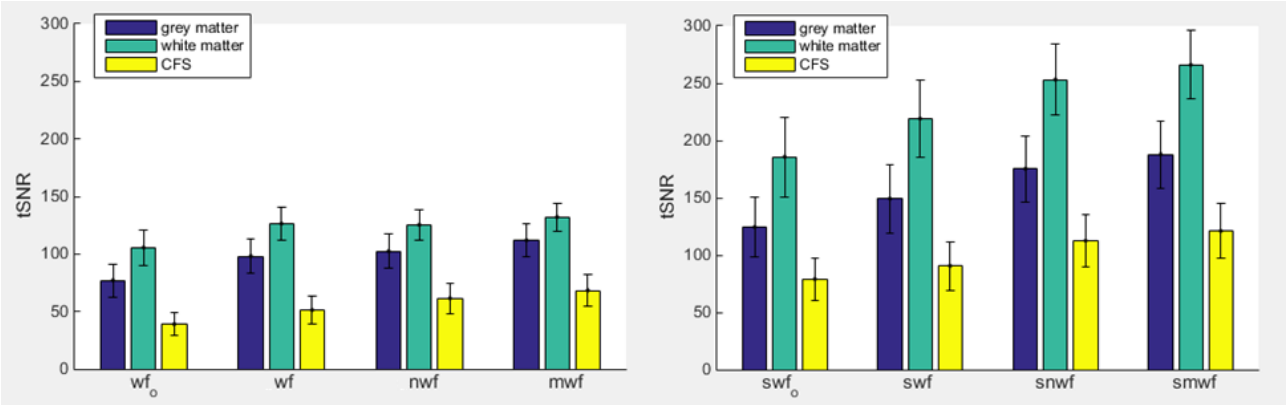


Figure 5.6 Mean and standard deviation of tSNR in mask of GM, WM, and CFS for the 4 pipelines before and after smoothing step.

As for the previous general ROI, also for ROIs derived from task we can identify the same pattern. In particular, the differences in tSNR among the different pipeline was more evident in OFA regions that was a low contrast region.

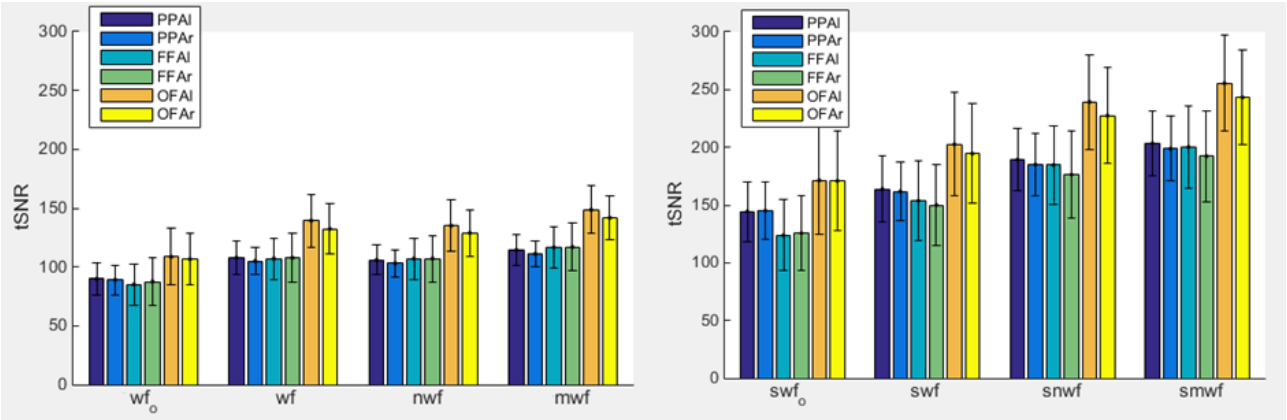


Figure 5.7 Mean and standard deviation of mean tSNR in ROIs derived from task (bilateral PPA, FFA, OFA) for the 4 pipelines before and after smoothing step.

Repeated measures ANOVA with a Greenhouse-Geisser correction determined that mean tSNR for each ROI differed statistically significantly among pipelines:

- grey matter: $F(1.493, 35.844) = 616.978$, $p < 0.000$; white matter: $F(1.493, 35.839) = 455.774$, $p < 0.000$; CSF: $F(1.412, 33.896) = 484.051$, $p < 0.000$;
- left PPA: $F(1.856, 44.537) = 325.240$, $p < 0.000$; right PPA: $F(1.384, 33.213) = 193.067$, $p < 0.000$; left FFA: $F(2.084, 50.024) = 186.838$, $p < 0.000$, right FFA: $F(1.726, 41.431) = 219.261$, $p < 0.000$; left OFA: $F(1.354, 32.491) = 143.965$, $p < 0.000$, right OFA: $F(1.294, 31.065) = 153.293$, $p < 0.000$).

Post hoc tests using the Bonferroni correction revealed that there was a significant different effect ($p<.000$) for each couples of pipelines.

5.5 Discussion

In task ROIs, 3) snwf (PhyP) and 4) smwf (RPPhyP) pipelines obtained a significant increase of task contrast estimate and a reduction of GLM residual term in comparison to 1) swf_o (none), 2) swf (RP). These results underlined an increase in task strength and GLM modelling for these pre-processing approaches. On the other site, the introduction of additional nuisance regressors increase the tSNR, and consecutive the BOLD signal variability, gradually from 1) swf_o (none) to 4) smwf (RPPhyP) pipelines. The increase was more evident after the smoothing step.

The introduction of realignment parameters as nuisance regression, 2) swf (RP) pipeline, permitted to obtain a more accurate model (Lund et al., 2005) (i.e. significant reduction of mean squared residuals) and increase in tSNR, but does not highlighted a significant task strength (beta contrast estimates) in all the task ROIs. This aspects was in line with previous studies, where the impact of motion regression was not significant in low-motion subjects (Evans et al., 2010) and in some cases it can reduce reduces the strength of task activation (Johnstone et al., 2006).

On the other side, introduction of physiological parameters as signal of no-interest in regression, 3) snwf (PhyP) pipeline, in comparison to the 2) swf (RP) pipeline achieved better results not only on model fitting (decrease of residual term), but also in beta contrast estimates and tSNR in all task ROIs identified in group analyses. Moreover, this pipeline permitted to identify significant regions of interest in bilateral amygdala that were under-threshold in previous pipeline. The amygdala is an important region previous identified in other studies using this task (Kleinhans et al., 2007; Vuilleumier et al, 2001). The results were similar to those reported by Behzadi and collegues (2007), where they showed a decrease of temporal standard deviation (tSTD) and an increase of active voxels comparing regression of physiological signal derived from images to no nuisance regressors. In our case, we additional underlined these results in comparison to other pipelines and the suggested order for temporal filtering and smoothing steps.

Finally, the introduction of motion parameters and physiological derived signals in regression, 4) smwf (RPPhyP) pipeline, in comparison to the 2) swf (RP) pipeline increased task strength and GLM residuals. Despite a slight increase in beta contrast parameters and mean squared residuals in ROIs compared to 3) snwf (PhyP) pipeline, there were no significant improvements on these parameters for this pipeline, excepted in beta contrast estimate in right PPA (house> faces, Table

5.1). These results were similar to previous 3) snwf (PhyP) pipeline. They supported the hypothesis that some structured noise may be improve the detectability of signal of interests, reaching a threshold beyond which there were only minimal improvements in quality of results (Yan et al. 2013, Varikuti et al., 2016). Moreover, the effects on tSNR in the 3) snwf (PhyP) and 4) smwf (RPPhyP) pipelines using derived physiological measure were in line to those reported by Hutton and others (2011) using physiological noise models based on cardio-respiratory information recorded during fMRI acquisition.

These results were important in a dataset with low movements, because stress the importance of considering the physiological noise influence on BOLD signal in grey matter and not only focalize attention on realignment parameters regression alone.

We can conclude that:

- the increase of the strength of task contrast (i.e. increase in mean and decrease of standard deviation in beta contrast estimated) and variability of BOLD signal in 3) snwf (PhyP) and 4) smwf (RPPhyP) pipelines indicate that the GLM was able to better model the variance in the data when signals of no-interest were regressed out, also if these signal were derived from images (i.e. WM and CFS signals);
- the quality of brain activation maps may be significantly limited by sub-optimal choices of data pre-processing steps (or "pipeline") in a clinical task-design;
- pre-processing choices have significant effects and optimized pipelines on the basis of state of art in resting state field may significantly improve the reproducibility of task-based fMRI results.

References

Behzadi Y, Restom K, Liau J, Liu TT. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage;37(1):90-101.

Evans JW1, Todd RM, Taylor MJ, Strother SC. (2010). Group specific optimisation of fMRI processing steps for child and adult data. Neuroimage;50(2):479-90.

Freire L, Mangin JF. (2001). Motion correction algorithms may create spurious brain activations in the absence of subject motion. Neuroimage;14(3):709-22.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI timeseries revisited. NeuroImage, 2(1), 45–53.

Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. (1996). Movement-related effects in fMRI time-series. Magn Reson Med;35(3):346-55.

Glover GH, Li TQ, Ress D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med;44(1):162-7.

Hallquist MN, Hwang K, Luna B. (2013). The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fMRI pre-processing reintroduces noise and obscures functional connectivity. Neuroimage;82:208-25.

Hutton C, Josephs O, Stadler J, Featherstone E, Reid A, Speck O, Bernarding J, Weiskopf N. (2011). The impact of physiological noise correction on fMRI at 7 T. Neuroimage;57(1):101-12.

Jo HJ, Saad ZS, Simmons WK, Milbury LA, Cox RW. (2010). Mapping sources of correlation in resting state FMRI, with artifact detection and removal. Neuroimage;52(2):571-82.

Johnstone T, Ores Walsh KS, Greischar LL, Alexander AL, Fox AS, et al. (2006) Motion correction and the use of motion covariates in multiple-subject fMRI analysis. Hum Brain Mapp 27: 779–788.

Jones TB, Bandettini PA, Birn RM. (2008). Integration of motion correction and physiological noise regression in fMRI. Neuroimage;42(2):582-90.

Kleinhans NM, Johnson LC, Mahurin R, Richards T, Stegbauer KC, Greenson J, Dawson G, Aylward E. (2007). Increased amygdala activation to neutral faces is associated with better face memory performance. Neuroreport;18(10):987-91.

Krüger G, Glover GH. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. Magn Reson Med;46(4):631-7.

Lund TE, Madsen KH, Sidaros K, Luo WL, Nichols TE. (2006). Non-white noise in fMRI: does modelling have an impact? Neuroimage;29(1):54-66.

Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? Neuroimage;44(3):893-905.

Pernet CR. (2014). Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers. Front Neurosci;8:1. Review.

Shaw ME, Strother SC, Gavrilescu M, Podzebenko K, Waites A, Watson J, Anderson J, Jackson G, Egan G. (2003). Evaluating subject specific preprocessing choices in multisubject fMRI data sets using data-driven performance metrics. Neuroimage;19(3):988-1001.

Strother S, La Conte S, Kai Hansen L, Anderson J, Zhang J, Pulapura S, Rottenberg D. (2004). Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. Neuroimage;23 Suppl 1:S196-207.

Triantafyllou C, Hoge RD, Krueger G, Wiggins CJ, Potthast A, Wiggins GC, Wald LL. (2005). Comparison of

physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. Neuroimage;26(1):243-50.

Varikuti DP, Hoffstaedter F, Genon S, Schwender H, Reid AT, Eickhoff SB. (2016). Resting-state test-retest reliability of a priori defined canonical networks over different preprocessing steps. Brain Struct Funct

Vergara VM, Mayer AR, Damaraju E, Hutchison K, Calhoun VD. (2016). The effect of preprocessing pipelines in subject classification and detection of abnormal resting state functional network connectivity using group ICA. Neuroimage. pii: S1053-8119(16)00243-3.

Vuilleumier P, Armony JL, Driver J, Dolan RJ. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron;30(3):829-41.

Weissenbacher A, Kasess C, Gerstl F, Lanzenberger R, Moser E, Windischberger C. (2009). Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. Neuroimage;47(4):1408-16.

Yan CG, Craddock RC, He Y, Milham MP. (2013). Addressing head motion dependencies for small-world topologies in functional connectomics. Front Hum Neurosci;7:910.

Zhang J, Anderson JR, Liang L, Pulapura SK, Gatewood L, Rottenberg DA, Strother SC. (2009). Evaluation and optimization of fMRI single-subject processing pipelines with NPAIRS and second-level CVA. Magn Reson Imaging;27(2):264-78.

Supplementary data

swf_o (none)							
Contrast	**Brain region**	**Side**	**x**	**y**	**z**	**z-score**	**n.voxels**
	FFA	Right	42	-46	-22	5.96	20
	OFA	Right	45	-61	5	5.55	47
	OFA	Left	-57	-67	5	4.91	9
swf (RP)							
Contrast	**Brain region**	**Side**	**x**	**y**	**z**	**z-score**	**n.voxels**
	FFA	Right	42	-46	-19	5.78	18
	OFA	Right	48	-58	8	5.75	35
	Amygdala	Right	27	-10	-10	4.85	2
snwf (PhyP)							
Contrast	**Brain region**	**Side**	**x**	**y**	**z**	**z-score**	**n.voxels**
	FFA	Right	42	-46	-22	5.73	13
	OFA	Left	-51	-70	8	5.7	6
	Amygdala	Right	27	-10	-10	5.7	9
	OFA	Right	48	-58	8	5.63	114
	Amygdala	Left	-30	-7	-10	5.53	9

smwf (RPPhyP)							
Contrast	**Brain region**	**Side**	**x**	**y**	**z**	**z-score**	**n.voxels**
	OFA	Right	51	-55	8	6.15	104
	FFA	Right	42	-43	-19	5.62	14
	Amygdala	Right	21	-1	-10	5.18	15
	Amygdala	Left	-24	-10	-10	5.11	5
	OFA	Left	-51	-70	8	5.04	2

Supplementary Table 5.1: Coordinates, z-score and number of voxel for clusters of activation in contrast Faces > Houses (voxel-level threshold FWE p<0.05).

swf_o (none)							
Contrast	**Brain region**	**Side**	**x**	**y**	**z**	**z-score**	**n.voxels**
	PPA	Right	24	-46	-13	Inf	332
	PPA	Left	-27	-61	-13	7.84	272
	Occipital cortex	Right	36	-76	20	6.44	145
	Occipital cortex	Left	-39	-82	-17	6.31	208
	Parietal Superior	Right	21	-76	47	5.31	5
swf (RP)							
Contrast	**Brain region**	**Side**	**x**	**y**	**z**	**z-score**	**n.voxels**
	PPA	Right	24	-46	-13	Inf	361
	PPA	Left	-27	-61	-13	7.79	289
	Occipital cortex	Right	36	-76	20	6.54	178
	Occipital cortex	Left	-33	-85	23	6.38	221
snwf (PhyP)							
Contrast	**Brain region**	**Side**	**x**	**y**	**z**	**z-score**	**n.voxels**
	PPA	Left	-27	-61	-13	Inf	284
	PPA	Right	27	-43	-13	Inf	384
	Occipital cortex	Right	36	-76	20	6.53	175
	Occipital cortex	Left	-27	-79	14	6.46	176
	Parietal Superior	Right	21	-76	47	5.04	2
smwf (RPPhyP)							
Contrast	**Brain region**	**Side**	**x**	**y**	**z**	**z-score**	**n.voxels**
	PPA	Left	-27	-61	-13	Inf	330
	PPA	Right	24	-46	-10	Inf	408
	Occipital cortex	Right	35	-76	20	6.75	215
	Occipital cortex	Left	-27	-82	23	6.5	195

Supplementary Table 5.2: Coordinates, z-score and number of voxel for clusters of activation in contrast Houses > Faces (voxel-level threshold FWE p<0.01).

Chapter 6

fMRI studies: applications

6.1 Introduction

Previous studies have demonstrated that patients with disorder of consciousness (DOC) maintain some minor neural processing. Here, we presented some works performed with functional magnetic resonance on patients with DOC applying improvements in pre-processing analysis on the basis of literature.

A sample of 168 patients and healthy participants were recruited at Neurological Institute Carlo Besta during CRC-Start Up Coma Research Centre project. All patients underwent a 1-week program of clinical multimodal assessment, enclosing clinical, neurophysiological, and neuroimaging evaluations. All MRI imaging data were acquired with a 32-channel head coil on an Achieva 3-T MR scanner (Philips Healthcare, NL, USA).

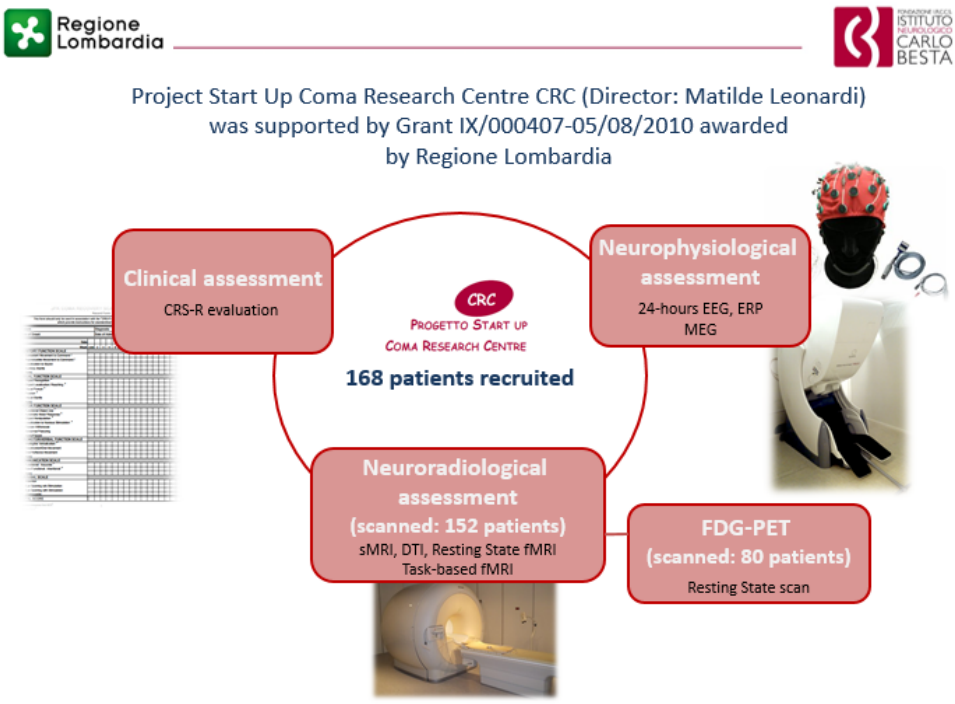


Figure 6.1 Organization of Start-Up Coma Research Centre CRC

6.2 Patients with disorder of consciousness

Consciousness is a complex concept that has two major components: awareness of environment and of self (i.e. the content of consciousness) and wakefulness (i.e. the level of consciousness).

Disorder of consciousness (DOC) is a heterogeneous condition. This disorder includes different aetiology, traumatic and non-traumatic brain injuries, ranging from vegetative state/unresponsive wakefulness syndrome (VS/ UWS) (Laureys et al., 2010) to minimally conscious state (MCS) (Giacino et al., 2002).

A clinical and behavioural diagnosis of residual cognitive function in patients with DOC, can be very difficult, due to the complexity of detecting a clear response. Indeed, it is possible that a subgroup of these patients may retain a level of awareness and sensory functions, but does not have the ability to communicate through an explicit output response assessed with behavioural scales, with a high rate of misdiagnosis that can reach 43% (Andrews et al., 1996). Actually, a diagnosis of patient as VS/UWS, lowest level in diagnosis, is supported by evidence of ***arousal*** (such as eye opening) but no signs of ***awareness*** of self or environment. When a VS/UWS patient with a non-traumatic etiology remains in this state for more than three months after brain injury, he/she is defined to be in a persistent vegetative state. In case of traumatic etiology the cut-off is extended to one year after brain injury. The term "persistent" refers to a chronic phase and implies an unfavourable prognosis about the possibility of improvement. A middle level of diagnosis is MCS (Giacino et al., 2002): patients who shows a preservation of both dimension of consciousness, arousal and awareness. The highest level of diagnosis is a patient classified as ***emerged from the minimally conscious state (EMCS)***, who is able to communicate with consistent and appropriate response.

Generally in clinical scale, VS/UWS patients present a preserved hypothalamic and brainstem autonomic functions for survival condition, but present no evidence of sustained, reproducible or voluntary behavioural response to sensory (visual, auditory, tactile, or noxious) stimuli, language processing and motor response. In contrast, MCS patients show inconsistent but reproducible evidence of minimal response. However, in some cases is very difficult to distinguish reflexive movements unrelated to the context from consistent responses. Finally, EMCS patients can therefore use multiple objects in an appropriate manner and their communication systems are adequate and consistent (Giacino et al. 2002).

Another particular class of DOC patients is composed by locked-in syndrome (LIS). It is a complete paralysis of the body resulting from a lesion localized in the brainstem (American Congress of

Rehabilitation Medicine, 1995). Cognitive functions are fully preserved if the lesion is only restricted to the brainstem. The only way that these patients have to communicate with their environment is through eye movement (blink). Despite the fact that the patients cannot move, their sensations are still intact and they are fully aware of their environment and themselves (Laureys et al., 2005a). Nearly 90% of LIS cases has vascular etiology, but there are also few cases with traumatic brain injury.

The application neuroimaging and neurophysiology evaluation to assess residual cognitive abilities in patients lacking behavioural responses, such as VS/UWS and MCS, has been one of the key point in the last few years (see Gosseries et al., 2014; Giacino et al., 2014 for recent reviews). In particular, recent studies have demonstrated that DOC patients show primary sensory areas and, in some cases, also higher-order sensory areas preservation with different extension in response to sensory stimulation (Coleman et al., 2009; Di et al., 2007; Monti et al., 2010). Preserved sensory processing has been demonstrated for visual (Giacino et al., 2006; Zhu et al., 2009), auditory (Coleman et al., 2009; Di et al., 2007) and tactile (Schiff et al., 2005; Eickhoff et al., 2008) modality.

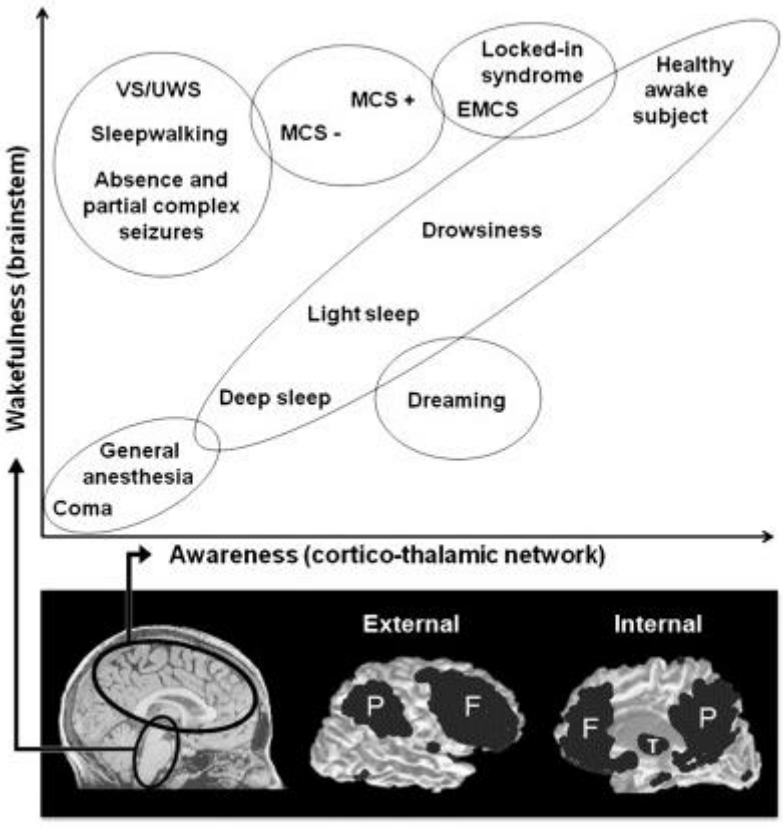


Figure 6.2 The two main components of consciousness: wakefulness and awareness (adapted from Laureys, 2005b)

Characteristic clinical features of disorders of consciousness						
Disorder	**Arousal and attention**	**Cognition**	**Receptive language**	**Expressive language**	**Visuoperception**	**Motor function**
Coma	No sleep–wake cycles*	None	None	None	None	Primitive reflexes only
Vegetative state	Intermittent periods of wakefulness*	None	None	None	Inconsistent visual startle	Involuntary movement only
Minimally conscious state	Intermittent periods of wakefulness	Inconsistent but clear-cut behavioural signs of self-awareness or environmental awareness	Inconsistent one-step command-following*	Aspontaneous and limited to single words or short phrases*	Visual pursuit* Object recognition*	Localization to noxious stimuli* Object manipulation* Automatic movement sequences*
Post-traumatic confusional state	Extended periods of wakefulness	Confused and disoriented*	Consistent one-step command-following	Sentence-level speech, often confused, perseverative Reliable yes–no responses*	Object recognition	Functional use of common objects*
Locked-in syndrome‡	Normal sleep–wake cycles	Normal to near-normal	Normal	Aphonic	Normal	Tetraplegia

*Key distinguishing features. ‡Locked-in syndrome is not a disorder of consciousness, but is included here for purposes of comparison with syndromes associated with significant disturbance in consciousness.

Table 6.1 Clinical characteristics of DOC patients (Source: Giacino et al., 2014)

6.2.1 Clinical scale: CRS-R

The clinical evaluation was performed with the coma recovery scale-revised (CRS-R) (Giacino et al., 2004), 23- items including six subscales addressing auditory, visual, motor, oromotor/verbal, communication and arousal functions. Hierarchically items associated with brainstem, subcortical and cortical processes, constitute the CRS-R subscales: the lowest item on each subscale represents reflexive activity, while the highest items represent cognitively-mediated behaviours. Scoring is standardized and is based on the presence or absence of behavioural responses to specific stimuli administered. For each subscale, there are threshold scores for which the patient can be classified from VS/UWS to MCS to emersion from MCS.

Scala per la funzione uditiva	
4	Movimenti consistenti su ordine*
3	Movimenti riproducibili su ordine*
2	Localizzazione del suono
1	Reazione di sussulto uditivo
0	Nessuna risposta
Scala per la funzione visiva	
5	Riconoscimento dell'oggetto*
4	Localizzazione dell'oggetto: raggiungimento*
3	Inseguimento visivo*
2	Fissazione*
1	Reazione di sussulto visivo
0	Nessuna risposta

Scala per la funzione motoria	
6	Uso funzionale dell'oggetto†
5	Risposte motorie automatiche*
4	Manipolazione degli oggetti*
3	Localizzazione dello stimolo nocicettivo*
2	Allontanamento in flessione
1	Postura anomala
0	Nessuna risposta/flaccidità
Scala per la funzione motoria orale/verbale	
3	Verbalizzazione comprensibile*
2	Vocalizzazione/movimenti orali
1	Movimenti orali riflessi
0	Nessuna risposta

Scala per la comunicazione	
2	Funzionale: Appropriata†
1	Non funzionale: intenzionale*
0	Nessuna risposta
Scala per la vigilanza	
3	Attenzione
2	Apertura degli occhi senza stimolazione
1	Apertura degli occhi con stimolazione
0	Non risvegliabile

* Indica Stato di Minima Coscienza.
† Indica emergenza dallo Stato di Minima Coscienza.

Figure 6.3 Italian CRS-R subscales.

The CRS-R was repeated 3-5 times during the week of the assessment and from different operators to obtain a reliable measure independent from arousal fluctuation of patient.

6.3 Central olfactory processing in patients with disorders of consciousness

in collaboration with Prof. Johan Lundstrom, Dept. of Clinical Neuroscience, Karolinska Institute – Stockholm (Nigri et al. 2015)

6.3.1 Background and aims

The sense of olfaction presents some anatomical and functional organization, characterizing it amongst the other senses (Carmichael et al., 1994). Receptor cells of the nasal mucosa detect the odour stimulus and project to tufed and mitral cells of the olfactory bulb, where happened a first detection and discrimination of odours. The output of the olfactory bulb reaches the piriform cortex and other structures. Form the piriform cortex, there are two direct main streams: one reach the orbitofrontal cortex (OFC) and the other ends in the subcortical limbic regions. The OFC serves as the first neocortical receiving areas for the olfactory pathway, where some study hypothesized that the conscious perception arise (Sherpherd, 2007; Li et al., 2010), and receives indirect projection from other primary cortex through the dorsomedial nucleus of thalamus (Sela et al., 2009). Instead, the subcortical limbic regions are involved in control of learning, memory, motivation, and emotion of odours and presumably mediates unconscious behaviour. Finally, the cerebellum pathway provides a general motor control of sniff generation and a sensory motor control specifically to optimize sensory processing (Mainland et al., 2005).

Considering the anatomical and functional structure of the olfactory system are highlighted four peculiarities that differentiate this sensory system from the others. First, the main feature of the olfactory system is the absence of thalamic relay in the direct pathway between the olfactory receptor layer, peripheral input, and olfactory projection sites in neocortex. Second, afferent projections from the nasal periphery remain ipsilateral all the way to brain hemisphere. Third, the spatial organization of olfactory system is structured in a more dispersed way: the primary cortex includes a large set of structures instead of a single region and secondary cortex does not involve areas adjacent to the primary as usually happens. Fourth, the first cortical processing for olfaction starts in paleocortex (including the olfactory bulb, anterior olfactory nucleus and piriform cortex (Lundstrom et al., 2011). These peculiar features, let us hypothesized that olfactory processing could be preserved in most DOC patients: odors stimuli may be a more direct tool to evoke and assess neural responses in DOC patients. This would open up the possibility of using odors as behavioural neuromodulation stimuli (Giacino et., 2014).

Our aim was to assess neural processing of olfactory stimuli in patients with DOC and in healthy participants with fMRI, to determine:

1. the degree of observed preservation of olfactory processing in the two diagnostic categories (i.e. VS/UWS and MCS patients);
2. the relation between neural activations and etiology or diagnostic category, respectively;
3. the correlation between neural activations in olfactory region of interest and clinical scores (CRS-R);
4. the differences in olfactory neural processing between DOC patients and the healthy participants.

6.3.2 Material and Methods

The recruited participants for the study were described in Table 6.2:

Participants	n. subjects	Age (median; range) years	Disease duration (median; range) months	Sex (M/F)
VS/UWS patients	32	57; 23-77	26; 3-146	20/12
MCS patients	10	44; 20-71	41; 11-170	3/7
Healthy subjects	28	36; 20-64	-	14/14

Table 6.2 Clinical characteristics of DOC patients.

VS/UWS and MCS patients, with different etiologies [hemorrhagic (HBI), traumatic (TBI), and anoxic (ABI) brain injury] were categorized according Aspen criteria (Giacino et al.,1997).

Healthy participants were included only to the neuroimaging part of the experiment.

Due to excessive head movements (see below), 3 healthy participants and 9 patients (6 VS/UWS and 3 MCS) were excluded from the subsequent analyses.

A block-design task with odors and baseline (air) conditions were administered to participants using an olfactometer that is MRI-compatible and controlled by a computer (Lundstrom et al., 2010). The two conditions were proposed in a random and balanced order to minimize the effect of adaptation and habituation (Poellinger et al., 2001). Two unpleasant odors were selected, because elicited more robust activations: 1-octenol-3-ol (odor quality of mushroom) and n-butanol (odor quality of white-board marker). More details on task were reported in Nigri et al., 2015, Nigri et al., 2013.

To reduce susceptibility artefacts and improve signal to noise ratio in orbitofrontal region and piriform cortex/amigdala area (Wilson et al., 2002), we used an EPI sequence with the following characteristics:

- TE = 20 ms: it is reduced in comparison to standard value of 30 ms;

- sagittal slices.

Statistical Parametric Mapping (SPM8) was used for the analysis of fMRI data, while SPSS 17.0 (SPSS Inc., Chicago, IL, USA) for statistical tests on fMRI derived measures in regions of interest (ROIs).

A standard fMRI data pre-processing was applied (i.e. realignment to the mean image, normalization to the MNI EPI template, spatial smoothing 6-mm FWHM, temporal low-pass filtering (hemodynamic response), and temporal high-pass filtering with a cut-off period of 128 s. For quality assessment for the identification of outliers subjects we used the user-defined threshold of 2 voxels on relative displacement of realignment parameters: subjects exceeding this threshold were excluded from further analyses. A scrubbing procedure was used, because it appears to improve task fMRI data results in comparison to other motion regression approaches (Siegel et al., 2014). Framewise displacement was calculated from the 6 head realignment parameters in each subject. A user-dependent threshold was set 1 mm and blocks with FD superior to this threshold in at least 5 consecutive volumes were excluded from the GLM analysis by scrubbing (Power et al., 2012). Scrubbing procedure removed 10 ± 11% (range: 0–39%) of the volumes across DOC patients. They were equally distributed between odor (10.6 ± 11.7%) and baseline (10.0 ± 11.8%) conditions. No volumes were discarded from healthy participants.

Three different analyses were conducted:

- ***single-subject whole-brain analyses***. Two ROIs were selected a priori on the basis of the previous literature: the ***piriform cortex*** associated to amygdala cortex (***PC/AMY***) and the ***lateral-medial orbitofrontal cortex*** (***lOFC***) (Seubert et al., 2012). Odor-based processing at whole-brain level in the defined ROIs was assessed with specific voxel-level (uncorrected $p < 0.005$) and cluster-level thresholds (minimum cluster size of 20 voxels).
- ***diagnostic and etiological group analyses***. To determine significant differences in neural processing of odor stimuli between the diagnostic (VS/UWS; MCS) and etiological (TBI; HBI; ABI) groups, separately, we used the following classification: each patient was identified with one of these categories based on the evaluation of presence/absence of functional activation in previous ROIs, PC/AMY and lOFC, during fMRI task: no activation, activation in PC/AMY, and activation in PC/AMY and lOFC. Moreover, to identify whole-brain significant differences among VS/UWS, MCS, and healthy participants groups on statistical maps, a one-way ANOVA was applied (VS/UWS versus MCS patients; DOC patients versus healthy participants).

- ***correlation with clinical data***. To determine if there was a relation between beta parameter estimates in ROIs are linked to the clinical data, Spearman's correlations were performed between CRS-R scores and beta parameter estimates in literature-defined ROIs. The literature-defined ROIs were centered in the activation peak coordinates of a meta-analysis conducted on olfactory tasks (Seubert et al., 2012) to avoid circularity error in correlation analysis (Vul et al., 2012): bilateral PC/AMY and OFC. Moreover, to evaluate the indirect effects of CRS-R scores on beta parameter estimates in bilateral PC/AMY and OFC via time of scan post-injury a mediation analysis was conducted (Price 1985).

6.3.3 Main results

Olfactory neural processing was at least partially preserved in a large portion of DOC patients.

This study produced three important results:

- First, we demonstrated that odor-related activation within ***primary olfactory regions*** was preserved in a large portion of VS/UWS patients (58%) and all MCS patients (100%), while within ***higher-order olfactory processing area (lOFC)***, an area previously linked to conscious experience of odor stimuli (Li et al., 2010) was significant in the majority of both VS/UWS patients (39%) and MCS patients (71%). Importantly, a difference in the extent of the activated olfactory network was identified among the various etiologies, but not the diagnostic category: the majority of the ABI patients had either no activation or limited activation within PC/AMY, whereas most of the TBI and HBI patients demonstrated significant activation in lOFC. The differences of olfactory preservation among the different etiology can be linked to the different pattern of damage related to the specific brain injury (i.e. more localized in TBI and HBI, while more diffuse in ABI).
- Secondly, a negative correlation was demonstrated between beta parameter estimates for odor stimuli within OFC and CRS-R scores in DOC patients. We hypothesized two different explanations for this behaviour: the first theory is related to a reduction in inhibitory inputs within the primary odor cortex, which serves an important hub for the regulation of the activity in the higher-order areas; the second hypothesis is linked to a continuous direct interaction, mediating conscious odor percept, between lower and higher cortical areas, as suggested by others (Baars 2013; Casarotto et al., 2016). Moreover, a recent study reported that neural processing of odors changed between awake and anesthetized animals: in the latter state the recurrent inhibition is significantly lower compared to former state (Rinberg et al., 2006).

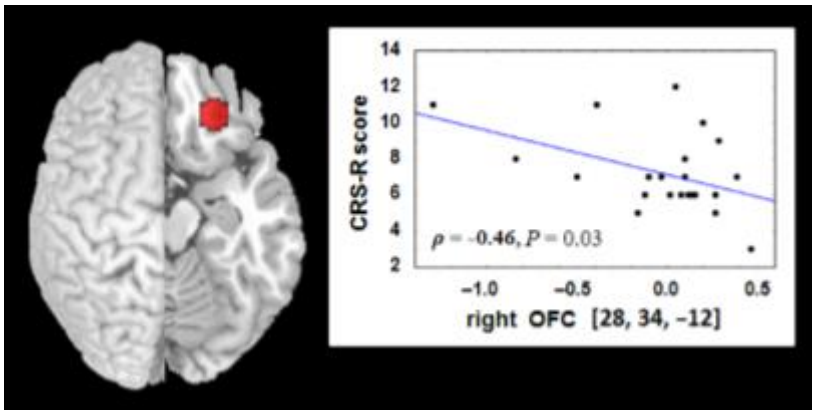


Figure 6.4 Correlation between contrast estimates in right OFC and CRS-R (Source: Nigri et al., 2015)

- Thirdly, we did not identified significant olfactory-related regions differing between healthy participants and DOC patients, while we detected a significant different responce for task between the two groups within the precuneus. This region is rarely associated with odor-related processing (Seubert et al., 2012), but it is one of the important hub of default mode network (DMN). A different activation in this region was reported by Crone and others (Crone et al., 2011), demonstrating that levels of consciousness was linked to a deactivation of the precuneus in DOC patients during the execution of a task. Usually the regions of DMN, related to ongoing mental processes, were deactivated when a subject was involved in a task. So, we theorized that only DOC patients with more preserved cognitive functions can deactivate the DMN regions, because they were able to break internal mental processes to make cognitive reserve available to focus attention on the execution of task.

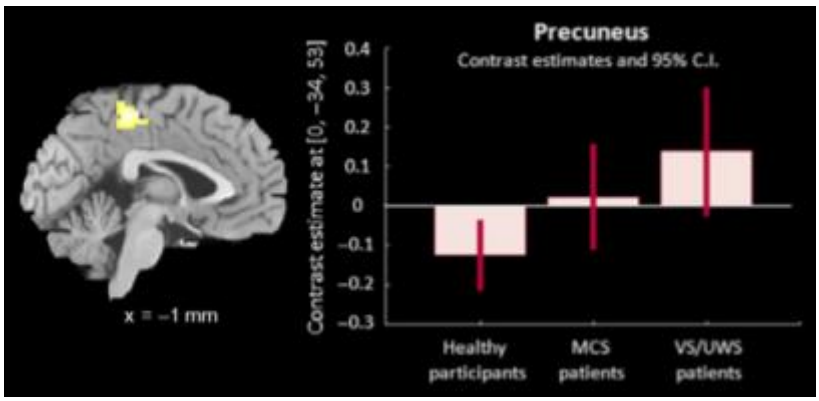


Figure 6.5 Contrast estimates in precuneus for groups of interest (Source: Nigri et al., 2015)

6.4 The neural correlates of lexical processing in disorder of consciousness

in collaboration with Prof. Stefano Cappa, and Eleonora Catricalà, School of Advanced Studies IUSS, Pavia, and, Division of Neuroscience, San Raffaele Scientific Institute, Milan; Giorgio Marotta, Department of Nuclear Medicine, Ca' Granda Ospedale Maggiore Policlinico, IRCCS Foundation, Milan (Nigri et al., 2016).

6.4.1 Background and aims

As reported in the introduction to DOC patients, the assessment of cognitive abilities, in particular of linguistic neural processing, in these patients, lacking behavioural responses, is one of the main goal (Gosseries et al., 2014, Giacino et al., 2014).

Linguistic residual processing was demonstrated in both VS/UWS and MCS patients during fMRI: Schiff et al. (2005), identified an extended neural activation in two MCS patients in the comparison between passive listening of sentences and of reversed speech; Fernandez-Espejo (2008) and Coleman (2009) showed a consistent neural response in subacute VS/UWS patients in response of complex auditory stimuli. A ***hierarchical linguistic task*** was proposed by Davis et al. (2007): they applied this task in healthy subjects awake and during propofol sedation. The results of cortical responses were different among the two states: while the involvement of inferior frontal and posterior temporal regions, observed in conscious subjects in response to ambiguous versus non-ambiguous sentences, disappeared with sedation, the superior temporal areas remained active during sentences versus noise. The application of this paradigm to DOC patients showed that 3 patients diagnosed VS/UWS (out of 7) and 2 patients categorized as MCS (out of 5) presented evidence of preserved speech processing (Coleman et al., 2007; Owen et al., 2005). Subsequently, the same task was applied in a larger cohort, where 19 patients either in VS/UWS or in MCS showed BOLD responses to meaningful speech (Coleman et al., 2009).

A similar exploration of linguistic network in ***semantic processing*** in DOC patients was performed using evoked potentials (reviewed in Rohaut et al., 2015). In particular, a large group of VS/UWS and MCS patients showed N400-like event-related potential (ERP) components, only MCS patients presented a late positive complex. Among the 3 MCS patients, who preserved both N400 and LPC, two of them preserved functional language abilities.

In the present study, during fMRI a hierarchical auditory stimulation paradigm was administered to a group ***of long-term DOC adult patients*** and healthy participants to assessment linguistic cognitive resources. The paradigm included pairs of pseudowords, unrelated words and semantically related

words, i.e. stimuli differing in lexical status (words vs. pseudowords) and semantic relatedness (related vs. unrelated). A first level of task was used to assess automatic neural response to the ***passive presentation*** of meaningful real words, while the second level was applied to evaluate the response to meaning relationships.

6.4.2 Materials and Methods

The recruited participants for the study were reported in Table 6.3.

Fourteen italian DOC patients (4VS/UWS, 10MCS) with different etiologies TBI, HBI, and ABI (CRC-Start Up Coma Research Centre project). To avoid confounding effect due to sedative drugs, they were not administered prior the scanning session.

Three MCS patients were excluded from subsequent analyses, due to excessive head movements during the fMRI scan (Harrison and Connolly 2013).

Participants	n. subjects	Age (median; range) years	Diagnosis	Disease duration (median; range) months	Sex (M/F)
DOC patients	11	57; 19-69	4VS/UWS, 7MCS	27; 5-252	4/7
Healthy subjects	18	46; 27-67	-	-	7/11

Table 6.3 Clinical characteristics of DOC patients.

For the selection of DOC patients the following inclusion criteria were used:

- the presence of brainstem auditory evoked potentials (BAEPs), and
- the absence of anatomical lesions based on T1-weighted scan in auditory regions (i.e. bilateral superior temporal gyrus (STG).

Healthy participants took part only to the neuroimaging portion of the experiment.

An event-related paradigm was used. It was administered during two separate fMRI scans for two reasons: to avoid a twice repetition of a single word in each run and to maintain a stable level of attention during both entire scanning sessions. Each condition was presented in 60 trials, and each trial started with an auditory prime stimulus (mean duration =644 ms; standard deviation (SD) = 114 ms), followed by the presentation of an auditory target stimulus (mean duration =699 ms; SD = 131 ms). Four experimental conditions were identified: pseudowords trials (PP: suto-cosvo), word-

pseudoword trials (WP: cow-cosvo), associatively related trials (WWr: couch–pillow) and unrelated trials (WWur: couch–bridge).

Figure 6.6 Hierarchical task (Source: Nigri et al., 2016)

On fMRI data, realignment, coregistration, normalization, and smoothing (Friston et al., 1995b) were performed. For quality assessment for the identification of outliers subjects we used the user-defined threshold of 2 voxels on relative displacement of realignment parameters: subjects exceeding this threshold were excluded from further analyses. Subsequently, a scrubbing procedure was applied to improve the detection of task-related activations obscured by motion and to reduce the number of false-positives (Power et al., 2012; Siegel et al.,2014). FD was used as quality assessment for head movements in each single subject: volumes exceeding a threshold of > 1 mm in at least 5 consecutive volumes were excluded from the subsequent analyses (Birn et al., 2004; Power et al., 2012). The scrubbing procedure removed 6 ± 4 % (range: 0–12 %) of the volumes in both fMRI sessions only in DOC patients group.

The following analyses were performed:

- ***single-subject whole-brain analysis***. The GLM included 5 regressors of interest, obtained with the convolution of hemodynamic response (Friston et al., 1995a) with the ideal response at target onset for each event-type (WWr, WWur, WP, PP) and baseline, and movement parameters as nuisance regressors. Inferences at single-subject level were drawn from activated clusters obtained with a voxel-level threshold of $p < .001$ uncorrected and a minimum cluster size of 5 voxels.
- ***group analysis***. Con images of patients and healthy participants, obtained from previous step, were entered into two one-way ANOVA, respectively, with group as between factor and the conditions of interest as within-factors.

- ***correlation with clinical data***. Pearson's correlation was applied between CRS-R scores and mean beta parameter estimates in anatomical defined ROIs. To avoid a circularity error in correlation analysis (Vul and Pashler 2012), the a-priori anatomical ROIs were identified based on previous literature (Morosan et al. 2001). They were defined using probabilistic atlas of Anatomy toolbox: auditory cortex (AC) including TE 1.0, TE.1.1, TE 1.2, TE 3.0. The extraction of beta constrast estimates were restricted to grey matter, due to the severe anatomical damage in these patients. In order to create single subject grey matter ROIs, the intersection between the probabilistic ROIs and the segmented grey matter map, obtained from normalization step, was used. This procedure is demonstrated in seed-based correlation improving the reliability of extracted estimates (Varikuti et al., 2016, Chapter 3).

Moreover, two other evaluations were performed:

- auditory function of DOC patients before fMRI was assessed with BAEPs and the acoustic long-latency responses (slow vertex responses, SVR). BAEPs were considered as "normal", "altered", or "absent".
- presence of a possible cortical metabolism in ROIs was evaluated with FDG-PET scan (except one VS/UWS patient not acquired). Standardized uptake value (SUV) map of each subject was thresholded (value=2) with value superior to the highest skull-surface cutaneous and subcutaneous tissues metabolism (Britz-Cunningham et al., 2012).

6.4.3 Main results

Three main results can be obtained with this study.

- First, significant BOLD signal changes were detected in a considerable subgroup of DOC patients (8 out of 11), also in subjects evalauted after a very long period post injury. These activations were observed not only in MCS patients (n=4), but also in VS/UWS patients (n = 4). Considering that a previous longitudinal fMRI study demonstrated decreased of BOLD single changes in longitudinal scanning sessions (Rousseau et al. 2008), this result is particular important in a long-term DOC population.

  Considering the hierarchical levels of the task, in line with our expectations, all the patients, except one, showing any activation in upper levels (7 out of 11), presented significant activations for this low-level contrast (LLA). Interestingly, we observed a pattern among these patients: 3 MCS (MCS2, MCS3, MCS4) patients with no evidence of fMRI activation showed a very low score for the auditory subscale of CRS-R (score 1), while MCS1, who

showed a unilateral cluster of activation in the left STG, presented a higher score in the same subscale (score 2).

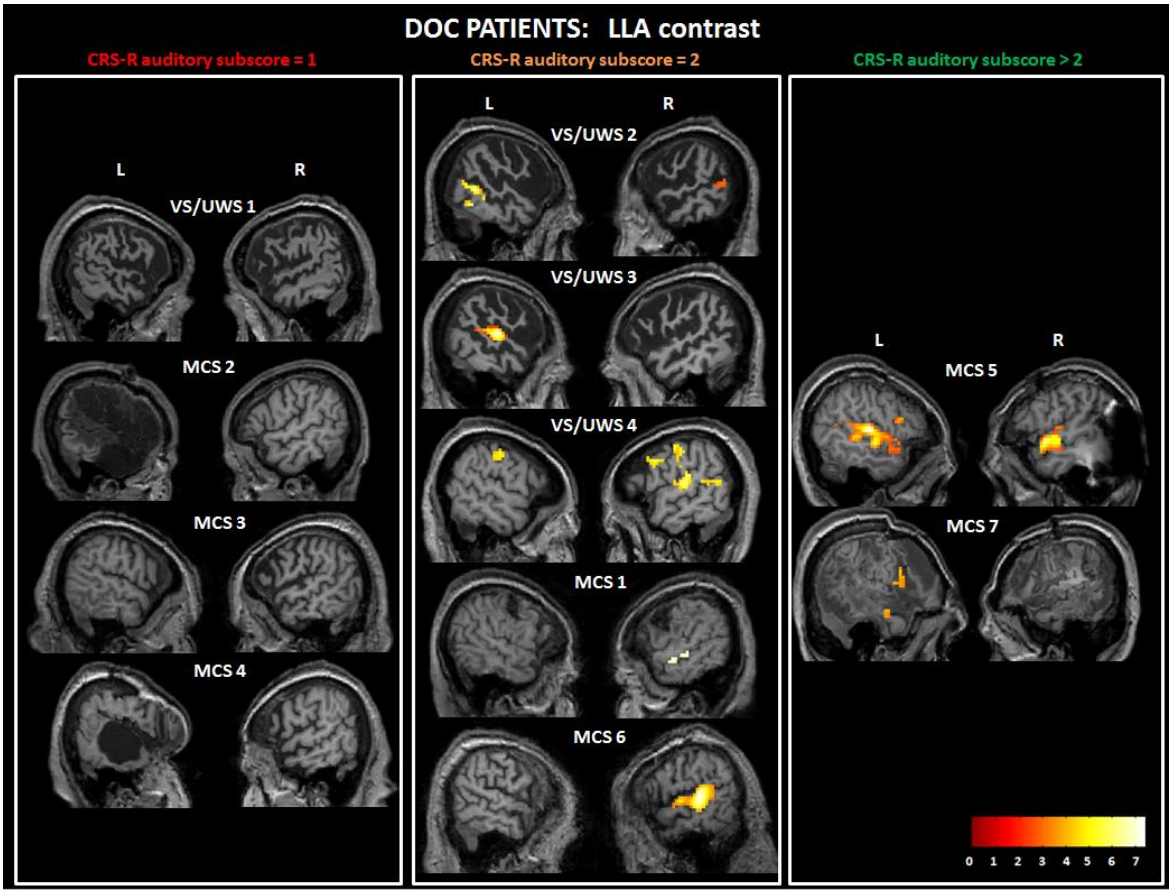


Figure 6.7 For LLA contrast, there is a relation between score of CRS-R auditory subscale and presence of functional activations: score 1 comprise only VS/UWS patients and score >2 only MCS patients. A "grey zone" (score 2), that comprised VS/UWS and MCS patients, was highlighted.

Notably, 3 patients (1VS/UWS, 2 MCS), among the patients showing activity for LLA contrast, presented consistent activation for pseudowordeffect (PWef) contrast (WP+PP>WWr+WWur; upper contrast), 1 MCS for PWef and Lex contrast and, importantly, 1 VS/UWS for lexical (Lex) (WWr + WWur > WP + PP) contrast. Importantly, 2 DOC patients (MCS) presented not only an involvement of the STG but also of inferior frontal gyrus. The only VS/UWS patient showing BOLD signal changes for the PWef contrast presented activations mainly localized in STG and medial temporal gyrus.

- Secondly, ***PET metabolism*** was preserved in regions were significant BOLD signal were identified in response to task. However, 3 MCS patients, although presenting supra-threshold metabolism, did not presented any significant activation for any contrast: we can

observe that a preservation of metabolism does not necessarily determine a functional processing of stimuli.

- Finally, a positive correlation was identified between ***clinical scores*** and beta contrast estimates extracted in bilateral AC during some linguistic conditions (WWur and WP + PP) (Figure 6.8).

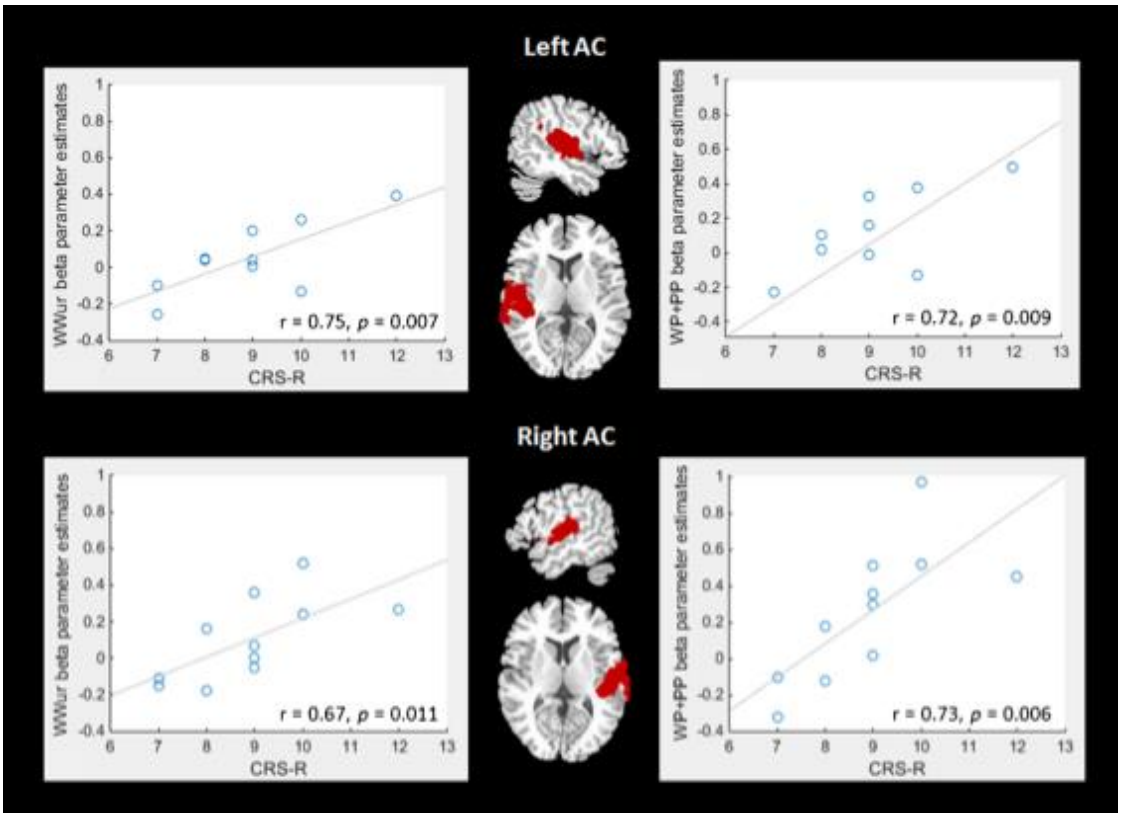


Figure 6.8 Correlation between contrast estimates and CRS-R (Source: Nigri et al., 2016).

Previous studies showed the importance of auditory-language network in DOC patients: language fMRI activity was found to be a reliable marker in predicting the outcome of these patients (Coleman et al., 2009; Wang et al., 2015) and functional connectivity in auditory resting state network, among other resting state networks, was found to achieve the best results in distinguishing between MCS and VS patients (Demertzi et al., 2015).

Moreover, these results highlighted the important role of neuroimaging studies, in particular of fMRI, in the identification of residual processing, "grey zone" (Figure 6.7), not clearly assessed with behavioural scale.

References

Andrews K, Murphy L, Munday R, Littlewood C. (1996). Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. BMJ. Jul 6;313(7048):13-6.

Baars BJ. (2013). Multiple sources of conscious odor integration and propagation in olfactory cortex. Front Psychol; 4: 930.

Birn,R.M.,Cox,R.W.,& Bandettini,P.A.(2004). Experimental designs and processing strategies for fMRI studies involving overt verbal responses. NeuroImage, 23(3), 1046–1058.

Britz-Cunningham, S., & Gerbaud, V. H. (2012). Functional anatomy of the FDG image. In A case-based approach to PET/CTin oncology (pp. 53–74). Cambridge: Cambridge University Press.

Carmichael ST, Clugnet MC, Price JL. (1994). Central olfactory connections in the macaque monkey. J Comp Neurol; 346: 403–434.

Casarotto S, Comanducci A, Rosanova M, et al. Stratification of unresponsive patients by an independently validated index of brain complexity. Ann Neurol. 2016 Sep 22.

Coleman, M. R., Rodd, J. M., Davis, M. H., Johnsrude, I. S., Menon, D. K., Pickard, J. D., & Owen, A. M. (2007). Do vegetative patients retain aspects of language comprehension? evidence from fMRI. Brain: A Journal of Neurology, 130(Pt 10), 2494– 2507.

Coleman MR, Davis MH, Rodd JM, et al. (2009). Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. Brain; 132: 2541–2552.

Davis,M.H., Coleman,M.R., Absalom,A.R., Rodd,J.M., Johnsrude,I. S., Matta, B. F., et al. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. Proceedings of the National Academy of Sciences of the United States of America, 104(41), 16032–16037.

Demertzi A, Antonopoulos G, Heine L, et al. (2015). Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. Brain. Sep;138(Pt 9):2619-31.

Di HB, Yu SM, Weng XC, et al. (2007). Cerebral response to patient's own name in the vegetative and minimally conscious states. Neurology; 68: 895–899.

Eickhoff SB, Dafotakis M, Grefkes C, et al. (2008). fMRI reveals cognitive and emotional processing in a longterm comatose patient. Exp Neurol; 214: 240–246.

Fernandez-Espejo, D., Junque, C., Vendrell, P., Bernabeu, M., Roig, T., Bargallo,N.,& Mercader, J.M.(2008).Cerebral response to speech in vegetative and minimally conscious states after traumatic brain injury. Brain Injury, 22(11), 882–890.

Friston, K. J., Frith, C. D., Turner, R., & Frackowiak, R. S. (1995a). Characterizing evoked hemodynamics with fMRI. NeuroImage, 2(2), 157–165.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995b). Analysis of fMRI timeseries revisited. NeuroImage, 2(1), 45–53.

Giacino JT, Zasler ND, Katz DI, et al. (1997). Development of practice guidelines for assessment and management of the vegetative and minimally conscious states. J Head Trauma Rehabil; 12: 79–89.

Giacino JT, Ashwal S, Childs N, et al. (2002). The minimally conscious state: definition and diagnostic criteria. Neurology; 58: 349–353.

Giacino, J. T., Kalmar, K., & Whyte, J. (2004). The JFK coma recovery scale-revised: Measurement characteristics and diagnostic utility. Archives of Physical Medicine and Rehabilitation, 85(12), 2020– 2029.

Giacino JT, Hirsch J, Schiff N, et al. (2006). Functional neuroimaging applications for assessment and rehabilitation planning in patients with disorders of consciousness. Arch Phys Med Rehabil; 87: S67–S76.

Giacino,J.T.,Fins,J.J.,Laureys,S.,& Schiff,N.D.(2014). Disorders of consciousness after acquired brain injury: the state of the science. Nature Reviews. Neurology, 10(2), 99–114.

Gosseries, O., Di, H., Laureys, S., & Boly, M. (2014). Measuring consciousness in severely damaged brains. Annual Review of Neuroscience, 37, 457 –478.

Harrison AH, Connolly JF. (2013). Finding a way in: a review and practical evaluation of fMRI and EEG for detection and assessment in disorders of consciousness. Neurosci Biobehav Rev; 37: 1403–1419.

Laureys S, Pellas F, Van Eeckhout P, Ghorbel S, Schnakers C, Perrin F, et al. (2005a). The locked-in syndrome : what is it like to be conscious but paralyzed and voiceless? Prog Brain Res.;150:495-511. Review.

Laureys S. (2005b). The neural correlate of (un)awareness: lessons from the vegetative state. Trends Cogn Sci. Dec;9(12):556-9.

Laureys S, Celesia GG, Cohadon F, et al. (2010). Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. BMC Med; 8: 68. 2.

Li W, Lopez L, Osher J, et al. (2010). Right orbitofrontal cortex mediates conscious olfactory perception. Psychol Sci; 21: 1454–1463.

Lundstrom JN, Gordon AR, Alden EC, et al. (2010). Methods for building an inexpensive computer-controlled olfactometer for temporally-precise experiments. Int J Psychophysiol; 78: 179–189.

Lundstrom JN, Boesveldt S, Albrecht J. (2011). Central processing of the chemical senses: an overview. ACS Chem Neurosci; 2: 5–16.

Mainland JD, Johnson BN, Khan R, Ivry RB, Sobel N. (2005). Olfactory impairments in patients with unilateral cerebellar lesions are selective to inputs from the contralesional nostril. J Neurosci. Jul 6;25(27):6362-71.

Monti MM, Vanhaudenhuyse A, Coleman MR, et al. (2010). Willful modulation of brain activity in disorders of consciousness. N Engl J Med; 362: 579–589.

Nigri A, Ferraro S, D'Incerti L, Critchley HD, Bruzzone MG, Minati L. (2013). Connectivity of the amygdala, piriform, and orbitofrontal cortex during olfactory stimulation: a functional MRI study. Neuroreport. Mar 6;24(4):171-5.

Nigri A, Ferraro S, Bruzzone MG, Nava S, D'Incerti L, Bertolino N, Sattin D, Leonardi M, Lundström JN; CRC - Coma Research Centre members. (2016). Central olfactory processing in patients with disorders of consciousness. Eur J Neurol. Mar;23(3):605-12.

Nigri A, Catricalà E, Ferraro S, Bruzzone MG, D'Incerti L, Sattin D, Sebastiano DR, Franceschetti S, Marotta G, Benti R, Leonardi M, Cappa SF; (2016). The neural correlates of lexical processing in disorders of consciousness. Brain Imaging Behav. Oct 13.

Owen, A. M., Coleman, M. R., Menon, D. K., Johnsrude, I. S., Rodd, J. M., Davis, M. H., et al. (2005). Residual auditory function in persistent vegetative state: a combined PET and fMRI study. Neuropsychological Rehabilitation, 15(3–4), 290–306.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functionalconnectivity MRI networks arise from subject motion. NeuroImage, 59(3), 2142–2154.

Price JL. (1985). Beyond the primary olfactory cortex: olfactory-related areas in the neocortex, thalamus and hypothalamus. Chem Senses; 10: 239–258.

Rinberg D, Koulakov A, Gelperin A. (2006). Sparse odor coding in awake behaving mice. J Neurosci; 26: 8857– 8865.

Rohaut, B., Faugeras, F., Chausson, N., King,J.R., Karoui,I.E., Cohen, L.,& Naccache,L.(2015).Probing ERPcorrelates of verbal semantic processing in patients with impaired consciousness. Neuropsychologia, 66, 279

Shepherd GM. (2007). Perspectives on olfactory processing, conscious perception, and orbitofrontal cortex. Ann N Y Acad Sci. Dec;1121:87-101.

Schiff ND, Rodriguez-Moreno D, Kamal A, et al. (2005). fMRI reveals large-scale network activation in minimally conscious patients. Neurology; 64: 514–523.

Sela L, Sacher Y, Serfaty C, Yeshurun Y, Soroker N, Sobel N. (2009). Spared and impaired olfactory abilities after thalamic lesions. J Neurosci. Sep 30;29(39):12059-69.

Seubert J, Freiherr J, Djordjevic J, et al. (2012). Statistical localization of human olfactory cortex. NeuroImage; 66C: 333–342.

Siegel, J. S., Power, J. D., Dubis, J. W., et al. (2014).Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. Human Brain Mapping, 35(5), 1981–1996.

Vul E, Pashler H. (2012). Voodoo and circularity errors. NeuroImage; 62: 945–948.

Wang F, Di H, Hu X, et al. (2015). Cerebral response to subject's own name showed high prognostic value in traumatic vegetative state. BMC Med. Apr 15;13:83.

Wilson JL, Jenkinson M, de Araujo I, et al. (2002). Fast, fully automated global and local magnetic field optimization for fMRI of the human brain. NeuroImage; 17: 967–976.

Zhu J, Wu X, Gao L, et al. (2009). Cortical activity after emotional visual stimulation in minimally conscious state patients. J Neurotrauma; 26: 677–688.

Conclusion and Future Work

In this thesis, we proposed auto-calibrated methods, clustering and MD, for the detection of outlier subjects/volumes and we compared the results with previous described indices based on user-dependent thresholds. These methods increased statistical power of GLM computation at single-subject and group level for the following reasons. First, the methods permitted to detect selectively subject and volumes with high spikes movements. Second, the auto-calibration of the two methods allows to homogenise the variability among subjects and volumes for movements at level of group and single-subject. This ensures a more robust GLM estimation in both low and high movement populations as we have observed in the reduction of GLM whole-brain mean squared residuals. These methods are particularly suitable to investigate populations with high movements (e.g. children, disorder of consciousness, movement disorders) usually not prone to be evaluated using the common procedure based on user-dependent thresholds. However, as well known, high movements synchronized with task are difficult to discard without a reduction of task-related signal, as previously described (Johnstone et al., 2006). Moreover, the proposed auto-calibrated method, as the other metrics, based on realignment parameters. This estimation could not be entirely accurate because sampling depended on TR, might not reflect properties relevant to all slices, and could be influenced by how data was processed prior/during to realignment (Andersson et al., 2001).

Clustering and MD methods based on realignment parameters provided an alternative data-driven approach in contrast to other methods based on user-dependent thresholds implemented in other fMRI analysis softwares, such as AFNI (3dToutcount; https://afni.nimh.nih.gov/afni), FSL (fsl_motion_outliers; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) and SPM (ArtRepair toolbox; (Mazaika et al., 2007). Through this approach, we obtained a quantification of the motion spurious artefacts specific for the investigated sample with an auto-calibrated definition of exclusion thresholds based on similarity among subjects/volumes. Further studies on resting state and task-fMRI data are necessary to test these temporal masks obtained with clustering and MD applied to different approaches of scrubbing (Churchill et al., 2015).

On the other hand, in a low-motion dataset, we demonstrated that an optimized pre-processing pipeline, using recent advances in resting state pre-processing field, can reduce the influence of spurious signals on BOLD signal related to task. The regression of only realignment parameters cannot achieve the best results in signal variability and strength of task. In case of low-motion dataset a more robust results can be obtained with the regression of physiological signals, derived from data itself, alone or combined with realignment parameters. These results underlined the choice of optimized pre-processing pipelines targeted on sample of interest. Also in this case, different orders of pre-processing steps need to be validate in larger cohorts (Churchill et al., 2015) or in high-motion datasets.